THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 1 Capitole (UT1 Capitole)

Présentée et soutenue par :
Christophe Gaillac

le mercredi 3 mars 2021

Titre :

Some Problems Related to Random Coefficients Models and Data Combination in Economics

---

École doctorale et discipline ou spécialité :
ED MPSE : Sciences Économiques

Unité de recherche :
UMR 5314 TSE RECHERCHE

Directeur/trice(s) de Thèse :

Eric Gautier, Professor at the University of Toulouse Capitole

Jury :

Marine Carrasco, University of Montréal	Examinateur
Enno Mammen, Heidelberg University	Examinateur
Jean-Pierre Florens, University of Toulouse Capitole	Rapporteur
Karine Van Der Straeten, Toulouse School of Economics, CNRS	Rapporteur
Xavier D'Haultfoeuille, CREST-ENSAE	Membre Invité

Contents

Remerciements. Cette thèse est une étape qui me permet surtout de mesurer à quel point j'ai été chanceux jusque-là.

Très peu de doctorants ont eu la chance d'avoir un directeur de thèse comme Éric Gautier. Je lui suis infiniment reconnaissant de m'avoir consacré autant de temps et d'énergie durant ces années, ainsi que de m'avoir donné l'envie de faire un doctorat en Économie en me donnant un sujet de thèse passionnant et stimulant. Je le remercie aussi très sincèrement pour les qualités pédagogiques de son encadrement, la pertinence de ses conseils, la finesse de ses raisonnements, et pour la patience dont il a fait preuve pour me tempérer et, je l'espère, m'avoir fait progresser.

Je suis vraiment très heureux que Marine Carrasco et Enno Mammen m'aient fait l'honneur d'accepter de rapporter cette thèse et je les remercie sincèrement pour leurs commentaires.

Je suis aussi reconnaissant à Jean-Pierre Florens, qui a su me donner le goût pour les problèmes inverses en première année de thèse, et à Karine Van Der Straeten pour avoir tout les deux accepté d'être membres de mon jury. Je tenais à la remercier chaleureusement avec Michel Le Breton, pour avoir gardé leurs portes ouvertes cette année, avoir pris le temps de comprendre mon job market paper et de m'aiguiller sur les applications.

J'ai peu de mots pour exprimer ma gratitude envers Xavier D'Haultfœuille. Je le remercie de m'avoir fait confiance à mon arrivée au CREST et tout au long de nos collaborations. Je lui suis redevable, entre beaucoup d'autres choses, de ses idées lumineuses et toute l'énergie positive qu'il a déployée, alors même qu'il est visiblement au four et au moulin. Je remercie aussi sincèrement Arnaud Maurel pour toutes ces qualités qu'il partage avec Xavier. J'ai beaucoup appris en travaillant avec eux deux.

Je tiens à remercier particulièrement deux de mes anciens collègues de bureau et amis, Jérémy L'Hour et Victor Lyonnet, avec qui je prends un grand plaisir à collaborer et à faire de l'épistémologie au quotidien. J'ai eu aussi la chance de faire un bout de chemin avec Yannick Guyonvarch, que je remercie sincèrement pour nos discussions économétriques matinales dans la bonne humeur, qui aboutiront sûrement.

J'ai pris beaucoup de plaisir à travailler avec Bruno Crépon, que je remercie pour nos riches discussions et nos projets enthousiasmants, qui ne sont pas dans cette thèse mais qui ont occupé une place importante. Je remercie aussi Jens Ludwig, Sendhil Mullainathan, tous les membres de la Team VADORE : Guillaume Bied, Phillipe Caillou, Elia Pérennès, et Michèle Sebag, ainsi que nos collègues de Pôle Emploi, notamment Chantal Vessereau, Cyril Nouveau, Emmanuel Chion, et David Buchner, pour leur ouverture d'esprit et leur enthousiasme pour faire avancer nos projets interdisciplinaires.

Je remercie du fond du cœur tou.te.s ceux.celles qui ont accepté de donner de leur temps pour m'écouter et me conseiller toujours avec justesse, notamment et sans prétendre à l'exhaustivité, Pierre Alquier, Andrii Babii, Aline Bonami, Stephane Bonhomme, Christian Bontemps, Frederico Bugni, Cristina Butucea, Nicolas Chopin, Marco Cuturi, Arnak Dalalyan, Laurent Davezies, Renato Gomes, Olivier De Groote, Emmanuel Guerre, Vishal Kamat, Maximilian Kasy, Louise Laage, Xavier Lambin, Elia Lapenta, Pascal Lavergne, Guillaume Lecué, Thierry Magnac, Franck Malherbet, Sophocles Mavroeidis, Nour Meddahi, Isabelle Méjean, François Poinas, Vincent Pons, Stefan Pollinger, Matthew Masten, Andrew Rhodes, Adam Rosen, Alain Trognon, Sacha Tsybakov, Arne Uhlendorff, Ao Wang, Nicolas Werquin, Martin Weidner, et Frank Windmeijer.

Je suis vraiment fier de faire partie du laboratoire de recherche de TSE, où j'ai eu la chance de croiser, de travailler, et d'être aidé par beaucoup de gens exceptionnels. Ma gratitude va à tous mes professeurs, collègues, et amis à TSE pour la qualité de leurs conseils et de leur soutien pendant toute la thèse, et en particulier pendant cette dernière année de préparation du marché. Je remercie en particulier Nour Meddahi pour son écoute et son énergie incroyable dépensée pour les doctorants de TSE. Merci aussi à Ludmila Namolovan et Elvire Jalran pour leur aide durant la thèse et pour l'organisation de la soutenance.

Je suis aussi vraiment fier d'avoir fait une longue escale au CREST et à l'ENSAE, qui m'a donné l'opportunité de faire partie des deux départements de Statistique et d'Economie, où les portes sont toujours ouvertes, quelque soit la question. Je remercie tous mes collègues et amis de ces deux départements pour m'avoir accueilli comme jeune chercheur puis assistant, m'avoir donné beaucoup de plaisir à travailler et à apprendre, et pour être une source continuelle d'inspiration. Je remercie Francis Kramarz, pour m'avoir fait découvrir et aimer l'économétrie en 1ère année à l'X et pour la générosité de son encadrement au CREST. J'ai une pensée particulière pour mes ancien.nes collègues à l'ENSAE ainsi que mes chef.fe.s, qui m'ont donné toutes les conditions pour pouvoir concilier ma recherche avec mon travail d'assistant : Pierre Biscourp, Laurent Davezies, Rosalinda Solotareff, Corentin Trévien, et Lionel Wilner.

J'ai évité jusque-là la comparaison de la thèse avec un ultra-trail, mais les deux ressentis sont très proches. Comme pour un ultra, je ne serai pas arrivé au bout sans ma famille et mes ami.e.s, et il me faudrait beaucoup plus que ces pages pour les remercier.

Je ne mesure pas la chance que j'ai d'avoir été soutenu par les familles Gaillac et Le Jouan. Merci pour tous leurs encouragements chaleureux, parfois amusés mais toujours de bon ton. Je suis aussi infiniment redevable à tou.te.s mes ami.e.s : les palois.e.s, les raideur.e.s et associé.e.s, de Fermat, et de la Sorbonne, qui ont été des appuis vitaux au quotidien pour me permettre de garder l'équilibre et d'avancer.

Enfin, j'ai une chance incommensurable d'être et d'avoir été aidé par Béatrice, Thierry, Alexandre, Nicolas, et Estelle, qui m'ont toujours poussé à me dépasser et à aller chercher plus loin sur tous les plans. Je conclue en remerciant Delphine pour sa fine compréhension de mes mécanismes psychologiques et pour avoir contribué très largement à faire de cette thèse un moment vraiment heureux.

Je dédie ce manuscrit à Guillaume Razimbaud qui, contre son habitude, aurait sûrement été au premier rang.

Résumé Cette thèse comporte cinq chapitres qui correspondent à des articles soumis ou en cours de publication. Les trois premiers étudient des modèles à coefficients aléatoires. Le troisième chapitre est aussi un problème de combinaison de données, ce qui fait le lien avec le quatrième. Le cinquième chapitre contient l'analyse des propriétés de la décomposition en valeurs singulières (SVD) d'un opérateur qui intervient dans le problème inverse du second chapitre.

Le premier chapitre étudie l'identification nonparamétrique de la distribution de l'hétérogénéité inobservée dans certains modèles à coefficients aléatoires lorsque les régresseurs ont une variation limitée, avec un support possiblement discret mais dénombrable. Nous exhibons un compromis dans l'identification de la distribution des coefficients aléatoires entre la taille des classes nonparamétriques que l'on considère pour ces distributions et la variation des régresseurs. Nous obtenons de nouveaux résultats dans un modèle linéaire où les coefficients, le terme d'erreur et les pentes, sont aléatoires et indépendants des regresseurs, mais aussi dans des modèles à coefficients aléatoires à choix binaire ou de données de panel.

Le second chapitre se concentre sur l'estimation de la densité jointe des coefficients dans le modèle linéaire à coefficients aléatoires. Nous considérons le cas de régresseurs dont le support est possiblement limité, *i.e.*, est un sous-ensemble strict de l'espace. Nous imposons que les coefficients, hormis l'intercept, n'aient pas de queues épaisses. Nous obtenons des bornes inférieures sur le risque minimax pour un large ensemble de régularités. Certaines classes de régularité autorisent des vitesses de convergence polynomiales voire quasi-paramétriques. Nous présentons un estimateur qui est optimal au sens minimax ainsi qu'une règle de sélection basée sur les données pour l'estimation adaptative.

Le troisième chapitre considère l'*inférence écologique*, qui est utilisée par exemple pour prévoir la probabilité de voter conditionnellement à la race, lorsque l'on observe seulement le taux de participation et la composition raciale de différents bureaux de votes. De manière plus générale, ce problème consiste, à partir d'un échantillon de distributions marginales (les taux de participation et les compositions raciale) de deux variables discrètes (le vote et la race) obtenues pour différents groupes (agrégées

par bureaux de votes), à prévoir la table de contingence sachant les distributions marginales. Ce problème tombe sous la nomenclature de la *combinaison de données*, puisqu'il utilise deux jeux de données qui ne peuvent pas être appareillés au niveau individuel (les résultats éléctoraux et le recensement). En utilisant la loi des probabilités totales, ce problème peut aussi se voir comme la prévision de coefficients aléatoires sachant la variable dépendante et des régresseurs dans un système linéaire à coefficients aléatoires avec des régresseurs bornés. Dans ce contexte, je caractérise l'ensemble identifié sans hypothèse supplémentaire et montre que l'hypothèse d'absence *d'effets contextuels* permet l'identification. Puis, je développe un estimateur qui est adaptatif et optimal au sens minimax. Enfin, j'applique mes méthodes pour estimer l'effet du porte-à-porte sur les résultats des candidats à la présidentielle pour deux catégories d'électeurs, construites à partir des votes passés, dans l'expérience randomisée de Pons (2018). Je montre, seulement à partir de données agrégées au niveau des bureaux de votes, que le porte-à-porte agit particulièrement en persuadant les électeurs indécis, plutôt qu'en mobilisant ceux qui sont déjà convaincus.

Le quatrième chapitre construit un nouveau test des anticipations rationnelles d'une variable individuelle, basé sur les distributions marginales des réalisations et des anticipations subjectives de celle-ci. Ce test peut être appliqué dans de nombreux contextes, notamment dans la situation fréquente où les réalisations et les anticipations sont observées dans deux jeux de données différents qui ne peuvent pas être appareillés. Nous montrons que le fait que l'on puisse rationaliser à partir des données l'hypothèse d'anticipations rationnelles est équivalent au fait que la distribution des réalisations domine au second ordre la distribution des croyances. L'hypothèse nulle peut alors être réécrite comme un système de plusieurs inégalités de moments et de contraintes d'égalité, pour lequel des tests ont été développés dans la littérature. Le test est robuste aux erreurs de mesures sous certaines hypothèses et peut être étendu pour tenir compte des chocs agrégés. Enfin, nous testons la rationalité des anticipations salariales. Tandis que les individus ne se trompent pas en moyenne quant à leurs salaires futurs, notre test rejette l'hypothèse d'anticipations rationnelles.

Le dernier chapitre considère la transformée de Fourier tronquée sur $[-c, c]$ et agis-

sant sur l'espace $L^2(\cosh(b|\cdot|))$ sur lequel elle est injective. Nous donnons des bornes non-asymptotiques supérieures et inférieures sur les valeurs singulières avec un comportement qualitatif similaire en m (l'index), b, et c. Les bornes inférieures sont utilisées pour obtenir des vitesses de convergence pour le prolongement analytique stable de fonctions potentiellement non limitées en fréquences dont la transformée de Fourier appartient à $L^2(e^{b|\cdot|})$. Nous donnons aussi des bornes sur la norme sup des fonctions singulières. Enfin, nous proposons une méthode d'implémentation numérique pour calculer la SVD et l'appliquons au prolongement analytique stable quand la fonction est observée avec erreur sur un intervalle. Dans l'application numérique, nous considérons deux cas. D'abord lorsque la fonction à extrapoler admet une transformée de Fourier qui n'est pas à support compact. Puis lorsqu'elle a un support compact mais qu'il est inconnu.

Abstract This PhD thesis consists of five chapters which correspond to different submitted or forthcoming articles. The first three study some random coefficients models. The third chapter, being also a data combination problem, draws the link with the fourth one. The fifth contains the analysis of the properties of the singular value decomposition of an operator which intervenes in the inverse problem of the second chapter.

The first chapter studies nonparametric identification of the distribution of the unobserved heterogeneity in some random coefficients models when regressors have limited variation, when their support is possibly discrete but countable. We exhibit a trade-off in the identification of the distribution of the random coefficients between the size of the nonparametric classes of law of the coefficients we consider and the variation of the regressors. We provide new results on the linear model where the coefficients, the error term and the slopes, are random and independent of the regressors, but also in some random coefficients models of binary choice or panel data.

The second chapter focuses on the linear random coefficient model. We consider regressors which can have a support which is a proper subset. We assume that the slopes do not have heavy-tails, but we do not make integrability assumptions on the error term. Lower bounds on the minimax risk for the estimation of the joint density of the random coefficients are obtained for a wide range of smoothness. Some allow for polynomial and nearly parametric rates of convergence. We present a minimax optimal estimator and a data-driven rule for adaptive estimation.

The third chapter considers the *ecological inference*, which is used for example to predict the probability to vote according to race using the turnout rate and the racial composition of different precincts. More generally, this problem consists, from a sample of marginal distributions (the turnout rates and the racial compositions) of two discrete variables (the vote and the race) obtained for different groups (aggregated by precincts), in predicting the contingency tables knowing the margins (the turnout rate and the racial composition). This is can be framed as a *data combination* problem, because it uses data from two datasets which can not be matched at an individual level (the electoral results and the census). This is also related to the prediction

of the random coefficients given the outcomes and regressors in a system of linear random coefficients equations with bounded regressors. I characterize the identified set without further assumptions and show that the *no contextual effects* assumption yields point identification. Then, I develop a minimax adaptive estimator. Finally, I apply my methods to estimate the effect of door-to-door visits on vote shares among two categories of voters, based on past votes, in the randomized experiment of Pons (2018). My results suggest, using precinct data only, that canvassing is especially effective through persuasion of undecided voters, rather than mobilization of convinced ones.

The fourth chapter build a new test of rational expectations based on the marginal distributions of realizations and subjective beliefs. This test is widely applicable, including in the common situation where realizations and beliefs are observed in two different datasets that cannot be matched. We show that whether one can rationalize rational expectations is equivalent to the distribution of realizations being a mean-preserving spread of the distribution of beliefs. The null hypothesis can then be rewritten as a system of many moment inequality and equality constraints, for which tests have been recently developed in the literature. The test is robust to measurement errors under some restrictions and can be extended to account for aggregate shocks. Finally, we apply our methodology to test for rational expectations about future earnings. While individuals tend to be right on average about their future earnings, our test rejects rational expectations.

The fifth chapter considers the Fourier transform truncated on $[-c, c]$ acting on the space $L^2(\cosh(b|\cdot|))$ on which it is injective. We give nonasymptotic upper and lower bounds on the singular values with similar qualitative behavior in m (the index), b, and c. The lower bounds are used to obtain rates of convergence for stable analytic continuation of possibly nonbandlimited functions whose Fourier transform belongs to $L^2(e^{b|\cdot|})$. We also derive bounds on the sup-norm of the singular functions. Finally, we propose a numerical method to compute the singular value decomposition and apply it to stable analytic continuation when the function is observed with error on an interval. In the numerical application we consider two cases. First, when the function to extrapolate does no admit a compactly supported Fourier transform. Second, when it is compactly supported but the support is unknown.

Chapter 1

Introduction

Heterogeneity of individual economic agents is a fundamental concept. Collecting data allows to access the outcome variables and some of the characteristics of these agents, constituting the observed heterogeneity. However, the amount of information we observe is always limited by the cost to obtain extensive data at large scale, the need to protect data privacy, or because there is information which is impossible to collect. Thus, there are heterogeneous elements which are known by the agents but remain unobserved to the econometrician. These can be modelled as random variables, or vectors when there are multiple sources of unobserved heterogeneity. Accounting for these variables in the economic modelling can lead to important quantitative differences. We adopt a nonparametric approach in all this PhD thesis.

The first three chapters study some random coefficients models, which have of multiple sources of unobserved heterogeneity. Here, it is crucial to weaken the assumptions used for identification and estimation to be credible in applications. The second part of the thesis, chapters three and four, consider some *data combination* problems. Data combination leverages two different datasets which cannot be matched at the individual level to learn about features of the joint distribution. This allows to test the researchers' working assumptions, like the rational expectations hypothesis, without collecting a full dataset. Here, the third chapter relates two parts of the thesis: it is both a data combination problem and a system of random coefficients equations.

Finally, the fifth chapter analyzes the properties of the singular value decomposition (SVD) of an operator intervening in the inverse problem of the second chapter. This SVD is also particularly well suited to perform stable analytic continuation.

The five chapters correspond to different submitted or forthcoming articles. This introductory chapter does not make a complete literature review of these topics, which is done in the subsequent chapters. It motivates the problems we consider and then details our contributions.

1.1 Nonparametric identification in some random coefficients models

1.1.1 Motivation

In many cases, the economist might be not only interested by the average effect of a covariate on an outcome, but also by the heterogeneity of this effect, assuming that there is heterogeneity. Consider for example the impact of the parents' income on children test scores or the impact of subsidies on consumption. To fix ideas, consider the linear random coefficients model,

$$Y_i = \alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{X}_i \tag{1.1}$$

$$(\alpha_i, \boldsymbol{\beta}_i) \text{ and } \boldsymbol{X}_i \text{ are independent,} \tag{1.2}$$

where $\boldsymbol{\beta}_i$ and $\boldsymbol{X}_i$ are vectors of size p. In (1.1)-(1.2), each individual i is affected differently by the regressors $\boldsymbol{X}_i$. Of course, using $\boldsymbol{\beta}_i = \mathbb{E}[\boldsymbol{\beta}_i] + (\boldsymbol{\beta}_i - \mathbb{E}[\boldsymbol{\beta}_i])$ and $\epsilon_i := \alpha_i + (\boldsymbol{\beta}_i - \mathbb{E}[\boldsymbol{\beta}_i])^\top \boldsymbol{X}_i$, it is related to the heteroscedastic linear regression model with one source of unobserved heterogeneity where the researcher is only interested by the average effect $\mathbb{E}[\boldsymbol{\beta}_i]$,

$$Y_i = \epsilon_i + \mathbb{E}[\alpha_i] + \mathbb{E}[\boldsymbol{\beta}_i]^\top \boldsymbol{X}_i, \qquad \mathbb{E}[\epsilon_i | \boldsymbol{X}_i] = 0.$$

Moreover, in (1.1)-(1.2), the dependence structure in the random vector $\boldsymbol{\gamma}_i := (\alpha_i, \boldsymbol{\beta}_i)$ is left unrestricted. In this model, the researcher can be interested in more than the average of the vector $\boldsymbol{\gamma}_i$ and even might want to recover the distribution $f_{\boldsymbol{\gamma}}$ of the unobserved heterogeneity. For example, a policymaker might be interested by the distribution of the marginal propensity to consume following a subsidy, to better calibrate the following ones.

A first approach is to assume that the true distribution $f_{\boldsymbol{\gamma}}$ belongs to a parametric family and then to estimate the model by maximum likelihood (see, *e.g.*, Dent and Hildreth, 1977; Imai et al., 2008). However, there is a misspecification risk and this assumption can drive the results (see, *e.g.*, Heckman and Singer, 1984). Another approach is to adopt a Bayesian point of view and to use a prior on the distribution of $\boldsymbol{\gamma}$ (see, *e.g.*, Griffiths et al., 1979; King, 1997; Imai et al., 2008). If the choice of the prior might not have an impact asymptotically, it can lead to different results in finite sample. A nonparametric approach remains agnostic on the distribution $f_{\boldsymbol{\gamma}}$. Taking into account these multiple sources of heterogeneity nonparametrically comes with various restrictions on the variation of the regressors as well as on the nonparametric class of densities $f_{\boldsymbol{\gamma}}$ to identify the distribution $f_{\boldsymbol{\gamma}}$ (*i.e.*, such that for all observed distribution of the data $\mathbb{P}_{\boldsymbol{X},Y}$, there exists a unique distribution of the unobserved heterogeneity $f_{\boldsymbol{\gamma}}$ which could have generated it through the model).

To give a glimpse of the trade-off between the variation of the regressors and the size of class of densities $f_{\boldsymbol{\gamma}}$ we can allow, we start with the deconvolution problem with two samples: one of the error and one of the sum of the signal and the error. This problem can be viewed as a particular linear random coefficients model (1.1)-(1.2), where the regressor X has dimension 1 and two points of support, $\{0, 1\}$ without loss of generality,

$$Y = \alpha + \beta X, \quad (\alpha, \beta) \perp X.$$

In this model the variation of the regressor is very limited and identification assumptions impose independence between α and β (see, *e.g.*, Carroll and Hall, 1988; Beran and Hall, 1992; Carrasco and Florens, 2011; Gaillac and Gautier, 2019*a*). However, in some cases there might be a deep underlying parameter Θ (*e.g.*, related to tastes, type of education, ideology, etc) through which the heterogeneous impact $\beta(\Theta)$ is related to the error term $\alpha(\Theta)$. When the support of X is richer but possibly discrete, we can identify the distribution of $\boldsymbol{\gamma}$ without independence.

Let us describe the trade-off on (1.1)-(1.2). Identification is based on the characterization of the distribution of $\boldsymbol{\gamma}$ from the knowledge of the conditional distribution of the outcome Y given the regressors $\boldsymbol{X}$, namely

$$\mathbb{P}_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x}) = \mathbb{P}_{\boldsymbol{\gamma}^\top \boldsymbol{x}}(\cdot) \quad (\text{using } (1.1) - (1.2)),$$

for all $\boldsymbol{x}$ in the support of $\boldsymbol{X}$, which we denote by $\mathbb{S}_{\boldsymbol{X}}$. This shows that these distributions are 1-dimensional projections of the distribution of $\boldsymbol{\gamma}$. For all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$, $\mathbb{P}_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ is characterized by its characteristic function, which is related to the Fourier transform of the distribution of $\boldsymbol{\gamma}$ by,[1] for all $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\left[e^{itY}\middle|\boldsymbol{X} = \boldsymbol{x}\right] &= \mathbb{E}\left[e^{it\boldsymbol{\gamma}^\top \boldsymbol{x}}\right] \quad (\text{using } (1.1) - (1.2)) \\
&= \int_{\mathbb{R}^{p+1}} e^{it\boldsymbol{g}^\top \boldsymbol{x}} f_{\boldsymbol{\gamma}}(\boldsymbol{g}) d\boldsymbol{g} \\
&= \mathcal{F}\left[f_{\boldsymbol{\gamma}}\right](t, t\boldsymbol{x}). \qquad (1.3)
\end{aligned}$$

Identification of $f_{\boldsymbol{\gamma}}$ amounts to recovering it from the knowledge of its Fourier transform on a double cone with apex 0, namely

$$\left\{t(1, \boldsymbol{x}),\ t \in \mathbb{R},\ \boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}\right\}.$$

This description of the problem underlines two points. First, if the regressors have support the space $\mathbb{R}^p$, then, using (1.3), the Fourier transform $\mathcal{F}\left[f_{\boldsymbol{\gamma}}\right]$ is known everywhere (see, *e.g* Beran and Hall, 1992; Hoderlein et al., 2010; Holzmann and Meister, 2019). Using the injectivity of the Fourier transform, identification is obtained with-

[1] The Fourier transform of f in $L^1(\mathbb{R}^p)$ is $\mathcal{F}[f] : \boldsymbol{x} \mapsto \int_{\mathbb{R}^p} e^{i\boldsymbol{b}^\top \boldsymbol{x}} f(\boldsymbol{b}) d\boldsymbol{b}$ and $\mathcal{F}[f]$ is also the Fourier transform in $L^2(\mathbb{R}^p)$.

out restrictions on the distribution f_{γ}. Second, if the regressors have limited variation, *i.e.*, a proper subset of $\mathbb{R}^p$, then one can to restrict the size of the class of possible distributions f_{γ} to obtain nonparametric identification (see, *e.g*, Beran and Hall, 1992; Beran and Millar, 1994; Masten, 2017; Gaillac and Gautier, 2019*a*). A degenerate situation where the support of the regressors has an empty interior is the nonlinear model:

$$Y_i = \gamma_{1,i} + \gamma_{2,i} X_i + \gamma_{3,i} X_i^2, \qquad \gamma_i \perp X_i.$$

Here, f_{γ} is not nonparametrically identified, whatever the support of the variable X_i. Thus, it is of interest to look for the type of nonlinear models which can lead to the nonparametric identification of f_{γ}.

1.1.2 Contribution

The first chapter of this thesis describes the trade-off in the identification of the distribution of the random coefficients between the size of the nonparametric classes of law of the coefficients and the variation of the regressors. We consider the random coefficients linear model, including the deconvolution, the random coefficients binary model, and some panel data models such as single-index panel data models. In these models, we find moment conditions on f_{γ} such that it is determined by the knowledge of the set of distributions

$$\{\mathbb{P}_{Y|\boldsymbol{X}=\boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}\} \tag{1.4}$$

when $\mathbb{S}_{\boldsymbol{X}}$ is a proper subset, possibly discrete. In the linear random coefficients model, this is related to the conditions for a measure to be determined by some 1-dimensional projections (see, *e.g.*, Cuesta-Albertos et al., 2007; De Jeu, 2003; Infusino, 2016). We give examples of parametric distributions satisfying these conditions, allowing heavy-tails. We also show how to include knowledge about the sign or the tails of the coefficients to weaken the identifying assumptions.

1.2 Nonparametric estimation in the linear random coefficients model

1.2.1 Motivation

The nonparametric estimation of the density of the random coefficients in the linear model has received considerable attention. This problem is closely related to computerized tomography (see, *e.g.*, Natterer, 1986), also known as medical X-ray tomography. This consists in recovering an image of the internal structure of an object only through the attenuation of the X-rays' intensity crossing the object. Initial and final intensities are related to the integral of the density f of the object's stucture on the line of the X-ray (see, *e.g.*, Alquier et al., 2011*a*). As noticed in Beran et al., 1996*a*, with one regressor in the linear random coefficients model (1.1)-(1.2), the distributions (1.4) are related to $f_{\alpha,\beta}$ through, for all $x \in \mathbb{S}_X$,

$$f_{Y|X}(\cdot|x) = \int_{\mathbb{R}} f_{\alpha,\beta}(\cdot - bx, b)db$$

hence are integrals indexed by $x \in \mathbb{S}_X$ along the line $\{(a,b) \in \mathbb{R}^2,\ (a,b)(1,x)^\top = y\}$. These problems are *linear inverse problems*, where the object of interest is related to an observable quantity by a linear operator. Indeed, considering a different renormalization

$$U := \frac{Y}{\sqrt{1 + \|\boldsymbol{X}\|_2^2}} \quad \text{and} \quad \boldsymbol{S} := \frac{1}{\sqrt{1 + \|\boldsymbol{X}\|_2^2}} \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix},$$

Hoderlein et al., 2010 show that the model (1.1)-(1.2) can be rewritten as

$$U = \boldsymbol{S}^\top \boldsymbol{\gamma}, \quad \boldsymbol{S} \perp \boldsymbol{\gamma}, \tag{1.5}$$

where $\boldsymbol{S} \in \mathbb{S}^p$ (see Figure 1-1). Then, the distribution of the outcome U conditional on $\boldsymbol{S}$ is related to $f_{\boldsymbol{\gamma}}$ via, for all $(\boldsymbol{s}, u) \in \mathbb{S}_{\boldsymbol{S}} \times \mathbb{R}$,

$$f_{U|\boldsymbol{S}}(u, s) = R[f_{\boldsymbol{\gamma}}](\boldsymbol{s}, u), \tag{1.6}$$

where R is the *Radon transform*, for all $(\boldsymbol{s}, u) \in \mathbb{S}^p \times \mathbb{R}$,

$$R[f](\boldsymbol{s}, u) = \int_{\{\boldsymbol{g} \in \mathbb{R}^{p+1}:\ \boldsymbol{g}^\top \boldsymbol{s} = u\}} f(\boldsymbol{g}) d\mathbb{P}_{u,s}(\boldsymbol{g}),$$

$\mathbb{S}^p$ is the unit sphere in $\mathbb{R}^{p+1}$, and $\mathbb{P}_{u,s}$ is the Lebesgue measure on the affine space $\{\boldsymbol{g} \in \mathbb{R}^{p+1} :\ \boldsymbol{g}^\top \boldsymbol{s} = u\}$.

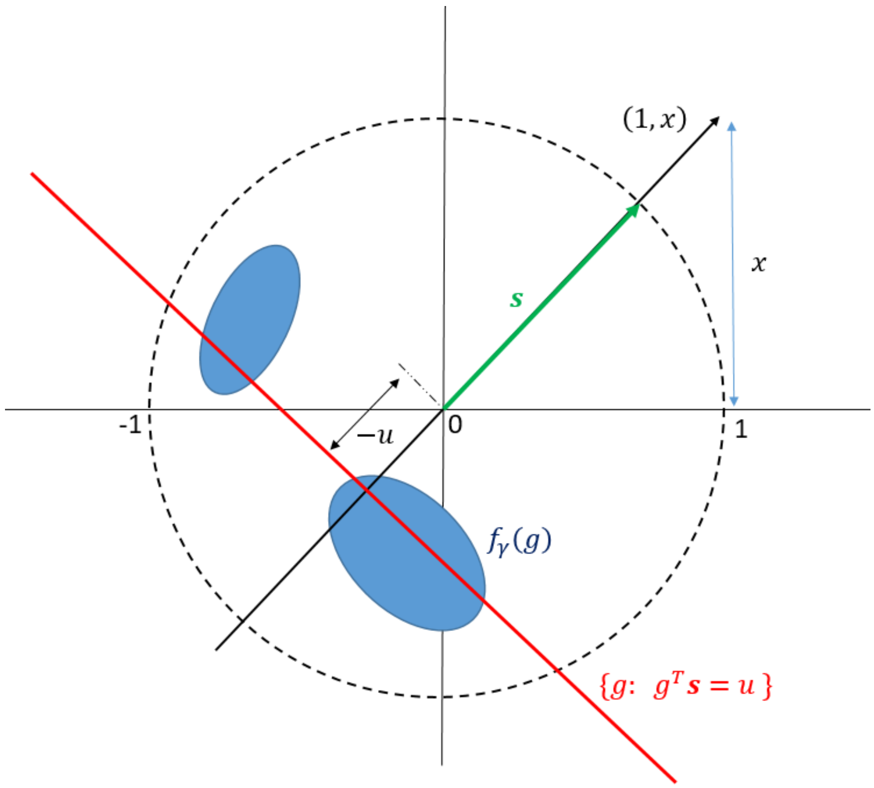


Figure 1-1: Parametrisation of the random coefficients model (1.5). The red line is the set of points $\boldsymbol{g}$ such that $\boldsymbol{g}^\top \boldsymbol{s} = u$, where $u < 0$ and $\boldsymbol{s}$, in green, belongs to the unit sphere in $\mathbb{R}^2$. The blue ellipses are a representation of the object f_γ, which is here a mixture of two uniforms.

This formulation (1.6) is used in Hoderlein et al., 2010 to build a nonparametric estimator of f_γ based on a kernel regularized inverse of the Radon transform, assuming that the regressors $\boldsymbol{X}$ have full support. Assuming that the density of the angles $f_{\boldsymbol{S}}$ is bounded away from 0 as well as Sobolev smoothness conditions, they show that the estimator converges at a polynomial rate. However, the former assumption imposes

Cauchy-type tails on the regressors $\boldsymbol{X}$ (see 4.3 in Hoderlein et al., 2010 and Holzmann and Meister, 2019 for an alternative estimator under light-tail assumptions). When regressors have limited variation in (1.1)-(1.2), the counterpart of this problem of is the *limited angle* tomography, when X-rays go only through a limited range of angles (see, *e.g.*, Frikel, 2013; Hohmann and Holzmann, 2015, and Figure 1-1). By analogy when $p = 1$ with a related Gaussian sequence space model studied in Hohmann and Holzmann, 2015, lighter tails of the regressors should impact the degree of ill-posedness of the problem. There, an analogous to the case where the regressors are bounded and $p = 1$ is a severely ill-posed problem where the rate of the minimax risk in Sobolev type ellipsoids relative to the right-singular functions of the Radon transform is logarithmic.

This might be the reason why, up to now, applied researchers have preferred models such as the quantile regression to describe unobserved heterogeneity (see, *e.g.*, Koenker and Bassett, 1978; Koenker, 2005). Quantile regression can be written as a linear random coefficients model where the coefficients are functions of a scalar uniform distribution,

$$Y = \boldsymbol{\beta}_U^\top \boldsymbol{X},$$
$$U|\boldsymbol{X} \sim \mathcal{U}[0,1].$$

However, it is hard to argue for such degeneracy of the heterogeneity. It is thus of importance to weaken the usual assumptions imposed on either the data or the model in the linear random coefficients model (1.1)-(1.2) to allow its wider use by practitioners. Specifically, because it is hardly ever the case in practice that regressors have a support which is the whole space, it is important to allow the regressors to have limited variation while authorizing $p > 1$ and showing that some polynomial rates of convergence can be attained in this context.

1.2.2 Contribution

The third chapter focuses on adaptive estimation of the density $f_{\boldsymbol{\gamma}}$ in the linear random coefficients model (1.1)-(1.2) when regressors have limited variation. We assume that the margins of $\boldsymbol{\beta}$ (but not of α) do not have heavy tails but can have noncompact support. This is an inverse problem (see (1.3)) whose operator is not compact when no integrability assumption is made on α.[2] This makes it difficult to prove rates of convergence, even for estimators which do not rely explicitly on the SVD such as the Tikhonov and Landweber method (see, *e.g.*, Carrasco et al., 2007). However, it can be decomposed as the composition of two operators which are easier to study. More precisely, assuming that there exists $x_0 > 0$ such that $[-x_0, x_0]^p \subseteq \mathbb{S}_{\boldsymbol{X}}$, we consider the decomposition of the left-hand side of (1.3), for all $(t, \boldsymbol{u}) \in \mathbb{R} \times [-1, 1]^p$,

$$\mathcal{F}\left[f_{\boldsymbol{\gamma}}\right](t, tx_0\boldsymbol{u}) = \mathcal{F}_{tx_0}\left[\mathcal{F}_{1\text{st}}\left[f_{\boldsymbol{\gamma}}\right](t, \cdot)\right](\boldsymbol{u}), \tag{1.7}$$

where $\mathcal{F}_{tx_0}$ is the truncated Fourier transform operator, which, for all $c \neq 0$, to a function f in a weighted L^2 space associates its Fourier transform truncated to $[-c, c]^p$, $\mathcal{F}[f](c\cdot) \in L^2([-1, 1]^p)$, and $\mathcal{F}_{1\text{st}}$ is the partial Fourier transform with respect to the first variable. This decomposition eases the analysis because the nonasymptotic properties of the SVD of the truncated Fourier transform operator are known in at least two cases of weights (see, *e.g.*, Bonami and Karoui, 2016; Gaillac and Gautier, 2019*a*). The later are related to tail assumptions on the coefficients $\boldsymbol{\beta}_j$ for $j = 1, \ldots, p$: either compact support or without heavy tails. Moreover, the SVD is fast to compute using numerical schemes developed recently (see, *e.g.*, Osipov et al., 2013; Gaillac and Gautier, 2019*a*).

Following (1.7), we build a three steps estimator. First, for $|t| \leq T$, we use an approximation of $\mathcal{F}_{1\text{st}}\left[f_{\boldsymbol{\gamma}}\right](t, \cdot)$ using the regularized inverse of the truncated Fourier operator $\mathcal{F}_{tx_0}$. It involves spectral cut-off. Second, because the singular values of $\mathcal{F}_{tx_0}$

[2] Similarly to Carrasco et al., 2007 in the deconvolution problem, we could alternatively consider an appropriate choice of reference spaces where the operator is compact.

go to 0 as t goes to 0, using the SVD for estimation is problematic and we would rely on few to none coefficients. Rather, we rely on an interpolation strategy for $t \in [\epsilon, \epsilon]$, where $0 < |t| < 1 < T$, which is of independent interest. Third, we use a regularized inverse of the partial Fourier transform with respect to the first variable to recover f_γ. We show that, for classes of sufficiently smooth functions, minimax polynomial rates of convergence can be obtained. We provide an adaptive estimator, *i.e.* whose parameter is automatically selected from the data, which attains the optimal rates of convergence up to a logarithmic term. The adaptive estimator is implemented in the R package RandomCoefficients, described in more details in the vignette Gaillac and Gautier (2019*d*).

1.3 Ecological inference as a data combination problem and a system of random coefficients equations

1.3.1 Motivation

Ecological inference uses data combined at the *group* level (see, *e.g.*, Robinson, 1950; King, 1997). Let us explain its principle using a simple yet striking illustration: the probability to vote for some candidates according to race for given precincts. Due to the secret ballot, this information is usually only accessible using surveys, which have several issues: there is attrition, misreporting, and they might not be available nationwide (see, *e.g.*, Burden and Kimball, 1998). On the contrary, the race $R \in \{1, \dots, d_R\}$ and the candidate choice $C \in \{1, \dots, d_C\}$ are observed nationwide and publicly, respectively in the census and the election returns, however aggregated at a group level g (usually the precincts). To put it differently, we do not have access to the individual discrete variables C and R, but to the margins of these variables, which are the vote shares and the racial composition of each precinct g, respectively

$$\boldsymbol{Y}_g := \begin{pmatrix} \mathbb{P}_g\left(C=1\right) \\ \vdots \\ \mathbb{P}_g\left(C=d_C\right) \end{pmatrix} \quad \text{and} \quad \boldsymbol{X}_g := \begin{pmatrix} \mathbb{P}_g\left(R=1\right) \\ \vdots \\ \mathbb{P}_g\left(R=d_R\right) \end{pmatrix}.$$

The fundamental indeterminacy of the ecological inference can be understood considering that the conditional distributions $\boldsymbol{B}_g$, or equivalently - as the margins are known - the contingency tables satisfy the law of total probability

$$\forall c = 1, \ldots, d_C, \quad \mathbb{P}_g\left(C=c\right) = \sum_{r=1}^{d_R} \mathbb{P}_g\left(C=c|R=r\right)\mathbb{P}_g\left(R=r\right), \tag{1.8}$$

which can be represented as

$$\underbrace{\begin{pmatrix} \mathbb{P}_g\left(C=1\right) \\ \vdots \\ \mathbb{P}_g\left(C=d_C\right) \end{pmatrix}}_{\boldsymbol{Y}_g} = \underbrace{\begin{pmatrix} \mathbb{P}_g\left(C=1|R=1\right) & .. & \mathbb{P}_g\left(C=d_c|R=1\right) \\ \vdots & & \vdots \\ \mathbb{P}_g\left(C=1|R=d_R\right) & .. & \mathbb{P}_g\left(C=d_c|R=d_R\right) \end{pmatrix}^\top}_{\boldsymbol{B}_g^\top} \underbrace{\begin{pmatrix} \mathbb{P}_g\left(R=1\right) \\ \vdots \\ \mathbb{P}_g\left(R=d_R\right) \end{pmatrix}}_{\boldsymbol{X}_g}$$

hence

$$\boldsymbol{Y}_g = \boldsymbol{B}_g^\top \boldsymbol{X}_g. \tag{1.9}$$

In the example, the matrix $\boldsymbol{B}_g$ contains the probabilities to vote for some candidate conditional on race. However, without additional restrictions, for each group, there are many possible tables compatible with the observed margins $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$. This is of particular importance as, among others, ecological inference is used in courts to rule according to the Voting Right Act, which prevents racial bloc voting (see, *e.g.*, King, 1997; Wakefield, 2004; Greiner and Quinn, 2010). Thus, it plays a central role in redistricting litigations, which is the process where political representatives adjust the boundaries of the electoral districts every decade in the US to take into account the evolution of the population. When politically biased, this is called *gerrymandering* and is prevented by the law, but is very hard to establish statistically due to the secret ballot. This question has been recently highlighted with the polarization of the US political debate (see, *e.g.*, McCarty et al., 2009). This problem is also encountered in other contexts in Economics, such as combining market level data with census data to perform demand analysis.

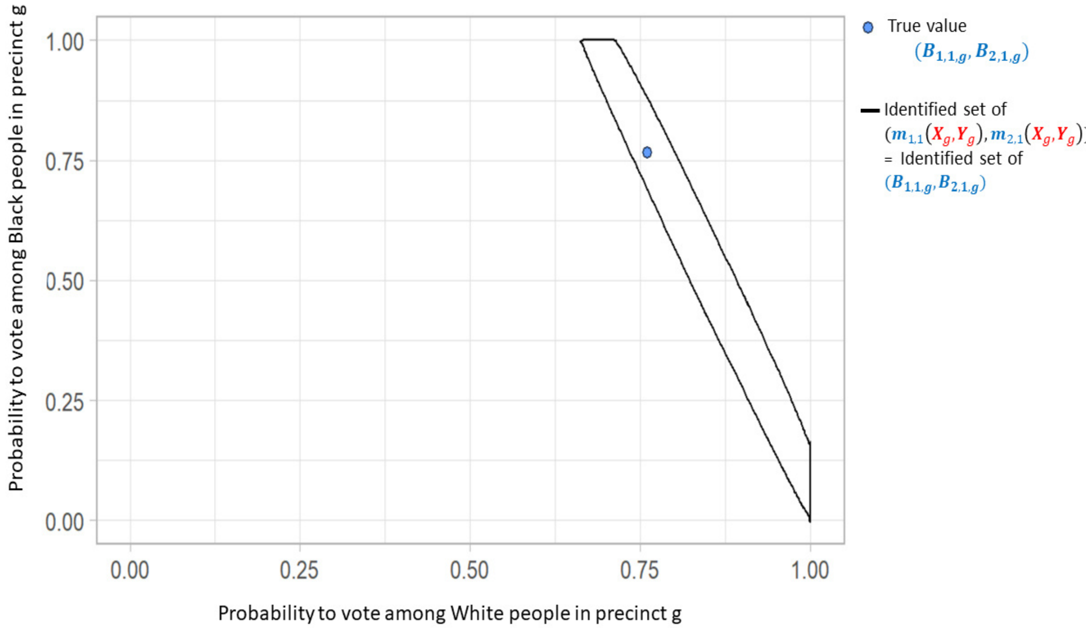
For a certain group g, the ecological inference problem is related to the *regression short and long* problem in econometrics (see, *e.g.*, Cross and Manski, 2002*a*), when the dependent variable C and the covariate R are discrete and observed in two different datasets, which can only be matched conditionning on a third discrete variable taking values $g = 1, \dots, G$ (*i.e.*, we observe the margins $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$). The point of view of the literature in econometrics (see, *e.g.*, Cross and Manski, 2002*a*; Molinari and Peski, 2006*a*; Fan et al., 2014, 2016; Manski, 2018; Jiang et al., 2020; Gaillac, 2020) is to focus on partial identification and to look for all the matrices $\boldsymbol{B}_g$ containing the conditional probabilities which are compatible with given margins $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$. Taking into account the constraints on the margins, this is the set[3]

$$\mathcal{I}(\boldsymbol{x}, \boldsymbol{y}) = \left\{ \boldsymbol{B} \in \mathcal{M}_{d_R, d_C}([0,1]),\ \boldsymbol{B}\mathbf{1} = \mathbf{1},\ \boldsymbol{y} = \boldsymbol{B}^\top \boldsymbol{x} \right\}. \tag{1.10}$$

However, the set $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y})$ is often too large for practitioners. Specifically, Figure 1-2 illustrates on an example, dealing with turnout among Black and White people in a particular precinct that this set is not informative about the way minorities vote (see, *e.g.* p359, Cross and Manski, 2002*a*, or the classical Dudley-Duncan bounds).

In Statistics and Political Science (see, *e.g.*, King, 1997; Wakefield, 2004; Imai et al., 2008), the ecological inference problem is related to the observation of a sample of these margins $(\boldsymbol{X}_g, \boldsymbol{Y}_g)$ for several groups g. Moreover, these matrices and margins are heterogeneous across groups. Thus, we treat the observed sample of margins for the groups, together with the unobserved and heterogeneous conditional distributions,

[3] Note that the set of associated joint distributions is known as the transportation polytope in the optimal transport literature (see, *e.g.*, Cuturi, 2012).



Notes: The identified set of $\boldsymbol{B}_g$ is delimited by the dark line. Racial composition is 72.2% White people, 24.4% Black people, and 3.4% others, and the turnout is 75.7%

Figure 1-2: Turnout among Black and White people in a precinct of the 2nd district of Florida in 2011, where I consider three race categories.

as random vectors and matrices drawn from a sampling distribution

$$(\boldsymbol{B}_g, \boldsymbol{X}_g, \boldsymbol{Y}_g) \sim \mathbb{P}_{\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y}}.$$

Thus, (1.9) together with the constraints on the margins, yield that $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y})$ satisfies the linear system of random coefficients equations

$$\forall c = 1, \ldots, d_C,\ \boldsymbol{Y}_c = \sum_{r=1}^{d_R} \boldsymbol{B}_{r,c} \boldsymbol{X}_r, \quad \forall r = 1, \ldots, d_R, \quad \sum_{c=1}^{d_C} \boldsymbol{B}_{r,c} = 1 \tag{1.11}$$

$$\forall c = 1, \ldots, d_C, \forall r = 1, \ldots, d_R, \quad \boldsymbol{B}_{r,c} \geq 0, \quad \boldsymbol{X}_{r,c} \geq 0, \quad \sum_{r=1}^{d_R} \boldsymbol{X}_r = 1 \quad . \tag{1.12}$$

Similarly to what have been introduced in Section 1.1 in the study of random coefficients models, the usual baseline assumption in the ecological inference literature, called *no contextual effects* (NCE), is the following independence in (1.11) between the random coefficients and the regressor

$$\text{(NCE)} \qquad \boldsymbol{B} \perp \boldsymbol{X}.$$

This can be viewed as the exogeneity of the regressor $\boldsymbol{X}$. In the voting example, this exogeneity assumption means that the probability that an individual of a given race go to the polls is independent of the racial composition of the precincts. We describe later some ways to relax this assumption, which might be strong for some applications.

The parameter of interest is the conditional expectation

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}) \in [0,1]^{d_R \times d_C} \mapsto \mathbb{E}\left[\boldsymbol{B} | (\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y})\right] \in \mathcal{M}_{d_R, d_C}([0,1]), \tag{1.13}$$

where $\mathcal{M}_{d_R,d_C}([0,1])$ are $d_R \times d_C$ matrices with elements in $[0,1]$ (see, *e.g.* King, 1997; Wakefield, 2004). In the example, this is the prediction of the probability to choose a candidate for people of a certain race given the turnout rate and racial composition. The parameter $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y})$ gives the best prediction of the heterogeneous conditional distribution $\boldsymbol{B}$ for given values of the margins $(\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y})$.

Again, standard approaches in the literature performing estimation of $\boldsymbol{m}$ are either parametric (see, *e.g.*, Imai et al., 2008) or Bayesian (King, 1997; Wakefield, 2004; Imai et al., 2008), but maintain parametric assumptions on the distribution of the rancom coefficients which can drive the results. The asymptotic properties of the nonparametric Bayesian method of Imai et al. (2008) are unknown and the method is limited, for computational reasons, to the case $d_R = d_C = 2$.

1.3.2 Contribution

The parameter $\boldsymbol{m}$ is a functional of the density $f_{\boldsymbol{B}}$, hence its study in the third chapter extends the first two chapters of this thesis. I relate the point of view of the econometrics literature with the random coefficients model representation. Indeed, without further assumptions, the identified set for $\boldsymbol{m}$, *i.e.*, the set of parameters compatible with the distribution of the data, is related to the one considered in Cross and Manski (2002*a*) and is exactly the one studied in Manski (2018) and Gaillac (2020). I do not include in this thesis Gaillac (2020) because it is not yet entirely polished.

My main contribution in this chapter is to obtain new nonparametric constructive point identification of $\boldsymbol{m}$ with two possibilities for the outcome $d_C = 2$ and an arbitrary number of covariate categories d_R. With more than two possibilities, I show how additional nonparametric restrictions on the dimension of the unobserved heterogeneity restore point identification in the linear system of random coefficients equations. This complements the nonidentification result of Masten (2017). The second contribution is to provide a nonparametric estimator which is optimal in the minimax sense and adaptive. Additional group level variables $\boldsymbol{Z}$ can be observed, such as the precincts compositions in terms of levels of income or education. This allows to assume NCE conditional on these variables, *i.e.*, $\boldsymbol{B} \perp \boldsymbol{X}|\boldsymbol{Z}$, and to perform inference on

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \mapsto \mathbb{E}\left[\boldsymbol{B}|(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\right]. \tag{1.14}$$

My method is the first to incorporate the additional variables $\boldsymbol{Z}$ nonparametrically. Another extension relaxes the NCE assumption when the researcher observes an instrument, adapting the control function approach of Masten and Torgovitsky (2013).

I conclude with an application whose methodology paves the way for others in Political Science and Economics. It shows that a direct application of my estimator provides an estimator of the effect of a treatment $T_g \in \{0, 1\}$ on the choice probabilities among different categories, in a clustered experiment by group g where the individual outcome C and covariate R are protected. Consider the potential outcome model of Rubin (1974), where we denote by $\boldsymbol{B}_g(0)$ the potential outcome for group g if not treated, $\boldsymbol{B}_g(1)$ the potential outcome if treated, and we observe only the treatment status $T_g \in \{0, 1\}$ and relalized outcome $\boldsymbol{B}_g(T_g)$. When the groups g are treated $T_g = 1$ randomly conditionally on the variables $\boldsymbol{W}_g$, *i.e.*, the *unconfounded-*

ness assumption at the group level g,

$$\boldsymbol{B}_g(1), \boldsymbol{B}_g(0) \perp T_g \quad | \boldsymbol{W}_g,$$

the treatment effect $\gamma_{r,1} = \mathbb{E}\left[\boldsymbol{B}_{r,1}(1) - \boldsymbol{B}_{r,1}(0)\right]$ on the probability to choose $C = 1$ among individuals of category $R = r$, namely $\boldsymbol{B}_{r,1}$, is

$$\gamma_{r,1} = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{B}_{r,1} | \boldsymbol{W}, T = 1\right]\right] - \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{B}_{r,1} | \boldsymbol{W}, T = 0\right]\right].$$

Thus, using the law of iterated expectations, $\gamma_{r,1}$ is related to $\boldsymbol{m}$ in (1.14), treating T and $\boldsymbol{W}$ as additional variables $\boldsymbol{Z} := (\boldsymbol{W}, T)$, via

$$\gamma_{r,1} = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{m}_{r,1}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, T) | \boldsymbol{W}, T = 1\right]\right] - \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{m}_{r,1}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, T) | \boldsymbol{W}, T = 0\right]\right].$$

Specifically, I consider the effect of door-to-door visits on vote shares among different categories of voters, based on past votes, in the experiment of Pons (2018). My results suggest that canvassing is especially effective through persuasion of undecided voters, rather than mobilization of already convinced ones.

1.4 Tests of rational expectations using data combination

1.4.1 Motivation

In a dynamic environment, individual economic agents base their decisions on observables but also on their beliefs about the future values of some other variables. The rational expectations (RE) assumption states that agents have expectations that do not systematically differ from the realized outcomes, and efficiently process all private information to form these expectations (see, *e.g.*, Muth, 1961). Specifically, this specification of the beliefs formation process allows to estimate dynamic microeconomic models without collecting subjective expectations data. Thus, rational expectations are a key building block in structural models despite longstanding critiques (see, *e.g.*, Pesaran, 1987; Manski, 2004). For example, standard life cycle models use the assumption that consumers have rational expectations about their future individual earnings, but sometimes also about their longevity (see, *e.g.*, Meghir and Pistaferri, 2011). In another context, standard job search models also assume that jobseekers have rational expectations about both the arrival rates of offers and the wage distribution of the latter (see, *e.g.*, McCall, 1970; Van den Berg, 2001).

If we observe realized outcomes and subjective expectations for the same individuals, we can easily test for RE. This test has already been conducted in several contexts such as Lovell (1986); Gourieroux and Pradel (1986); Dominitz (1998); Gennaioli et al. (2015). However, it is a common situation in practice to observe individual realizations and subjective beliefs in two different datasets that cannot be matched (see, *e.g.*, Delavande, 2008; Arcidiacono, Hotz and Kang, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; Stinebrickner and Stinebrickner, 2014*a*; Gennaioli, Ma and Shleifer, 2016; Kuchler and Zafar, 2019; Boneva and Rauh, 2018; Biroli, Boneva, Raja and Rauh, 2020). A striking example in the US are earnings, where for the Survey of Consumer Expectations of the New York Fed, expectations and realizations can typically only be matched for a subset of the respondents. For consumption, there is yet no large scale survey in the US which collects matched expectations and realizations. It is thus important to be able to test RE in this type of data combination context.

Let us formalize the problem. Denote by Y the individual outcome and by $\psi = \mathcal{E}\left[Y|\mathcal{I}\right]$ its subjective expectation, where $\mathcal{I}$ denotes the σ-algebra corresponding to the agent's information set and using the subjective expectation operator $\mathcal{E}\left[\cdot|\mathcal{I}\right]$ (*i.e.*, for any random variable U, $\mathcal{E}\left[U|\mathcal{I}\right]$ is a $\mathcal{I}$-measurable random variable which can be understood as taking expectation over a subjective distribution of U conditioned on information in $\mathcal{I}$). Denote by $\mathbb{E}\left[Y|\mathcal{I}\right]$ the conditional expectation operator generated by the true data generating process.We are interested in testing the rational expectations (RE) hypothesis

$$(RE): \quad \psi = \mathbb{E}\left[Y|\mathcal{I}\right]. \tag{1.15}$$

. Note that, as described in more details in D'Haultfoeuille et al. (2020), this is compatible with heterogeneity in the information different agents use to form their expectations.

Let us describe the fundamental equality arising when testing RE. Under RE, using the law of iterated expectations and that ψ is a $\mathcal{I}$-measurable random variable, we have $\mathbb{E}[Y|\psi] = \mathbb{E}[\mathbb{E}[Y|\mathcal{I}]|\psi]$, hence using (1.15),

$$\mathbb{E}[Y|\psi] = \psi. \tag{1.16}$$

Conversely, if (1.16) holds, then for $\mathcal{I} = \sigma(\psi)$, we have (1.15). Thus, testing RE without restrictions on the information set $\mathcal{I}$ boils down to testing (1.16).

1.4.2 Contribution

In the fourth chapter of this thesis we only consider the context where we observe Y in one dataset and ψ in another one. We can (point-) identify the marginal distributions but not their joint distribution: this is a data combination problem. Implications of (1.16) used in the literature are equality in mean $\mathbb{E}[\psi] = \mathbb{E}[Y]$ or the constraint on the variances $\mathbb{V}(Y) \geq \mathbb{V}(\psi)$. But there are (infinitely) many other implications of RE and we show in D'Haultfoeuille et al. (2020) how to use them to test RE.

More precisely, we consider the test of whether one can rationalize RE, which is the relevant hypothesis in this data combination context, namely

H_0 : there exists a pair of random variables (Y', ψ') and a sigma-algebra $\mathcal{I}'$ such that $\sigma(\psi') \subset \mathcal{I}',\ Y' \sim Y,\ \psi' \sim \psi$ and $\mathbb{E}[Y'|\mathcal{I}'] = \psi'$.

Note that whether we can rationalize RE does not mean that the true pair of random variables (Y, ψ) satisfies (1.16). However, rejecting H_0 implies that RE does not hold ($\mathbb{E}[Y|\mathcal{I}] \neq \psi$). The main insight of this chapter is that testing H_0 is equivalent to testing whether Y is a mean preserving spread of ψ, *i.e.*, if F_ψ dominates at the second order F_Y and $\mathbb{E}[Y] = \mathbb{E}[\psi]$. Using the terminology of Rothschild and Stiglitz

(1970), H_0 holds if and only if realizations Y are riskier than beliefs ψ. Then, we show that this can equivalently be rewritten as a system of infinitely many moment inequalities. This allows us to apply the instrumental functions approach of Andrews et al. (2017) to test for such inequalities. The extensions of the baseline test that we consider include accounting for the availability of additional covariates or aggregate shocks.

We apply our framework to test for rational expectations about future earnings using individual expectations from the Labor Market module of the Survey of Consumer Expectations. While a naive test of equality of means between earnings beliefs and realizations shows that earnings expectations are realistic in the sense of not being significantly biased, thus not rejecting the rational expectations hypothesis, our test does reject rational expectations at the 1% level. The results of our test also indicate that the RE hypothesis is more credible for certain subpopulations than others. A companion R package RationalExp and the user's guide D'Haultfœuille et al. (2018a) ease the implementation of our test in other contexts.

1.5 Stable out-of-band extrapolation and analytic continuation

1.5.1 Motivation

The truncated Fourier operator $\mathcal{F}_c$ described in Section 1.5 acting on $L^2(\cosh(b\cdot))$ appears in at least two inverse problems which have applications in economics.[4]

The first problem is *out-of-band extrapolation*, which consists in recovering a function from the observation with error of its Fourier transform truncated to $[-c, c]$ (see, *e.g.*, Chapter 11.5 in Bertero and Boccacci, 1998). We observe the function which is

[4]We recall that it is the operator which to $f \in L^2(\cosh(b\cdot))$, the L^2 space equipped with $\langle f, g\rangle_{L^2(\cosh(b\cdot))} = \int_{\mathbb{R}} f(x)\overline{g}(x)\cosh(bx)dx$, associates $\mathcal{F}[f](c\cdot) \in L^2(-1, 1)$.

defined by

$$g_\delta(x) = \mathcal{F}_c[f](x) + \delta\xi(x), \quad \text{for } a.e.\ x \in (-1,1),\ f \in L^2(\cosh(b\cdot)), \tag{1.17}$$

where $\xi \in L^2(-1,1)$ is a deterministic noise, $\|\xi\|_{L^2(-1,1)} \le 1$ and $\delta > 0$ is a noise level. We consider approximating f on $L^2(\mathbb{R})$ from g_δ on $(-c,c)$. Specifically, this can be viewed as a simplified version of the problem appearing in the estimation of the density of the random coefficients in the linear random coefficients model of Section 1.5 when $p = 1$.

The second problem is *stable analytic continuation*, which consists in extrapolating an analytic square integrable function f from its observation with error on $[-c,c]$. This has a wide range of applications in economics, such as analytical continuation of an outcome's density (see also references in Gaillac and Gautier, 2019*a*). We observe the function on $(-c,c)$, for $c > 0$, which is defined by

$$f_\delta(cx) = f(cx) + \delta\xi(x), \quad \text{for } a.e.\ x \in (-1,1), \quad \mathcal{F}[f] \in L^2(\cosh(b\cdot)), \tag{1.18}$$

where $\xi \in L^2(-1,1)$ is a deterministic noise, $\|\xi\|_{L^2(-1,1)} \le 1$, $\delta > 0$ is a noise level.[5] We consider approximating $f_0 = f$ on $L^2(\mathbb{R})$ from f_δ on $(-c,c)$. This problem can be analyzed using the formula, for a.e. $x \in (-1,1)$,

$$f(cx) = \frac{1}{2\pi}\mathcal{F}_c\left[\mathcal{F}\left[f(-\cdot)\right]\right](x). \tag{1.19}$$

This decomposes the extrapolation problem as the composition of two operators: first the operator $\mathcal{F}_c$, then the Fourier transform.

The formulation with (1.18) allows to consider nonbandlimited functions, *i.e.* which have a Fourier transform which does not have a compact support, which is a more convincing set-up for probability densities for example. (1.18) is a more usual problem when, instead of $L^2(\cosh(b\cdot))$, we consider functions having their Fourier transform supported in $[-1/b, 1/b]$. Then, the singular values of the associated op-

[5] We also consider a shift $x_0 \in \mathbb{R}$ of the interval $[-c,c]$.

erator $\mathcal{F}_c$ are the Prolate Spheroidal Wave functions (PSWF). Their properties are well studied (see, *e.g.*, Landau and Pollak, 1962; Osipov et al., 2013). However, such an approach is prone to criticism when the researcher does not have *a priori* information on the bandlimits $[-1/b, 1/b]$ or when she questions the assumption that f is bandlimited.

The decomposition (1.19) suggests a two-step regularized inverse using the singular value decomposition (SVD) of $\mathcal{F}_c$ to recover $\mathcal{F}[f(-\cdot)]$ and then taking Fourier inverse to recover f. This underlines that this SVD plays a key role here and *a fortiori* in the out-of-band extrapolation problem (1.17). However, its properties remained unknown, except from those established in Widom (1964), *i.e.*, a logarithmic equivalent for the decay of the singular values to zero and commutation between a symmetric integral operator $\mathcal{Q}_{c/b}$ obtained by applying the truncated Fourier operator $\mathcal{F}_c$ to its adjoint and a second order differential operator.

1.5.2 Contribution

We provide nonasymptotic upper and lower bounds on the singular values with similar qualitative behavior in m (the index), b, and c. The lower bounds are used to obtain rates of convergence for stable analytic continuation of possibly nonbandlimited functions whose Fourier transform belongs to $L^2(\cosh(b\cdot))$. They are also used to establish the rates of convergence of the estimator in Gaillac and Gautier (2019*c*) when the slopes of the random coefficients model do not have compact support. Upper bounds on the singular values are also used in Gaillac and Gautier (2019*c*) to establish lower bounds on the minimax rate of convergence in the latter context. We also derive bounds on the sup-norm of the singular functions which are used to establish the asymptotic results for the adaptive estimator in Gaillac and Gautier (2019*c*). Finally, we propose a numerical method to compute the SVD and apply it to stable analytic continuation. The computation of the SVD relies on the differential operator $\mathcal{L}[\psi] = -(p\psi')' + q\psi$ with $p(\cdot) = \cosh(4c) - \cosh(4c\cdot)$, $q(\cdot) = 3c^2\cosh(4c\cdot)$, and domain $\mathcal{D} \subset \{\psi \in L^2(-1,1) : \ \mathcal{L}[\psi] \in L^2(-1,1)\}$ with boundary conditions of continuity at ± 1. The latter commutes with $\mathcal{Q}_{c/b}$, hence has the same eigenfunctions (see, *e.g.*, Morrison, 1962; Widom, 1964). Working with the differential operator is useful because its eigenvalues increase quadratically while those of the integral operator decrease exponentially, which creates numerical difficulties in the computations. Thus, the computation of these eigenfunctions is performed solving the eigenproblem for $\mathcal{L}[\psi]$, called a Sturm-Liouville problem,

$$-(p\psi')' + (q - \lambda)\psi = 0$$

with $\lambda \in \mathbb{C}$, which is *singular* because p^{-1} is infinite at the endpoints, hence requiring specific numerical algorithms (see, *e.g.*, Ledoux, 2007).

Chapter 2

Variation of Observed and Unobserved Heterogeneity and Identification in Some Random Coefficients Models

Joint with Eric Gautier, Toulouse School of Economics.

Abstract

This paper studies point identification of the distribution of the coefficients in some random coefficients models with exogenous regressors when their support is a proper subset, possibly discrete but countable. We exhibit various trade-offs between restrictions involving (generalized) moments of the random coefficients and the support of the regressors. These are obtained by considering functions of the data which characterize the distribution of the random coefficients over a set of points. A simple instance involves analytic functions and their sets of uniqueness. We consider linear models which include deconvolution, the binary choice model, and panel data models such as single-index panel data models and an extension of the Kotlarski lemma.

Keywords: Identification, Random Coefficients, Quasi-analyticity, Deconvolution.

2.1 Introduction

The determinants of the preferences of the economic agents remain largely unobserved to the econometrician. This unobserved heterogeneity is a central topic in economics as it impacts the redistributive effects of public policies. Estimating this distributional impact requires to identify the whole distribution of the unobserved heterogeneity. An important and growing literature thus aims at incorporating, identifying, and estimating nonparametrically the unobserved heterogeneity into general classes of economic models (see, *e.g.*, Berry and Haile, 2009; Gautier and Kitamura, 2013; Fox, 2017; Hoderlein et al., 2017; Masten, 2017; Chernozhukov et al., 2019; Cooprider et al., 2020). Imposing parametric assumptions on the law of the random coefficients is a widely used approach in the analysis of random coefficients models (see, *e.g.*, Berry et al., 1995; Train, 2009) but can seriously drive the results (see Heckman and Singer, 1984; Breunig, 2019). Moreover, economic theory rarely motivates these restrictions of the unobserved heterogeneity. For this reason, this paper considers nonparametric identification of its distribution. In applications, regressors may only have limited variation. Thus, we aim at showing how to identify unobserved heterogeneity nonparametrically with regressors having limited support, possibly discrete.

Consider the example of the linear random coefficients model, where the coefficients - intercept and slopes - are random with a law in a nonparametric class and independent from the regressors. Here, for each point of support of the regressors, we observe the distribution of the outcome conditional on the regressors. Using the linearity of the equation and assuming independence between the regressors and the random coefficients, the data consists in a set of 1-dimensional projections of the unobserved distribution. Thus, there is a trade-off between the variation the regressors, which yields different projections, and the size of the class of distributions we consider. We explicit the link between the nonparametric classes of laws of the random coefficients we use and restrictions which are implied on their support, their sign, or their tails. We show that there exist similar trade-offs for other models with multiple sources of unobserved heterogeneity, such as the random coefficients binary model and panel data models.

The first contribution of this paper is to provide identification of several random coefficients models using moment conditions on the distribution with regressors having limited variation. We show that the support of the latter can possibly be discrete but countable and we describe the trade-off with the integrability assumptions on the coefficients distribution. We study the linear random coefficients model, which is identified in the literature assuming that the regressors have a support larger than the support of the law of the random coefficients (see, *e.g.*, Beran and Hall, 1992; Beran et al., 1996*a*; Hoderlein et al., 2010). An important innovation is that we obtain identification using moments conditions on the marginals of the distribution of the unobservables except from the random intercept, which yield the so-called *quasi-analyticity* of the Fourier transform of the random coefficients. Thus, we remove entirely the assumptions regarding the marginal distribution of the random intercept made in Masten (2017) and let the dependence structure between the random intercept and the random slope unrestricted (see Gaillac and Gautier, 2019*c*, for a data-driven estimation). We also study the random coefficients binary model (see, *e.g.*, Ichimura and Thompson, 1998; Gautier and Kitamura, 2013; Gautier and Le Pennec, 2018) and the linear and single-index panel data models, where the whole distribution of unobserved heterogeneity can also be identified using regressors with limited support, even discrete if we assume smoothness and integrability assumptions on the joint distribution.

The second contribution is our improvements made on deconvolution problems where characteristic functions can vanish (see Carrasco and Florens, 2011) and on the well-known Kotlarski lemma (see Kotlarski, 1967; Evdokimov and White, 2012). The third contribution of the paper is to provide identification of a class of nonlinear random coefficients models, using their decompositions on well-chosen families of functions under the restriction that only a finite number the coefficients are random. The motivations for this model are to recover the whole distribution of the nonparametric elasticities or marginal effects.

The paper is organized as follows. Section 2.2 introduces a motivating example then settles the nonparametric identification principles and the useful tools. In Section 2.3, we provide the main identification results: on linear random coefficients models, including deconvolution with identifiable error, on the random coefficients binary model, and on some panel data models with random coefficients (*i.e.*, an extension of the Kotlarski lemma, the single index panel data model and the linear panel data models where regressors are monomials of a baseline scalar regressor). Section 2.4 contains the proofs. The supplemental appendix provides more details about the main tools used for identification, namely sets of uniqueness of homogeneous polynomials, analytic, and quasi-analytic classes, and Fourier transform of measures in sections 2.5.2 and 2.5.3.

Notations. Bold letters are used for vectors, capital letters for indeterminates of polynomials or random variables/vectors. For a real number r, $\boldsymbol{r}_c$ is the vector, the dimension of which will be clear from the text, where each entry is r. The notation $\mathbb{N}$ and $\mathbb{N}_0$ are used for the positive and nonnegative integers, $\mathbb{R}_+ = \{x \in \mathbb{R} : \ x > 0\}$, $\mathbb{C}(X)$ denotes the set of rational functions, and $\mathbb{1}\{\cdot\}$ the indicator function. $\mathbb{K}[X_1, \dots, X_p]$ is the ring of polynomials of p variables with coefficients in the ring $\mathbb{K}$. A^B is the set of applications from B to A. The notation $|\cdot|_q$ for $q \in [1, \infty]$ stands for the ℓ_q norm of a vector with components in $\mathbb{C}$ while $L^q(\mathcal{S}, \mu)$ for $q \in [1, \infty]$ are the q integrable functions on $\mathcal{S}$ with respect to the measure μ and the norms $\|\cdot\|_{L^q(\mathcal{S},\mu)}$. When μ is the Lebesgue measure we drop it from the notation. $\mathbb{S}^p$ is the unit sphere in $\mathbb{R}^{p+1}$. $(\boldsymbol{e}(j))_{j=1}^p$ is the canonical basis of $\mathbb{R}^p$.

For $\boldsymbol{\beta} \in \mathbb{C}^p$, $(f_m)_{m\in\mathbb{N}}$ functions with values in $\mathbb{C}$, $k \in \mathbb{N}_0$, and $\boldsymbol{m} \in \mathbb{N}_0^p$, denote by $\boldsymbol{m}! = \prod_{j=1}^p \boldsymbol{m}_j!$, $|\boldsymbol{m}| = \sum_{j=1}^p \boldsymbol{m}_j$, $\boldsymbol{\beta}^{\boldsymbol{m}} = \prod_{j=1}^p \boldsymbol{\beta}_j^{\boldsymbol{m}_j}$, $\boldsymbol{\beta}^k = \left(\boldsymbol{\beta}_1^k, \dots, \boldsymbol{\beta}_p^k\right)^\top$, $|\boldsymbol{\beta}|^{\boldsymbol{m}} = \prod_{j=1}^p |\boldsymbol{\beta}_j|^{\boldsymbol{m}_j}$, and $f_{\boldsymbol{m}} = \prod_{j=1}^p f_{\boldsymbol{m}_j}$. For a differentiable function f of real variables, $f^{(\boldsymbol{m})}$ denotes $\prod_{j=1}^p \frac{\partial^{\boldsymbol{m}_j}}{\partial \boldsymbol{x}_j^{\boldsymbol{m}_j}} f$ and $\mathrm{supp}(f)$ its support. Δ denotes the Laplacian. $\{M_m\}$, $\{0_m\}$ and $\{1_m\}$ are respectively $(M_m)_{m\in\mathbb{N}_0}$ and the sequences identically 0 and 1. We write $\{\boldsymbol{M}_m\}$ when $\boldsymbol{M}_m = (\boldsymbol{M}_{j,m})_{j=1}^p$. $\mathfrak{M}_c(\mathcal{S})$, $\mathfrak{M}(\mathcal{S})$, and $\mathfrak{M}_1(\mathcal{S})$ are the sets of complex, nonnegative, and probability measures on a Borel measurable set $\mathcal{S}$. For $\mu \in \mathfrak{M}_c(\mathbb{R}^p)$ and $\boldsymbol{m} \in \mathbb{N}_0^p$, $|\mu|$ is its *total variation*, $s_\mu(\boldsymbol{m}) := \int_{\mathbb{R}^p} \boldsymbol{x}^{\boldsymbol{m}} d\mu(\boldsymbol{x})$ the moments, $s_{|\cdot|,\mu}(\boldsymbol{m}) := \int_{\mathbb{R}^p} |\boldsymbol{x}^{\boldsymbol{m}}| d\mu(\boldsymbol{x})$ the *absolute moments*.

The Fourier transform of $\mu \in \mathfrak{M}_c(\mathbb{R}^d)$ (resp. f in $L^q(\mathbb{R}^d)$ for $q = 1, 2$) is $\mathcal{F}[\mu] : \boldsymbol{x} \mapsto \int_{\mathbb{R}^d} e^{i\boldsymbol{b}^\top \boldsymbol{x}} d\mu(\boldsymbol{b})$ (resp. $\mathcal{F}[f]$). For a random vector $\boldsymbol{X}$, $\mathbb{P}_{\boldsymbol{X}}$ is its law, $F_{\boldsymbol{X}}$ its CDF, $f_{\boldsymbol{X}}$ its density, $\varphi_{\boldsymbol{X}} := \mathcal{F}[\mathbb{P}_{\boldsymbol{X}}]$, and $\mathbb{S}_{\boldsymbol{X}}$ its support. $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{X})]$ is the expectation of $f(\boldsymbol{X})$ under $\mathbb{P}$, $\otimes$ the product of measure, $\mathbb{P}_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$, and $\varphi_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x}) := \mathcal{F}\left[\mathbb{P}_{Y|\boldsymbol{X}=\boldsymbol{x}}\right](\cdot)$. We use the symbol $\perp$ to denote independence.

2.2 Tools for nonparametric identification of the distribution of the unobserved heterogeneity

2.2.1 An illustrating example : the linear random coefficients model

Consider the linear random coefficients model

$$Y = \boldsymbol{\gamma}^\top \boldsymbol{f}(X), \tag{2.1}$$

$$\boldsymbol{\gamma} \text{ and } X \text{ are independent,} \tag{2.2}$$

where the random coefficients $\boldsymbol{\gamma}$ belong to $\mathbb{R}^{p+1}$, X is scalar, and $\boldsymbol{f}(X)$ is a vector of transformations of X, $\boldsymbol{f}(X) := (1, \boldsymbol{f}_2(X), \dots, \boldsymbol{f}_{p+1}(X))$. In this model, we are interested in describing the heterogeneity when $\boldsymbol{f}(X) = (1, X)$, or in recovering the whole distribution of the elasticity with respect to X, $\boldsymbol{\gamma}^\top(\boldsymbol{f}'(x))x/\boldsymbol{\gamma}^\top \boldsymbol{f}(x)$, for $x \in \mathbb{S}_X$, or the marginal effects $\boldsymbol{\gamma}^\top(\boldsymbol{f}'(x))$, assuming that both quantities exist. We give here a sense of how identification depends on the transformation $\boldsymbol{f}$, the support of the regressor X, and the moments of $\mathbb{P}_{\boldsymbol{\gamma}}$.

Starting from (2.1)-(2.2), we have

$$\left(\mathbb{P}_{Y|X=x}\right)_{x\in\mathbb{S}_X} = \left(\mathbb{P}_{\boldsymbol{\gamma}^\top \boldsymbol{f}(x)}\right)_{x\in\mathbb{S}_X},$$

which are 1-dimensional projections of the distribution of $\boldsymbol{\gamma}$ indexed by $x \in \mathbb{S}_X$. Thus, we want to characterize the distribution $\mathbb{P}_{\boldsymbol{\gamma}}$ from the knowledge of these 1-dimensional projections. For all $x \in \mathbb{S}_X$, the distributions of the projections $\mathbb{P}_{\boldsymbol{\gamma}^\top \boldsymbol{f}(x)}$ are characterized by their characteristic functions, which are related to the Fourier transform of the unobserved distribution $\mathbb{P}_{\boldsymbol{\gamma}}$, for all $t \in \mathbb{R}$ and $x \in \mathbb{S}_X$,

$$\begin{aligned}
\mathcal{F}\left[\mathbb{P}_{Y|X=x}\right](t) &= \mathbb{E}\left[e^{itY}\middle|X=x\right] \\
&= \mathbb{E}\left[e^{it\boldsymbol{\gamma}^\top \boldsymbol{f}(x)}\right] \quad (\text{using } (2.1)-(2.2)) \\
&= \int_{\mathbb{R}^{p+1}} e^{it\boldsymbol{g}^\top \boldsymbol{f}(x)} d\mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{g}) \\
&= \mathcal{F}\left[\mathbb{P}_{\boldsymbol{\gamma}}\right](t\boldsymbol{f}(x)). \qquad (2.3)
\end{aligned}$$

Specifically, assuming that they exist, the derivatives of the characteristic function of $\mathbb{P}_{Y|X=x}$ at 0 are identified,[1] for all $k \in \mathbb{N}_0$ and $x \in \mathbb{S}_{\boldsymbol{X}}$,

$$\begin{aligned}
(-i)^k \mathcal{F}^{(k)}\left[\mathbb{P}_{Y|X=x}\right](t)\Big|_{t=0} &= (-i)^k \partial_t^{(k)} \mathcal{F}\left[\mathbb{P}_{\boldsymbol{\gamma}}\right](t\boldsymbol{f}(x))\Big|_{t=0} \\
&= \int_{\mathbb{R}^{p+1}} \left(\sum_{j=1}^{p+1} \boldsymbol{g}_j \boldsymbol{f}_j(x)\right)^k d\mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{g}), \\
&= \sum_{\boldsymbol{j}\in\mathbb{N}_0^{p+1}:\ |\boldsymbol{j}|_1=k} \begin{pmatrix} k \\ \boldsymbol{j}_1,\dots,\boldsymbol{j}_{p+1} \end{pmatrix} c_{\boldsymbol{j}} \boldsymbol{f}(x)^{\boldsymbol{j}} = P(\boldsymbol{f}(x)),
\end{aligned}$$

where $c_{\boldsymbol{j}} := \int_{\mathbb{R}^{p+1}} \boldsymbol{g}^{\boldsymbol{j}} d\mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{g})$ and

$$P(Z_1,\dots,Z_{p+1}) = \sum_{\boldsymbol{j}\in\mathbb{N}_0^{p+1}:\ |\boldsymbol{j}|_1=k} \begin{pmatrix} k \\ \boldsymbol{j}_1,\dots,\boldsymbol{j}_{p+1} \end{pmatrix} c_{\boldsymbol{j}}\ Z_1^{\boldsymbol{j}_1}\dots Z_{p+1}^{\boldsymbol{j}_{p+1}}. \qquad (2.4)$$

[1] They are also the moments of the projections $\mathbb{P}_{Y|X=x}$.

They are evaluations of homogeneous polynomials at $\boldsymbol{f}(x)$ whose coefficients are, up to a positive multiplicative constant, the moments of $\mathbb{P}_{\boldsymbol{\gamma}}$.

Several remarks are in order. First, if $\mathbb{P}_{\boldsymbol{\gamma}}$ is characterized by its moments, it is interesting to ask for conditions on the transformation $\boldsymbol{f}$ and the variation of the regressor X such that we are able to recover all the moments of $\mathbb{P}_{\boldsymbol{\gamma}}$. Indeed, for some $\boldsymbol{f}$, even if the support of X is $\mathbb{R}$, the distribution $\mathbb{P}_{\boldsymbol{\gamma}}$ in some models of the form (2.1)-(2.2) is not identified. To show this, consider model (2.1)-(2.2) and assume that there exists a homogeneous polynomial Q of degree $k > 0$, such that $Q(\boldsymbol{f}(x)) = 0$ for all $x \in \mathbb{S}_X$. Consider a compactly supported probability distribution $\mathbb{P}^*_{\boldsymbol{\gamma}}$ on $[-1,1]^{p+1}$ bounded by below by $c > 0$ and define the function $q(\boldsymbol{g}) = \prod_{j=1}^{p+1} \boldsymbol{g}_j \exp(-1/(\boldsymbol{g}_j^2 - 1))\mathbb{1}\{\boldsymbol{g}_j \in [-1,1]\}$ which is infinitely differentiable on $\mathbb{R}^{p+1}$. Then, $\mathbb{P}_{\boldsymbol{\gamma}} := \mathbb{P}^*_{\boldsymbol{\gamma}} + ch/\|h\|_\infty d\boldsymbol{g}$, where $h = Q(\partial_1, \dots, \partial_{p+1})q$ is a probability. Moreover, using the properties of the Fourier transform and that Q is homogeneous, $\mathbb{P}_{\boldsymbol{\gamma}}$ satisfies, for all $x \in \mathbb{S}_X$ and $t \in \mathbb{R}$,

$$\mathcal{F}\left[\mathbb{P}_{\boldsymbol{\gamma}}\right](t\boldsymbol{f}(x)) = \mathcal{F}\left[\mathbb{P}^*_{\boldsymbol{\gamma}}\right](t\boldsymbol{f}(x)) + t^k Q(\boldsymbol{f}(x))\mathcal{F}\left[h\right](t\boldsymbol{f}(x)). \tag{2.5}$$

Because $\{\boldsymbol{f}(x),\ x \in \mathbb{R}\}$ is a subset of the set of zeros of Q, we obtain from (2.5) that $\mathbb{P}^*_{\boldsymbol{\gamma}}$ and $\mathbb{P}_{\boldsymbol{\gamma}}$ are distinct measures leading to the same observables. A simple particular case is

$$Y = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 X + \boldsymbol{\gamma}_3 X^2,$$

where the distribution of $\boldsymbol{\gamma}$ is not nonparametrically identified even if the support of X is $\mathbb{R}$, as $\{(1, x, x^2),\ x \in \mathbb{R}\}$ is a subset of the set of zeros of $Q(Z_1, Z_2, Z_3) = Z_2^2 - Z_1 Z_3$. Building on these insights, this paper provides conditions on $\boldsymbol{f}$ to obtain linear random coefficients models of type (2.1)-(2.2) which are identified (see Section 2.3.1).

Second, assuming that $\mathbb{P}_{\boldsymbol{\gamma}}$ is characterized by its moments restricts the moments of the intercept $\boldsymbol{\gamma}_1$, which is sometimes undesirable. Consider the case $Y = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 X$ where, to relax assumptions on $\boldsymbol{\gamma}_1$, one can notice from (2.3) that the identification of $\mathbb{P}_{\boldsymbol{\gamma}}$ amounts to recovering it from the knowledge of its Fourier transform on the cone

$$\{t(1,x), t \in \mathbb{R}, x \in \mathbb{S}_X\}.$$

For all $t \neq 0$, using $\mu_t(\cdot) := \int_{\mathbb{R}} e^{it\boldsymbol{g}_1} d\mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{g}_1, \cdot)$, we have, for all $x \in \mathbb{S}_X$,

$$\begin{aligned} \mathcal{F}\left[\mathbb{P}_{\boldsymbol{\gamma}}\right](t, tx) &= \int_{\mathbb{R}} e^{it\boldsymbol{g}_2 x} d\mu_t(\boldsymbol{g}_2) \\ &= \mathcal{F}\left[\mu_t\right](tx). \end{aligned} \tag{2.6}$$

We consider moment conditions on $\boldsymbol{\gamma}_2$ implying that, for all $t \neq 0$, the characteristic function of the complex measure μ_t belongs to (quasi-)analytic classes. Assumption $R_{L,4}$ (i) yields that $\{tx, x \in \mathbb{S}_X\}$ is a set of uniqueness of such classes. Using (2.6), this yields that, for all $t \neq 0$, the partial Fourier transform with respect to the first variable $\mu_t = \mathcal{F}_{1\text{st}}\left[\mathbb{P}^*_{\boldsymbol{\gamma}}\right](t, \cdot)$ is identified hence, using the continuity of the Fourier transform at 0, that $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified. This allows to obtain identification without assumptions on $\boldsymbol{\gamma}_1$ with X having possibly discrete support.

2.2.2 Nonparametric identification principles

This paper uses basic elements of nonparametric identification which we introduce (see Matzkin, 2007*a*). Consider models defined by a function v (possibly a vector of functions) relating vectors of outcomes $\boldsymbol{Y} \in \mathcal{Y}$ to a vector of unobserved and observed factors, denoted respectively by $\boldsymbol{\gamma} \in \Gamma$ and $\boldsymbol{X} \in \mathcal{X}$, where $\Gamma = \mathbb{R}^{p+1}$ or $\Gamma = \mathbb{S}^p$ and $\mathcal{X} = \mathbb{R}^p$ or $\mathcal{X} = \mathbb{S}^p$. The sets $\mathcal{Y}$, Γ, and $\mathcal{X}$ are equipped with the Borel σ-field. The equation

$$v(\boldsymbol{Y}, \boldsymbol{\gamma}, \boldsymbol{X}) = 0, \tag{2.7}$$

defines the collection $\left(\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right)\right)_{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}}$ of laws of observables $\mathbb{P}_{\boldsymbol{Y},\boldsymbol{X}}$ generated by $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}$, where $\mathcal{R}$ accounts for the model restrictions. We consider the following base restriction.

Restriction (Exogeneity assumption)**.** *Define*

$$\mathcal{R}_1 = \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} :\ \mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} = \mathbb{P}_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{X}}\}.$$

Everything in this paper holds if we replace independence by independence given $\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a random vector from which we could have observations simultaneously with those of $\boldsymbol{X}$ and the outcomes or which could be identifiable from a model for $\boldsymbol{X}$ obtained by a control function approach. When $\mathcal{R} \subseteq \mathcal{R}_1$, we denote by $\mathbb{P}^*_{\boldsymbol{\gamma}}$ the true parameter.

Definition 1 (Nonparametric identification of $\mathbb{P}^*_{\boldsymbol{\gamma}}$)**.** *The distribution $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under the restriction $\mathcal{R} \subseteq \mathcal{R}_1$ if for all distributions $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}$ such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$ then $\mathbb{P}^*_{\boldsymbol{\gamma}} = \mathbb{P}_{\boldsymbol{\gamma}}$.*

Finding such a set $\mathcal{R}$ amounts to finding sufficient conditions for identification. Showing that a condition is necessary is only relatively useful. Indeed, if for $\mathcal{R} \subseteq \mathcal{R}''$, $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}$ but not $\mathcal{R}''$, it does not mean $\mathcal{R}$ is sharp as there could exist $\mathcal{R} \subseteq \mathcal{R}' \subseteq \mathcal{R}''$ such that $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}'$. In this paper, identification of $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is relative to $\mathcal{R} \subseteq \mathcal{R}_1$ which restricts the class of marginals $\mathbb{P}_{\boldsymbol{\gamma}}$ of $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}$ and imposes that $\mathbb{S}_{\boldsymbol{X}}$ is rich enough. To obtain identification without restricting the class of $\mathbb{P}_{\boldsymbol{\gamma}}$ usually requires that $\mathbb{S}_{\boldsymbol{X}} = \mathcal{X}$, which is too demanding for a dataset. Hence, this paper studies middle ground restrictions where $\mathbb{S}_{\boldsymbol{X}}$ could be a proper subset, possibly discrete, while $\mathbb{P}_{\boldsymbol{\gamma}}$ belongs to a restricted but nonparametric class. For models which involve an index $(1, \boldsymbol{X}^\top)\boldsymbol{\gamma}$, denoting by $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top)^\top$, we have, for all $\underline{\boldsymbol{x}} \in \mathbb{R}^p$ and $M \in GL(\mathbb{R}^p)$, where $GL(\mathbb{K}^p)$ is the group of invertible $p \times p$ matrices with coefficients in $\mathbb{K}$,

$$\alpha + \boldsymbol{\beta}^\top \boldsymbol{X} = \alpha + \boldsymbol{\beta}^\top \underline{\boldsymbol{x}} + \boldsymbol{\beta}^\top M^{-1} M \left(\boldsymbol{X} - \underline{\boldsymbol{x}}\right), \tag{2.8}$$

and, when $\underline{\boldsymbol{x}}$ and M are known, there is a one to one mapping between $\mathbb{P}_{\alpha+\boldsymbol{\beta}^\top \underline{\boldsymbol{x}},(M^{-1})^\top \boldsymbol{\beta}}$ and $\mathbb{P}_{\boldsymbol{\gamma}}$. Hence, we could, by reparametrization, study identification when $\mathbb{S}_{\boldsymbol{X}}$ is replaced by any convenient invertible affine transformation. Sometimes, the results in this paper also involve treating in a specific order the regressors but, based on this discussion, the order is irrelevant.

2.2.3 Quasi-analyticity, sets of uniqueness, and conditions on the moments

We introduce here the main concepts used to obtain identification. Let $\mathcal{S}$ be a subset of $\mathbb{C}$, $C(\mathcal{S})$ and $C^{\infty}(\mathcal{S})$ be the continuous and infinitely differentiable functions at every point in $\mathcal{S}$ with values in $\mathbb{C}$. Given $\mathcal{S} \subseteq \mathbb{R}$, $b \in \mathbb{R}_{+}$, and $\{M_m\} \in (0, \infty]^{\mathbb{N}_0}$, let us introduce useful classes of functions which have explicit controls on their derivatives,

$$C^{\{M_m\}}(\mathcal{S}, b) := \left\{ f \in C^{\infty}(\mathcal{S}) : \exists c : \forall m \in \mathbb{N}_0, \left\| f^{(m)} \right\|_{L^{\infty}(\mathcal{S})} \leq c b^{|m|} M_m \right\}. \quad (2.9)$$

Due to the constant c, we can assume that $M_0 = 1$ in (2.9) and $C^{\{M_m\}}(\mathcal{S}) := \bigcup_{b \in \mathbb{R}_{+}} C^{\{M_m\}}(\mathcal{S}, b)$. The analytic functions $\mathcal{A}(\mathcal{S}) \subseteq C^{\infty}(\mathcal{S})$ on $\mathcal{S} \subseteq \mathbb{R}^p$ or $\mathcal{S} \subseteq \mathbb{C}^p$ are the $C^{\infty}(\mathcal{S})$ functions with a convergent Taylor series around every point in $\mathcal{S}$. Note that the analytic functions on $\mathbb{R}$ correspond to the class $C^{\{m!\}}(\mathbb{R})$ (see, *e.g.*, Theorem 19.9 in Rudin, 1973). Because $\mathcal{A}(\mathcal{S})$ is a small subset of $C^{\infty}(\mathcal{S})$ (see Appendix 2.5.6), we use larger classes.

Definition 2 (**Quasi-analytic classes**). *Let $\mathcal{S} \subseteq \mathbb{R}$, then the class of functions $\mathcal{C}(\mathcal{S})$ is said to be quasi-analytic if there exists $x_0 \in \mathcal{S}$ such that, for all $f \in \mathcal{C}(\mathcal{S})$,*

$$f^{(m)}(x_0) = 0, \quad \forall m \in \mathbb{N}_0$$

implies that $f = 0$ on $\mathcal{S}$.

Thus, quasi-analytic classes are functions being characterized by the knowledge of their derivatives at a point. There exist quasi-analytic functions which are not analytic functions. Proposition 1 below or the Denjoy-Carleman Theorem (see Theorem 1.7 in Infusino, 2016) give necessary and sufficient conditions on $\{M_m\}$ for $C^{\{M_m\}}(\mathbb{R})$ to be a quasi-analytic class.

Our identification strategy uses sets of uniqueness for the classes of functions we have just introduced. We give several examples of such sets in Appendix 2.5.2.

Definition 3 (**Set of uniqueness**). *Let $\mathfrak{F}(\mathcal{S})$ be a vector space of functions on $\mathcal{S}$.*

$U \subseteq \mathcal{S}$ *is a* set of uniqueness *for* $\mathfrak{F}(\mathcal{S})$ *if every function of* $\mathfrak{F}(\mathcal{S})$ *which is zero on* U *is identically zero on* $\mathcal{S}$.

Determinate measures and how they relate to quasi-analytic classes is the last important concept. Define the classes of measures which admit finite absolute moments,

$$\mathfrak{M}_c^* \left(\mathbb{R}^p\right) := \left\{\mu \in \mathfrak{M}_c \left(\mathbb{R}^p\right) : \ \forall m \in \mathbb{N}, \ \sum_{j=1}^p s_{|\cdot|,|\mu|}(2m\boldsymbol{e}(j)) < \infty \right\};$$

$$\mathfrak{M}^* \left(\mathbb{R}^p\right) := \mathfrak{M}_c^* \left(\mathbb{R}^p\right) \cap \mathfrak{M} \left(\mathbb{R}^p\right), \ \mathfrak{M}_1^* \left(\mathbb{R}^p\right) := \mathfrak{M}^* \left(\mathbb{R}^p\right) \cap \mathfrak{M}_1 \left(\mathbb{R}^p\right).$$

Hereafter, $\mathcal{M}$ is $\mathfrak{M}_c^* \left(\mathbb{R}^p\right)$, $\mathfrak{M}^* \left(\mathbb{R}^p\right)$, or subsets with restriction on the support. Our identification strategies use determinate measures.

Definition 4 (**Determinate measures**)**.** *Let* $\mathcal{M} \subseteq \mathfrak{M}_c^*(\mathbb{R}^p)$*. A measure* μ *is determinate in* $\mathcal{M}$ *if, when* $\mu, \nu \in \mathcal{M}$ *are such that, for all* $\boldsymbol{m} \in \mathbb{N}_0^p$*,* $s_\mu(\boldsymbol{m}) = s_\nu(\boldsymbol{m})$*, then* $\mu = \nu$*.*

For a measure $\mu \in \mathfrak{M}^*(\mathbb{R})$, identification results in this paper assume moment restrictions of the type $s_\mu(m) \leq M_m$, where $\{M_m\}$ is a log-convex sequence, *i.e.*, a sequence of nonnegative numbers such that, for all $m \in \mathbb{N}$, $M_m^2 \leq M_{m-1}M_{m+1}$, which satisfies $M_0 = 1$.[2] More generally, Theorem 2.10 in Infusino (2016) obtained from Petersen (1982) ensures that the determinacy of a measure on $\mathbb{R}^p$ can be assessed by the determinacy of projections in a set of p directions spanning $\mathbb{R}^p$. Proposition 1 below gives a criterion on the bounds on the moments $s_\mu(m) \leq M_m$ of a measure μ when $p = 1$ ensuring its determinacy and that its Fourier transform belongs to a quasi-analytic class.

Proposition 1. *Let* $\{M_m\} \in (0, \infty]^{\mathbb{N}_0}$ *and be log-convex sequence.*

1. *Let* $\mu \in \mathfrak{M}_c^*(\mathbb{R})$*. If, for all* $m \in \mathbb{N}_0$*, the absolute moments of* $|\mu|$ *are bounded*

[2]For positive log-convex sequences, this equivalently means that $\{M_{m-1}/M_m\}$ is nonincreasing. Also, for such a sequence, if $\inf\{k \in \mathbb{N} : \ M_k = \infty\} < \infty$, then we have, for all $m \geq \min\{k \in \mathbb{N} : \ M_k = \infty\}$, $M_m = \infty$. We can relax this condition and rely on log-convex regularisation of the sequences, but we differ this point to the Appendix for simplicity.

$s_{|\cdot|,|\mu|}(m) \leq M_m$, and $\{M_m\}$ satisfies

$$\sum_{m\in\mathbb{N}} \frac{1}{M_{2m}^{1/(2m)}} = \infty, \tag{2.10}$$

then μ is determinate in $\mathfrak{M}_c^*(\mathbb{R})$ and its Fourier transform belongs to $C^{\{M_m\}}(\mathbb{R},1)$, which is a quasi-analytic class.

2. *Let $\mu \in \mathfrak{M}_c^*(\mathbb{R}_+)$. If, for all $m \in \mathbb{N}_0$, the absolute moments of $|\mu|$ are bounded $s_{|\cdot|,|\mu|}(m) \leq M_m$, and $\{M_m\}$ satisfies*

$$\sum_{m\in\mathbb{N}} \frac{1}{M_{m}^{1/(2m)}} = \infty, \tag{2.11}$$

then μ is determinate in $\mathfrak{M}_c^*(\mathbb{R}_+)$ and its Fourier transform belongs to $C^{\{M_m\}}(\mathbb{R},1)$, which is a quasi-analytic class.

The first part of the two statements of Proposition 1 is the Carleman theorem (see Theorem 10 in Appendix for other equivalent conditions or Theorem 2.5 in Infusino, 2016). Condition (2.10) is the so-called Carleman condition of the Hamburger moment problem and (2.11) is the weaker Stieltjes condition for the Stieltjes moment problem.

It is legitimate to ask what are in practice the measures satisfying (2.10), hence which are determinate in $\mathfrak{M}^*(\mathbb{R})$. To answer this question, there exist integral criteria for $\mu \in \mathfrak{M}_1^*(\mathbb{R})$ with a density f which are easier to check and interpret than (2.10). When f is positive and the *Krein condition* $\int_{\mathbb{R}} -\log(f(x))/(1+x^2)dx < \infty$ holds, μ is not determinate in $\mathfrak{M}_1^*(\mathbb{R})$ (see Lin, 1997); while, if f is also even, differentiable, and there exists $x_0 > 0$ such that, for all $x \geq x_0 > 0$ and $x \mapsto f(x)$ decreases to 0, $x \mapsto -xf'(x)/f(x)$ increases to infinity (so-called *Lin conditions*), and $\int_{\mathbb{R}} -\log(f(x))/(1+x^2)dx = \infty$ then (2.10) holds.

Examples. Based on these criteria, Student's t with $0 < \nu < \infty$ degrees of freedom, the generalized gamma $GG(a,b,p)$ for $0 < a < 1/2$ of density $ab^px^{ap-1}\exp(-bx^a)/\Gamma(p)$ on $\mathbb{R}^+$, any positive power of the lognormal, the law of N^{2n+1} for all $n \in \mathbb{N}$, $|N|^r$ for $r > 4$, X^m for all $m \in \mathbb{N} \setminus \{1,2\}$, Y^r for all $|r| > 2$, where N is a Gaussian, X a

Laplace, gamma or logistic, and Y an inverse Gaussian random variables, are not determinate respectively in the space of probabilities with appropriate assumptions on the support. However, $|N|^r$ for all $0 < r \leq 4$, X^m for $m = 1, 2$, Y^r for all $-2 \leq r \leq 2$, $GG(a, b, p)$ for all $a \geq 1/2$ (thus the χ^2 with any degrees of freedom) are determinate in the space of probabilities with the appropriate assumptions on the support (see Stoyanov, 2000; Pakes et al., 2001; Kleiber and Stoyanov, 2013, for more examples and references).

2.3 Main identification results

Theorems 1, 2, 3, 7, 9, and 11 use restrictions involving log-convex sequences, which is slightly less general than working with determinate measures, but it can be easily adapted.

2.3.1 The linear random coefficients model

Hereafter in this section, we consider the linear random coefficients model

$$Y = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}, \tag{2.12}$$

under $\mathcal{R}_1$ which particularizes (2.7) with $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top)^\top$.

Deconvolution

Deconvolution problems with two samples, one of the error and one of the sum of the signal and the error, can be viewed as a particular linear random coefficients model, where the regressor X has dimension 1 and two points of support:

$$Y = \alpha + \beta X,$$

under $\mathcal{R}_1$, which particularizes (2.7) with $\boldsymbol{\gamma} = (\alpha, \beta)^\top$.

Assumption 1. *Let* $\{M_m\}$ *be a log-convex sequence. Define the restriction*

$$\mathcal{R}_D(\{M_m\}) := \{\mathbb{P}_{\gamma,X} : \ \mathbb{P}_{\gamma,X} = \mathbb{P}_{\gamma} \otimes \mathbb{P}_X \ \textit{and} \ R_D \ \textit{(i)-(iii)}\},$$

R_D *(i)* $\{0,1\} \subseteq \mathbb{S}_X$*;*

R_D *(ii)* $\mathbb{P}_{\gamma} = \mathbb{P}_{\alpha} \otimes \mathbb{P}_{\beta}$ *and, for all* $m \in \mathbb{N}_0$, $\mathbb{E}[|\beta|^m] \le M_m$.

R_D *(iii) if* $\mathbb{S}_\beta \subseteq \mathbb{R}_+$, $\{M_m\}$ *satisfies* (2.10) *and else, if* $\mathbb{S}_\beta \not\subseteq \mathbb{R}_+$, $\{M_m\}$ *satisfies* (2.11).

Focusing on $\{0,1\}$ in R_D (i) is not restrictive as one can use affine rescaling of X. Because $\mathbb{P}_{\gamma} = \mathbb{P}_{\alpha} \otimes \mathbb{P}_{\beta}$, $\mathbb{P}_{Y|X}(\cdot|1)$ is the law of the sum of two independent random variables, where α plays the role of the noise in deconvolution problems, and $\mathbb{P}_{Y|X}(\cdot|0)$ is the law of the noise α. The following theorem is a particular case of Theorem 2, where (2.12) is generalized to higher dimensions of X and β. We show that in arbitrary dimensions, when $\mathbb{S}_X$ is richer but possibly discrete, *i.e.*, we can vary the intensity of the noise, the independence in condition R_D (ii) could be removed, and we can identify the distribution of γ without independence and assumption on the distribution of the noise. Else, when $\mathbb{S}_X = \{0,1\}$ and α and β are arbitrarily dependent, we can obtain sharp bounds on F_β (see Gautier and Hoderlein, 2015 and the references therein).

Theorem 1. $\mathbb{P}^*_{\gamma}$ *for* (2.12) *is identified under* $\mathcal{R}_D(\{M_m\})$*, where* $\{M_m\}$ *is a log-convex sequence satisfying* $M_0 = 1$.

Theorem 1 extends Theorem 2.1 in Beran and Hall (1992), where they assume that the interior of the support of $\boldsymbol{X}$ is nonempty. Estimation in the deconvolution model when, instead of R_D (ii)-(iii), the assumption that the interior of the sets of zeros of φ_α is empty is maintained has been studied in Carrasco and Florens (2011). By continuity of the Fourier transform of probabilities, the set of zeros of a characteristic function is a closed set and assuming that its interior is empty is equivalent to assuming it consists of isolated points. Under $\mathcal{R}_D(\{M_m\})$, φ_α can have zeros on arbitrary large sets at the expense of a stronger assumption on $\mathbb{P}_\beta$. Indeed, φ_β has isolated zeros (see Proposition 1 and Example QA.2 in the Appendix), else it would be identically 0. This would contradict the fact that its value at 0 is 1. There are classical examples of characteristic functions with compact support (*e.g.*, $t \mapsto (1-|t|^r)1\!\!1\{|t| \le 1\}$ for $0 < r \le 1$). The sufficient condition of Polya (see, *e.g.*, Theorem 6.5.3 in Chung, 2002) provides generic examples of laws which do not have absolute first moment. The idea to handle $\mathcal{R}_D(\{M_m\})$ is the same as in Remark 1 in Kotlarski (1967) but the maintained assumption is weaker because we allow for nonanalytic characteristic functions. Meister (2007) considers the estimation of compactly supported densities of a signal in the deconvolution model where the known distribution of the noise can have a characteristic function which set of zeros consists of isolated points. This is a more restrictive framework than Carrasco and Florens (2011) because the characteristic function is analytic (see (P3.3) in the Appendix).

Identification under independence of the marginals of $\mathbb{P}^*_{\boldsymbol{\gamma}}$

Due to $\mathcal{R}_1$, we have $\mathbb{P}_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x}) = \mathbb{P}_{\alpha+\boldsymbol{\beta}^\top\boldsymbol{x}}$. Hence, $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under

$$\mathcal{R}_{L,1} := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} : \ \mathbb{S}_{\boldsymbol{X}} = \mathbb{R}^p\}$$

by the Cramer-Wold Theorem. When the support of the regressors $\boldsymbol{X}$ is a proper subset of $\mathbb{R}^p$, $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified if we assume independence in the vector $\boldsymbol{\gamma}$ (see Beran and Hall, 1992; Gautier and Hoderlein, 2015). For example, independence of the marginals of $\mathbb{P}_{\boldsymbol{\gamma}}$ gives rise to a deconvolution problem and identification of $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is possible when the support of $\boldsymbol{X}$ contains only $p+1$ points.

Assumption 2. *Let, for all $k = 1, \dots, p$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences, $\boldsymbol{M}_{k,0} = 1$, and*

$$\mathcal{R}_{L,2}\left(\{\boldsymbol{M}_{k,m}\}_{k=1}^p\right) := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} : \ R_{L,2} \ (i)\text{-}(ii)\},$$

where

$R_{L,2}$ (*i*) $\{\boldsymbol{0}_c, \boldsymbol{x}(1), \dots, \boldsymbol{x}(p)\} \subseteq \mathbb{S}_{\boldsymbol{X}}$ *and, up to a permutation of the vectors,* $(\boldsymbol{x}(1), \dots, \boldsymbol{x}(p))$

is an upper triangular matrix with nonzero diagonal elements;

$R_{L,2}$ *(ii)* $\mathbb{P}_{\boldsymbol{\gamma}} = \mathbb{P}_{\boldsymbol{\alpha}} \otimes \bigotimes_{k=1}^{p} \mathbb{P}_{\boldsymbol{\beta}_k}$, *for all* $k = 1, \ldots, p$ *and* $m \in \mathbb{N}_0$, $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\beta}}}[|\boldsymbol{\beta}_k|^m] \leq \boldsymbol{M}_{k,m}$, *if* $\mathbb{S}_{\boldsymbol{\beta}_k} \subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ *satisfies* (2.11) *and else, if* $\mathbb{S}_{\boldsymbol{\beta}_k} \not\subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ *satisfies* (2.10).

The following Theorem 2 extends Theorem 1. We allow for different classes of determinate probabilities. For example, some coefficients could be assumed positive.

Theorem 2. $\mathbb{P}^*_{\boldsymbol{\gamma}}$ *is identified under* $\mathcal{R}_{L,2}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p})$, *where, for all* $k = 1, \ldots, p$, $\{\boldsymbol{M}_{k,m}\}$ *are log-convex sequences satisfying* $\boldsymbol{M}_{k,0} = 1$.

The restriction $\mathcal{R}_{L,2}$ (ii) can be replaced by $\mathbb{P}_{\boldsymbol{\gamma}} = \mathbb{P}_{\boldsymbol{\alpha}} \otimes \bigotimes_{k=1}^{p} \mathbb{P}_{\boldsymbol{\beta}_k}$ and φ_{α} and $\varphi_{\boldsymbol{\beta}_k}$ for all $k = 1, \ldots, p-1$ are such that the interior of their sets of zeros is empty. We give the proof with the one of Theorem 2.

Identification when $\mathbb{P}^*_{\boldsymbol{\gamma}}$ has restricted support

Independence in $R_{L,2}$ (ii) precludes cases where $\boldsymbol{\gamma}$ is a function of a deep heterogeneity parameter $\boldsymbol{\Theta}$, hence of the form $\boldsymbol{\gamma}(\boldsymbol{\Theta})$. An alternative to independence is to restrict the support of $\mathbb{P}_{\boldsymbol{\gamma}}$. By the results in Beran and Millar (1994), $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under the following restriction $\mathcal{R}_{L,4}$.

Restriction. *Define the restriction*

$$\mathcal{R}_L := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} : \ R_L \ (i)\text{-}(ii)\},$$

R_L *(i) The interior of the support of* $\boldsymbol{X}$ *is nonempty;*

R_L *(ii) The support of* $\boldsymbol{\gamma}$ *is compact.*

As we shall see, it is possible to remove the unpleasant assumption that the support of the intercept α is compact, which is unusual for an error term. Theorem 3 below is a consequence of Proposition 2 (see also Masten, 2017), which are about identification of a determinate measure by projections. It treats symmetrically the random intercept and slopes and allows for weaker assumptions on the support of the regressors than the restriction $\mathcal{R}_L$.

Assumption 3. *Let, for $k = 1, \ldots, p+1$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences such that $\boldsymbol{M}_{k,0} = 1$. Define the restriction*

$$\mathcal{R}_{L,3}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p+1}) := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} : \ R_{L,3} \ (i)\text{-}(ii)\},$$

$R_{L,3}$ *(i)* *$\prod_{k=1}^{p} V_k \subseteq \mathbb{S}_{\boldsymbol{X}}$, where, for all $k = 1, \ldots, p$, V_k contains an infinite number of points.*

$R_{L,3}$ *(ii)* *For all $k = 1, \ldots, p+1$ and $m \in \mathbb{N}_0$, $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\gamma}}}[|\boldsymbol{\gamma}_k|^m] \leq \boldsymbol{M}_{k,m}$ and, if $\mathbb{S}_{\boldsymbol{\gamma}_k} \subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ satisfies (2.11) and else, if $\mathbb{S}_{\boldsymbol{\gamma}_k} \not\subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ satisfies (2.10).*

Theorem 3. *Let, for $k = 1, \ldots, p+1$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences such that $\boldsymbol{M}_{k,0} = 1$. $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}_{L,3}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p+1})$.*

Usual assumptions made on the support of the regressors (see, *e.g.*, Meister, 2007; Hoderlein et al., 2017) prevent from allowing their support to be discrete. We can replace $R_{L,5}$ (i) by the more general condition: the support of $\boldsymbol{X}$ is a set of uniqueness of homogeneous polynomials in p variables. Examples of latter sets are gathered in Appendix 2.5.2. Note that we allow the support of $\boldsymbol{X}$ to be discrete and without an accumulation point.

Identification without any restrictions on the intercept $\mathbb{P}^*_\alpha$

$R_{L,3}$ (ii) still places restrictions on $\mathbb{P}^*_\alpha$ which we now entirely remove. We use that, under the restriction $\mathcal{R}_1$, for all $t \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$,

$$\mathcal{F}\left[\mathbb{P}_{Y|\boldsymbol{X}=\boldsymbol{x}}\right](t) = \mathcal{F}\left[\mathbb{P}_{\boldsymbol{\gamma}}\right](t, t\boldsymbol{x}).$$

Assumption 4. *Let, for $k = 1, \ldots, p$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences such that $\boldsymbol{M}_{k,0} = 1$. Define the restriction*

$$\mathcal{R}_{L,4}\left(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p}\right) := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} : \ R_{L,4} \ (i)\text{-}(ii)\},$$

where

$R_{L,4}$ (*i*) $\prod_{k=1}^{p} V_k \subseteq \mathbb{S}_{\boldsymbol{X}}$*, where, for all* $k = 1, \ldots, p$*,* $V_k \subset \mathbb{S}_{\boldsymbol{X}_k}$ *contains an accumulation point.*

$R_{L,4}$ (*ii*) *For all* $k = 1, \ldots, p$ *and* $m \in \mathbb{N}_0$*,* $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\beta}}}[|\boldsymbol{\beta}_k|^m] \le \boldsymbol{M}_{k,m}$ *and, if* $\mathbb{S}_{\boldsymbol{\beta}_k} \subseteq \mathbb{R}_+$*,* $\{\boldsymbol{M}_{k,m}\}$ *satisfies* (2.11) *and else, if* $\mathbb{S}_{\boldsymbol{\beta}_k} \not\subseteq \mathbb{R}_+$*,* $\{\boldsymbol{M}_{k,m}\}$ *satisfies* (2.10)*.*

Theorem 4. $\mathbb{P}^*_{\boldsymbol{\gamma}}$ *is identified under* $\mathcal{R}_{L,4}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p})$*, where, for* $k = 1, \ldots, p$*,* $\{\boldsymbol{M}_{k,m}\}$ *are log-convex sequences such that* $\boldsymbol{M}_{k,0} = 1$*.*

The distribution of $\boldsymbol{\beta}_k$ with $R_{L,4}$ can have heavy tails. Using Proposition 4, $R_{L,4}$ (ii) can also be replaced by, for all $k = 1, \ldots, p$, $\mathbb{E}\left[e^{\rho|\boldsymbol{\beta}_k|}\right] \le R$ or $\mathbb{S}_{\boldsymbol{\beta}_k} \subseteq [-\rho, \rho]$, where $\rho > 0$ and $R_{L,4}$ (i) by, $\prod_{k=1}^{p} V_k \subseteq \mathbb{S}_{\boldsymbol{X}}$, where, for all $k = 1, \ldots, p$, $V_k = \mathbb{Z}h_k \subseteq \mathbb{S}_{\boldsymbol{X}_k}$, with $\rho h_k/2 < 1$. Many classical discrete distributions satisfy this condition. Identification does not rely on the knowledge of $\{\boldsymbol{M}_{k,m}\}$ and R.

Nonlinear models with a decomposition admitting a finite number of random coefficients

Consider the nonlinear random coefficients model

$$Y = \sum_{j \in \mathbb{Z}} \boldsymbol{\gamma}_j(\boldsymbol{\Theta}) \boldsymbol{f}_j(X), \tag{2.13}$$

where $\boldsymbol{\Theta}$ is deep underlying heterogeneity with arbitrary dimension, possibly infinite, X is scalar and independent of $\boldsymbol{\Theta}$, and $\mathcal{X} := [-x_0, x_0] \subseteq \mathbb{S}_X$ for $0 < x_0 \le \infty$. We assume that there are only a finite number of random coefficients in this decomposition: there exists $j_0 \in \mathbb{N}$, such that $(\boldsymbol{\gamma}_j)_{|j|>j_0}$ are deterministic. We also assume that, for a.e. $\boldsymbol{\theta} \in \mathbb{S}_{\boldsymbol{\Theta}}$, $\sum_{|j| \le j_0} \boldsymbol{\gamma}_j(\boldsymbol{\theta})^2 + \sum_{|j|>j_0} \boldsymbol{\gamma}_j^2 < \infty$.

Thus, this model considers a restriction of the heterogeneity in the model

$$Y = g(X, \boldsymbol{\Theta}), \tag{2.14}$$

where g is unknown and a.s in $\boldsymbol{\theta} \in \mathbb{S}_{\boldsymbol{\Theta}}$, $x \mapsto g(x, \boldsymbol{\theta})$ can be decomposed in $L^2(-x_0, x_0)$ as,

$$g(x, \boldsymbol{\theta}) = \sum_{j \in \mathbb{Z}} \boldsymbol{\gamma}_j(\boldsymbol{\theta}) \boldsymbol{f}_j(x), \tag{2.15}$$

where $\boldsymbol{\gamma}(\boldsymbol{\theta})$ is an infinite dimensional measure. The difference with (2.13) is that we assume that only low frequencies $|j| \leq j_0$ are heterogenous, and not the high ones.[3] The motivation is to recover the distribution of the elasticity in $\partial_x f(x, \Theta) x / f(x, \Theta)$ or of the marginal effects $\partial_x f(x, \Theta)$ for $x \in [-x_0, x_0]$, assuming that they exist. A important case motivating the decomposition (2.15) is when $(\boldsymbol{f}_j)_{j \in \mathbb{Z}}$ form a Riesz basis of $L^2(-x_0, x_0)$ with $\boldsymbol{f}_0 = 1$ (see Definition 7). Here, the coefficients $\boldsymbol{\gamma}_j(\boldsymbol{\theta})$ are the inner products of $x \mapsto g(x, \boldsymbol{\theta})$ with the unique biorthogonal system (see sections 1.7 and 1.8 in Young, 2001).

Let us introduce the following concept.

Definition 5 (Algebraically independent functions)**.** *Functions $(\boldsymbol{f}_j)_{j \in \mathbb{Z}}$ are algebraically independent over $\mathbb{R}$ if, for all finite subset I of $\mathbb{Z}$, and $P \in \mathbb{R}[X_i, i = 1, \ldots, |I|]$, $P((\boldsymbol{f}_j)_{j \in I}) = 0$ implies that $P = 0$.*

Basic examples of algebraically independent functions over $\mathbb{R}$ are given in Waldschmidt (2000). The following examples are interesting:

$\mathfrak{F}$ (i) The family with only three elements $(\boldsymbol{f}_0, \boldsymbol{f}_1, \boldsymbol{f}_3) = (1, x, e^x)$ hence (2.15) is a finite sum;

$\mathfrak{F}$ (ii) $\boldsymbol{f}_0 = 1$, $\boldsymbol{f}_j(x) = x^{\lambda_j}$, for all $j \in \mathbb{Z} \setminus \{0\}$, where $\lambda_j = j + 1/r^{|j|}$ and $r \in (1, \infty)$ is transcendental over $\mathbb{Z}$ (*i.e.*, it is not the root of any polynomials with coefficients in $\mathbb{Z}$, *e.g.*, e or π). One can also consider $\boldsymbol{f}$ as the Gram-Schmidt orthonormalisation of this family to obtain a basis of $L^2(-x_0, x_0)$, motivating the decomposition (2.15);

$\mathfrak{F}$ (iii) $\boldsymbol{f}_0 = 1$, $\boldsymbol{f}_j(x) = e^{i\pi\lambda_j x/x_0}$, for all $j \in \mathbb{Z} \setminus \{0\}$, where $\lambda_j = j + 1/(4r^{|j|})$ and $r \in (1, \infty)$ is transcendental over $\mathbb{Z}$. Proposition 3 in Appendix shows that

[3] Note that one could consider the more general model where the random components of $\boldsymbol{\gamma}(\boldsymbol{\theta})$ are another arbitrary group of indexes j.

these specific exponential families of functions $\boldsymbol{f}$ are algebraically independent functions and form a Riesz basis of $L^2(-x_0, x_0)$, for $x_0 \in \mathbb{R}_+$ (see Definition 7), which motivates the decomposition (2.15). Here, the coefficients $\boldsymbol{\gamma}_j(\boldsymbol{\theta})$ are the inner products of $x \mapsto f(x, \boldsymbol{\theta})$ with the unique biorthogonal system (see sections 1.7 and 1.8 in Young, 2001).

Assumption 5. *Let, for $k \in \mathbb{Z}$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences such that $\boldsymbol{M}_{k,0} = 1$. Define*

$$\mathcal{R}_{NL}(\{\boldsymbol{M}_{k,m}\}_{k\in\mathbb{Z}}) := \mathcal{R}_1 \cap \{\mathbb{P}_{\boldsymbol{\gamma},X} : \; R_{NL} \text{ (i)-(ii)}\},$$

R_{NL} (i) (a) *There exists $x_0 > 0$ such that $\mathcal{X} := [-x_0, x_0] \subseteq \mathbb{S}_X$ and f_X is bounded by below on $\mathcal{X}$;*

(b) *Or, j_0 is known, $\boldsymbol{\gamma}_j = 0$ for $|j| > j_0$, and $\mathbb{S}_X$ contains an infinite number of points.*

R_{NL} (ii) *For all $|k| \le j_0$ and $m \in \mathbb{N}_0$, $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\gamma}}}[|\boldsymbol{\gamma}_k|^m] \le \boldsymbol{M}_{k,m}$ and, if $\mathbb{S}_{\boldsymbol{\gamma}_k} \subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ satisfies (2.11) and else, if $\mathbb{S}_{\boldsymbol{\gamma}_k} \not\subseteq \mathbb{R}_+$, $\{\boldsymbol{M}_{k,m}\}$ satisfies (2.10);*

R_{NL} (iii) *$(\boldsymbol{f}_j)_{j\in\mathbb{Z}}$ are algebraically independent functions over $\mathbb{R}$*

The condition on R_{NL} (i) (a) on f_X is mild as one can take a small x_0 such that its holds. Under the restriction R_{NL} (i) on the support of $\mathbb{S}_X$, if $(\boldsymbol{f}_j)_{j\in\mathbb{Z}}$ are algebraically independent functions over $\mathbb{R}$, then

$$\{(\boldsymbol{f}_{-j_0}(x), \ldots, 1, \ldots, \boldsymbol{f}_{j_0}(x)), x \in \mathbb{S}_X\} \tag{2.16}$$

is a set of uniqueness of homogeneous polynomials on $\mathbb{R}[X_1, \ldots, X_{2j_0+1}]$, which is a more general condition.

Theorem 5. *Let, for $k \in \mathbb{Z}$, $\{\boldsymbol{M}_{k,m}\}$ be log-convex sequences such that $\boldsymbol{M}_{k,0} = 1$. $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}_{NL}(\{\boldsymbol{M}_{k,m}\}_{k\in\mathbb{Z}})$.*

2.3.2 The random coefficients binary choice model

Consider the equation

$$Y = \mathbb{1}\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{X} \geq 0\}.$$

Assuming that $\mathbb{P}(|(\alpha, \boldsymbol{\beta}^\top)^\top|_2 = 0) = 0$ and this can be equivalently written as

$$Y = \mathbb{1}\{\boldsymbol{\gamma}^\top \boldsymbol{S} \geq 0\},$$

where $\boldsymbol{\gamma} := (\alpha, \boldsymbol{\beta}^\top)^\top / |(\alpha, \boldsymbol{\beta}^\top)^\top|_2$ and $\boldsymbol{S} := (1, \boldsymbol{X}^\top)^\top / |(1, \boldsymbol{X}^\top)^\top|_2$. Clearly $|(1, \boldsymbol{X}^\top)^\top|_2 \geq 1$ and the support of $\boldsymbol{S}$ is a closed subset of the hemisphere $H^+ := \{\boldsymbol{s} \in \mathbb{S}^p : \ \boldsymbol{s}_1 \geq 0\}$. In this section, we consider identification of the density $f^*_{\boldsymbol{\gamma}}$ of $\mathbb{P}^*_{\boldsymbol{\gamma}}$ with respect to σ, which is the surface measure on $\mathbb{S}^p$. We maintain $\mathcal{R}_1$ with $\Gamma = \mathbb{S}^p$ and consider the following restriction.

Assumption 6.

$$\mathcal{R}_{BC,0} := \mathcal{R}_1 \cap \left\{\mathbb{P}_{\boldsymbol{\gamma}, \boldsymbol{X}} \in \mathfrak{M}_1(\mathbb{S}^p \times \mathbb{R}^p) : \ d\mathbb{P}_{\boldsymbol{\gamma}} = f_{\boldsymbol{\gamma}} d\sigma, \ f_{\boldsymbol{\gamma}}(\boldsymbol{u}) f_{\boldsymbol{\gamma}}(-\boldsymbol{u}) = 0 \text{ for a.e. } \boldsymbol{u} \in \mathbb{S}^p\right\}.$$

It is shown in Gautier and Le Pennec (2018) that $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under the restriction

$$\mathcal{R}_{BC,1} := \mathcal{R}_{BC,0} \cap \{\mathbb{S}_{\boldsymbol{X}} = \mathbb{R}^p\}.$$

It is assumed in Gautier and Kitamura (2013) that the support of $\boldsymbol{\gamma}$ lies in an (unknown) hemisphere, namely, that there exists $\boldsymbol{n}$ in $\mathbb{S}^p$ such that $\mathbb{P}(\boldsymbol{n}^\top \boldsymbol{\gamma} \geq 0) = 1$. This assumption first appeared in Ichimura and Thompson (1998) and is too strong for some applications. Indeed, if $\boldsymbol{n} \in H^+$, then we have $\mathbb{P}(Y = 1 | \boldsymbol{S} = \boldsymbol{n}) = 1$, else $\mathbb{P}(Y = 1 | \boldsymbol{S} = -\boldsymbol{n}) = 0$. This means that there exist limits of values of the regressors such that in the limit everyone chooses $Y = 1$ or in the limit everyone chooses $Y = 0$. It is stronger than $f_{\boldsymbol{\gamma}}(\boldsymbol{u}) f_{\boldsymbol{\gamma}}(-\boldsymbol{u}) = 0$ for a.e. $\boldsymbol{u}$ in $\mathbb{S}^p$ which does not imply "unselected samples".

Let us now consider the case where the support of $\boldsymbol{X}$ is a proper subset of $\mathbb{R}^p$ or equivalently the support of $\boldsymbol{S}$ is a proper subset of H^+. Similarly to the previous section, identification can be achieved if the conditional expectation $\boldsymbol{x} \mapsto \mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$ belongs to a class of analytic functions or a quasi-analytic class (see the definition on the sphere in Appendix 2.5.7) and the support of $\boldsymbol{X}$ is the associated set of uniqueness. This can be achieved by imposing restrictions on $\mathbb{P}^*_{\boldsymbol{\gamma}}$ as we now present. $\mathcal{H}^\infty(\mathcal{S})$ denotes the bounded analytic functions on $\mathcal{S}$.

Assumption 7. *Let $0 < \epsilon < 1$. Define the restriction*

$$\mathcal{R}_{BC,2}(\epsilon) := \mathcal{R}_{BC,0} \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{S}} \in \mathfrak{M}_1(\mathbb{S}^p \times \mathbb{S}^p) : \ R_{BC,2} \ \textit{(i)-(ii)}\},$$

where

$R_{BC,2}$ *(i)* $\mathbb{S}_{\boldsymbol{X}}$ *is a set of uniqueness of* $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C}^p : \ |\mathrm{Im}(\boldsymbol{z})|_2 < \epsilon\}\right)$;

$R_{BC,2}$ *(ii)* $\mathbb{E}\left[Y|\boldsymbol{X} = \cdot\right]$ *belongs to* $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C}^p : \ |\mathrm{Im}(\boldsymbol{z})|_2 < \epsilon\}\right)$.

For functions on $\mathbb{S}^p$, the Laplacian Δ has eigenspaces $H_{m,p}$, eigenvalues $\zeta_{m,p} = -m(m+p-1)$, and $Q_{m,p}f(\cdot) = \int_{\mathbb{S}^p} q_{m,p}(\cdot, \boldsymbol{y}) f(\boldsymbol{y}) d\sigma(\boldsymbol{y})$ is the orthogonal projection of f onto $H^{m,p}$ for all $m \in \mathbb{N}_0$.

Theorem 6. *For all $0 < \epsilon \leq 1/2$, $\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}_{BC,2}(\epsilon)$ and a sufficient condition for $R_{BC,2}$ (ii) is*

$$\overline{\lim}_{m \to \infty} \|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\|^{1/m}_{L^1(\mathbb{S}^p)} < 1/(1 + 2\epsilon). \tag{2.17}$$

Clearly, a set of uniqueness of $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C}^p : \ |\mathrm{Im}(\boldsymbol{z})|_2 < \epsilon\}\right)$ is a set of uniqueness of the superset $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C}^p : \ |\mathrm{Im}(\boldsymbol{z})|_\infty < \epsilon\}\right)$. Hence, a sufficient condition for $R_{BC,2}$ (i) is that $U_p \subseteq \mathbb{S}_{\boldsymbol{X}}$ where for all $j = 2, \ldots, p$, $U_j := \bigcup_{\boldsymbol{u} \in U_{j-1}} \left\{(\boldsymbol{u}^\top, v)^\top, \ v \in V_j(\boldsymbol{u})\right\}$, where $V_j(\boldsymbol{u})$ and U_1 are sets of uniqueness of $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C} : \ |\mathrm{Im}(\boldsymbol{z})| < \epsilon\}\right)$. A particular case is a product of sets of uniqueness of $\mathcal{H}^\infty\left(\{\boldsymbol{z} \in \mathbb{C} : \ |\mathrm{Im}(\boldsymbol{z})| < \epsilon\}\right)$. We give examples A.1, QA.1, and Q.A.2 in Appendices 2.5.2 and 2.5.2, where some are discrete.

.

2.3.3 Some panel data models with random coefficients

Extension of the Kotlarski lemma

In some nonparametric panel data random coefficients models considered in this paper, we use the following structure. Consider the equation

$$\boldsymbol{Y}_t = \delta + \boldsymbol{\epsilon}_t,\ t \in \{1, 2\}, \tag{2.18}$$

where the vector of unobserved heterogeneity is $\boldsymbol{\gamma} := (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \delta)$.

Assumption 8. *Let $\{M_m\}$ be a log-convex sequence. Define the restriction*

$$\mathcal{R}_K(\{M_m\}) := \left\{\mathbb{P}_{\boldsymbol{\gamma}} = \mathbb{P}_{\boldsymbol{\epsilon}_1} \otimes \mathbb{P}_{\boldsymbol{\epsilon}_2} \otimes \mathbb{P}_{\delta},\ \forall m \in \mathbb{N}_0\ \mathbb{E}_{\mathbb{P}_{\boldsymbol{\epsilon}_1}}\left[|\boldsymbol{\epsilon}_1|^m\right] \le M_m,\ \textit{and}\ \mathbb{E}_{\mathbb{P}_{\boldsymbol{\epsilon}_1}}[\boldsymbol{\epsilon}_1] = 0\right\}.$$

Lemma 1 in Kotlarski (1967) assumes all characteristic functions do not vanish and in Remark 1 it is written that this can be extended to the case where all characteristic functions are analytic. The next result shows that these assumptions are too strong and identification can be achieved when none of the characteristic functions are analytic and $\boldsymbol{\epsilon}_2$ and δ might not have finite first absolute moments. Evdokimov and White (2012) present a similar result under alternative assumptions, but assuming that the characteristic function of $\boldsymbol{\epsilon}_1$ is analytic.

Theorem 7. *$\mathbb{P}^*_{\boldsymbol{\gamma}}$ is identified under $\mathcal{R}_K(\{M_m\})$ where $\{M_m\}$ is a log-convex sequence which satisfies $M_0 = 1$ and (2.10).*

Again, we refer to Theorem 10 in the Appendix for conditions equivalent to (2.10).

A single-index panel data model with two periods.

Consider the equation

$$\boldsymbol{Y}_t = f(\boldsymbol{\gamma}^\top \boldsymbol{X}_t) + \boldsymbol{\eta}_t,\ t = 1, 2,$$

where f is increasing.

Assumption 9. *Let* $\{M_m\}$ *a log-convex sequence and* $0 < \epsilon < 1$*. Define the restriction*

$$\mathcal{R}_{SI}(\{M_m\}, \epsilon) := \mathcal{R}_{BC,0} \cap \{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}_1,\boldsymbol{X}_2} \in \mathfrak{M}_1 (\mathbb{S}^p \times \mathbb{R}^p \times \mathbb{R}^p) : \ R_{SI} \ (i)\text{-}(iii), R_{BC,2} \ (i)\text{-}(ii)\} ,$$

where

$R_{SI}(i)$ $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{\eta}} = \mathbb{P}_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{\eta}_1} \otimes \mathbb{P}_{\boldsymbol{\eta}_2}$*, for all* $m \in \mathbb{N}_0$, $s_{|\cdot|,\mathbb{P}_{\eta_1}}(m) \leq M_m$, $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\eta}_1}}[\boldsymbol{\eta}_1] = 0$*;*

$R_{SI}(ii)$ $\{(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{S}_{\boldsymbol{X}_1,\boldsymbol{X}_2} : \boldsymbol{x}_1 = \boldsymbol{x}_2\} \neq \emptyset$*;*

$R_{SI}(iii)$ $|\boldsymbol{\gamma}|_2 = 1$*.*

Restriction R_{SI} (ii) means that there exists “stayers" in the population, for which the value of the covariate stays the same accross periods, which is a mild assumption. Theorem 8 below relies on theorems 6 and 7.

Theorem 8. $\mathbb{P}^*_{\boldsymbol{\gamma},\boldsymbol{\eta}}$ *is identified under* $\mathcal{R}_{SI}(\{M_m\}, \epsilon)$*, where* $\{M_m\}$ *is a log-convex sequence satisfying* $M_0 = 1$ *and* (2.10) *and* $0 < \epsilon \leq 1/2$*.*

Linear panel data model where regressors are monomials of a baseline scalar regressor

Consider the equation

$$\boldsymbol{Y}_t = \alpha + \sum_{j=1}^{T} \boldsymbol{\beta}_j \boldsymbol{X}_t^j + \boldsymbol{\epsilon}_t, \ t = 1, \ldots, T, \tag{2.19}$$

where $\boldsymbol{X}_t$ is a scalar regressor and denote by $\boldsymbol{\gamma} := (\boldsymbol{\epsilon}^\top, \alpha, \boldsymbol{\beta}^\top)^\top$. For each t, we have

$$\mathbb{S}_{\boldsymbol{X}_t,\ldots,\boldsymbol{X}_t^T} \subseteq \left\{\mathbf{u} \in \mathbb{R}^p : \ \mathbf{u}_2 = \mathbf{u}_1^2, \ldots, \mathbf{u}_T = \boldsymbol{u}_1^T\right\} ,$$

hence the restriction $R_{L,5}$ (i) does not hold. We show in this section that the availability of T periods allows nonparametric identification.

Remark 1. *Note that* (2.19) *could be generalized to* $\boldsymbol{Y}_t = \alpha + P(\boldsymbol{X}_t) + \boldsymbol{\epsilon}_t$*, for all* $t = 1, \ldots, T$*, where* $P(\boldsymbol{X}_t) = \sum_{j=1}^{T} \boldsymbol{\beta}_j \boldsymbol{X}_t^{\theta(j)}$ *and* $\theta \in \mathbb{N}_0^{\mathbb{N}}$ *and increasing. However,* *using our identification strategy would yield to consider the so-called generalized Vandermonde matrices, whose theoretical properties, in particular their inverse, are not yet well known.*

Restriction. *Let* $\{M_m\} \in (0,\infty]^{\mathbb{N}_0}$, *where* $\{M_m\}$ *is a log-convex sequence. Define the restriction*

$$\mathcal{R}_{LP,0}\left(\{M_m\}\right) := \mathcal{R}_1 \cap \left\{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathfrak{M}_1\left(\mathbb{R}^{3T+1}\right) : \ R_{LP,0} \ \textit{(i)-(ii)}\right\},$$

where

$R_{LP,0}$ *(i)* $\mathbb{P}_{\boldsymbol{\gamma}} = \bigotimes_{j=1}^{T} \mathbb{P}_{\boldsymbol{\epsilon}_j} \otimes \mathbb{P}_{\alpha,\boldsymbol{\beta}}$ *and for all* $m \in \mathbb{N}_0$, $s_{|\cdot|,\mathbb{P}_{\epsilon_1}}(m) \le M_m$ *and* $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\epsilon}_1}}[\boldsymbol{\epsilon}_1] = 0$;

$R_{LP,0}$ *(ii)* $\mathcal{X}_1 := \{\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}} : \boldsymbol{x}_1 = \dots = \boldsymbol{x}_T\} \neq \emptyset$.

Using $R_{LP,0}$ (ii) there exists $r > 0$ such that $r\mathbf{1}_c \in \mathcal{X}_1$ and conditionning on $\boldsymbol{X} = r\mathbf{1}_c$ and using $\delta := \alpha + \sum_{k=1}^{T} \boldsymbol{\beta}_k r^k$ yield, for all $\boldsymbol{t} \in \mathbb{R}^T$,

$$\varphi_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{t}|r\mathbf{1}_c) = \varphi_\delta\left(\sum_{j=1}^{T} \boldsymbol{t}_j\right) \prod_{j=1}^{T} \varphi_{\boldsymbol{\epsilon}_j}(\boldsymbol{t}_j). \tag{2.20}$$

This shows that identification of $\mathbb{P}_{\boldsymbol{\epsilon}}$ is an extension Theorem 7 to T periods (see the proof of Theorem 11). Note that the restriction $R_{LP,0}$ (ii) is weaker than assuming that the covariates are centered $\mathbf{0}_c \in \mathbb{S}_{\boldsymbol{X}}$. The restriction $R_{LP,0}$ (ii) is also maintained in Cooprider et al. (2020), where they focus on the marginals of $\boldsymbol{\gamma}$ without imposing $\mathcal{R}_1$ but considering only the individuals whose $\boldsymbol{X}$ belong to $\mathcal{X}_1$, which are the stayers.

Assumption 10. *Let* $\{M_{m,j}\}_{j=1,2}$ *be two log-convex sequences which satisfy* $M_{0,j} = 1$. *Define*

$$\mathcal{R}_{LP,T}\left(\{M_{m,j}\}_{j=1,2}\right) := \mathcal{R}_{LP,0}(\{M_{m,1}\}) \cap \left\{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathfrak{M}_1\left(\mathbb{R}^{3T+1}\right) : \ R_{LP} \ \textit{(i)-(ii)}\right\},$$

R_{LP} *(i) For all* $t = 1, \dots, T$, *the support of* $\boldsymbol{X}_t$ *contains an accumulation point;*

R_{LP} *(ii) For all* $m \in \mathbb{N}_0$, $\mathbb{E}[|\boldsymbol{\beta}_T|^m] \le M_{m,2}$ *and, if* $\mathbb{S}_{\boldsymbol{\beta}_T} \subseteq \mathbb{R}_+$, $\{M_{m,2}\}$ *satisfies* (2.11) *and else, if* $\mathbb{S}_{\boldsymbol{\beta}_T} \not\subseteq \mathbb{R}_+$, $\{M_{m,2}\}$ *satisfies* (2.10).

Denote by $\boldsymbol{\gamma}^\top = (\alpha, \boldsymbol{\beta})$. Like Theorem 4, Theorem 9 allows to make no assumptions on $\mathbb{P}_\alpha$, and only on $\mathbb{P}_{\boldsymbol{\beta}_T}$.

Theorem 9. $\mathbb{P}^*_{\boldsymbol{\gamma}}$ *is identified under* $\mathcal{R}_{LP,T}(\{M_{m,j}\}_{j=1,2})$*, where* $\{M_{m,j}\}_{j=1,2}$ *are two log-convex sequences satisfying* $M_{0,j} = 1$.

Theorem 9 focuses on identification results without restrictions on $\mathbb{P}^*_{\alpha,\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{T-1}}$. We refer to Appendix 2.5.7 for complementary results which impose restrictions on $\mathbb{P}^*_{\alpha,\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{T-1}}$ but allow for weaker assumptions on the supports of $\boldsymbol{X}_t$ for $t = 1, \ldots, T$.

2.4 Appendix

2.4.1 Identification results of a measure from its projections on 1-dimensional spaces

Proposition 2 below relaxes the set of projections in the Cramér-Wold theorem for determinate measures (see Theorem 3.1 in Cuesta-Albertos et al., 2007). The image of $\mu \in \mathfrak{M}_c(\mathbb{R}^p)$ by φ measurable is denoted by $\varphi_*\mu$. 1-dimensional projections are denoted by $\Pi[\boldsymbol{s}]_*\mu$, where $\Pi[\boldsymbol{s}](\boldsymbol{x}) = \boldsymbol{s}^\top \boldsymbol{x}$ and $\boldsymbol{s} \in \mathbb{S}^p$. Projections on $\boldsymbol{s}^\perp$ are denoted by $\Pi[\boldsymbol{s}^\perp]_*\mu$.

Proposition 2. *Let* U *be a set of uniqueness of homogeneous polynomials in* $\mathbb{R}[X_1, \ldots, X_p]$ *and* $\mathcal{M} \subseteq \mathfrak{M}^*_c(\mathbb{R}^p)$*. Let* $\mu_1, \mu_2 \in \mathfrak{M}_c(\mathbb{R}^p)$*,* μ_1 *be determinate in* $\mathcal{M}$*, and* $\Pi[\boldsymbol{u}]_*\mu_1 = \Pi[\boldsymbol{u}]_*\mu_2$ *for all* $\boldsymbol{u} \in U$*. If* $\mu_2 \in \mathcal{M}$*, we have* $\mu_1 = \mu_2$*. If* $\mathcal{M} \subseteq \mathfrak{M}^*(\mathbb{R}^p)$*, then we have, for all* $m \in \mathbb{N}$*,* $\sum_{j=1}^p s_{|\cdot|,\mu_2}(2m\boldsymbol{e}(j)) < \infty$.

Proof of Proposition 2. Take $k \in \mathbb{N}_0$ and define the homogeneous polynomial

$$P(X_1, \ldots, X_{p+1}) = \int_{\mathbb{R}^{p+1}} \left(\sum_{j=1}^{p+1} X_j \boldsymbol{y}_j \right)^k d(\mu_1 - \mu_2)(\boldsymbol{y}).$$

We have, for all $\boldsymbol{s} \in \mathbb{S}^p$, $P(\boldsymbol{s}) = \int_{\mathbb{R}} t^k d(\Pi[\boldsymbol{s}]_*(\mu_1 - \mu_2))(t)$. Thus, we have $P(\boldsymbol{s}) = 0$ for all $\boldsymbol{s} \in U$, hence $P(\boldsymbol{s}) = 0$ for all $\boldsymbol{s} \in \mathbb{S}^p_{\mathbb{R}}$. Hence, we have for all $\boldsymbol{x} \in \mathbb{R}^{p+1}$ and $k \in \mathbb{N}_0$, $\int_{\mathbb{R}^{p+1}} (\boldsymbol{x}^\top \boldsymbol{y})^k d\mu_1(\boldsymbol{y}) = \int_{\mathbb{R}^{p+1}} (\boldsymbol{x}^\top \boldsymbol{y})^k d\mu_2(\boldsymbol{y})$. This allows to conclude that $\mu_1 = \mu_2$ because $\mu_1, \mu_2 \in \mathcal{M} \subseteq \mathfrak{M}_c^*(\mathbb{R}^p)$.
To handle the case where $\mathcal{M} \subseteq \mathfrak{M}(\mathbb{R}^p)$, it suffices to show that, for all $j = 1, \ldots, p$, $s_{|\cdot|,\mu_2}(d\boldsymbol{e}(j)) < \infty$. This can be done like in the proof of Theorem 3.1 in Cuesta-Albertos et al. (2007). □

Counterexamples exist when the assumptions in Proposition 2 fail (see theorems 3.5 and 3.6 in Cuesta-Albertos et al., 2007 and Theorem 5.4 in Belisle et al., 1997). They do not imply that the restrictions are sharp in the sense that no larger set delivers identification.

2.4.2 Proofs

Proof of Proposition 1. Let us denote by $f = \mathcal{F}[\mu]$. Start by proving the part 1. Because $\mu \in \mathfrak{M}_c^*(\mathbb{R})$, we have $f \in C^\infty(\mathbb{R})$ but also clearly $f \in C^{\{M_m\}}(\mathbb{R}, 1)$ because, for all $m \in \mathbb{N}_0$ and $x \in \mathbb{R}$, $\left|f^{(m)}(x)\right| \leq s_{|\cdot|,|\mu|}(m)$. Theorem 10 yields that $C^{\{M_m\}}(\mathbb{R}, 1)$ is quasi-analytic. Thus, using that for all $m \in \mathbb{N}_0$, $f^{(m)}(0) = i^m s_\mu(m)$, μ is determinate in $\mathfrak{M}_c^*(\mathbb{R})$.
For the part 2, we adapt the proof of Theorem 4.1 in Chalendar and Partington (2007). Define the measure μ_1 on $\mathbb{R}$ by $d\mu_1(t) = d\mu(t^2)$. Thus, we have, for all $m \in \mathbb{N}_0$ such that there exists $n \in \mathbb{N}_0$ $m = 2n$, $s_{|\cdot|,|\mu_1|}(m) = 2s_{|\cdot|,|\mu|}(n)$ hence $\sum_{m\in\mathbb{N}} 1/(s_{|\cdot|,|\mu_1|}(2m))^{1/(2m)} = \sum_{m\in\mathbb{N}} 1/(2s_{|\cdot|,|\mu|}(m))^{1/(2m)}$ and $\sum_{m\in\mathbb{N}} 1/(s_{|\cdot|,|\mu|}(m))^{1/(2m)} \geq \sum_{m\in\mathbb{N}} 1/M_m^{1/(2m)} = \infty$. Hence, applying the first part of the proposition to μ_1, μ is determinate in $\mathfrak{M}_c^*(\mathbb{R}_+)$. Because we have, for all $(m, x) \in \mathbb{N}_0 \times \mathbb{R}$,

$$\left|\mathcal{F}^{(m)}[\mu](x)\right| = \frac{1}{2}\left|\mathcal{F}^{(2m)}[\mu_1](x)\right| \leq 2^{2m-1} s_{|\cdot|,|\mu_1|}(2m) \leq 2^{2m} s_{|\cdot|,|\mu|}(m),$$

$\mathcal{F}[\mu] \in C^{\{M_m\}}(\mathbb{R}, 1)$, hence $\mathcal{F}^{(m)}[\mu](0) = i^m s_\mu(m)$ yields that

$$C^{\{M_m\}}(\mathbb{R}, 1) \bigcap \{\mathcal{F}[\mu],\ \mu \in \mathfrak{M}_c^*(\mathbb{R}_+)\}$$

is quasi-analytic. $\square$

Proof of Theorem 1. This is a particular case of Theorem 2. $\square$

Proof of Theorem 2. We consider $\mathbb{P}_{\gamma,\boldsymbol{X}} \in \mathcal{R}$ such that $\mathbb{P}\left(\mathbb{P}_{\gamma,\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\gamma} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$ and show that $\mathbb{P}_{\gamma} = \mathbb{P}^*_{\gamma}$. Due to the independence assumptions, under both restrictions, we have for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ and t in $\mathbb{R}$,

$$\varphi_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = \varphi_{\alpha}(t) \prod_{j=1}^{p} \varphi_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j) = \varphi^*_{\alpha}(t) \prod_{j=1}^{p} \varphi^*_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j). \tag{2.21}$$

Because the value at 0 of a characteristic function is 1, taking $\boldsymbol{x} = \mathbf{0}_c$, yields $\varphi_{\alpha} = \varphi^*_{\alpha}$. Under $\mathcal{R}'_{L,2}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p})$, because φ^*_{α} is continuous and equal to 1 at 0, we obtain that

$$\prod_{j=1}^{p} \varphi_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j) = \prod_{j=1}^{p} \varphi^*_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j). \tag{2.22}$$

holds for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ and $t \in (-t_0, t_0)$ for t_0 small enough. Taking $\boldsymbol{x} = \boldsymbol{x}(1)$ yields, for all $t \in (-t_0, t_0)$, $\varphi_{\boldsymbol{\beta}_1}(\boldsymbol{x}(1)_1 t) = \varphi^*_{\boldsymbol{\beta}_1}(\boldsymbol{x}(1)_1 t)$, hence $\varphi_{\boldsymbol{\beta}_1}(t) = \varphi^*_{\boldsymbol{\beta}_1}(t)$ for all $t \in (-t_0/\boldsymbol{x}(1)_1, t_0/\boldsymbol{x}(1)_1)$. Hence $\mathbb{P}_{\boldsymbol{\beta}_1}$ and $\mathbb{P}^*_{\boldsymbol{\beta}_1}$ have same moments. Using Assumption $R_{L,2}$ (ii') and Proposition 1, $\mathbb{P}_{\boldsymbol{\beta}_1}$ and $\mathbb{P}^*_{\boldsymbol{\beta}_1}$ are determinate in $\mathfrak{M}^*_1(\mathbb{R})$, thus $\mathbb{P}_{\boldsymbol{\beta}_1}$ and $\mathbb{P}^*_{\boldsymbol{\beta}_1}$ are equal by definition. We conclude by iterating this procedure.

Consider the alternative restriction mentioned after Theorem 2. By (2.21), we have, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ and t in the complement of the set of zeros of φ^*_{α}, that (2.22) holds. Now, using that the complement of the set of zeros of φ^*_{α} is dense and that both $t \mapsto \prod_{j=1}^{p} \varphi_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j)$ and $t \mapsto \prod_{j=1}^{p} \varphi^*_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j)$ are continuous, we obtain that, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ and $t \in \mathbb{R}$, (2.22) holds. Taking now $\boldsymbol{x} = \boldsymbol{x}(1)$ yields, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{\beta}_1}(\boldsymbol{x}(1)_1 t) = \varphi^*_{\boldsymbol{\beta}_1}(\boldsymbol{x}(1)_1 t)$, hence $\varphi_{\boldsymbol{\beta}_1} = \varphi^*_{\boldsymbol{\beta}_1}$. Hence, with the same arguments as before, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$ and $t \in \mathbb{R}$, $\prod_{j=2}^{p} \varphi_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j) = \prod_{j=2}^{p} \varphi^*_{\boldsymbol{\beta}_j}(t\boldsymbol{x}_j)$, and we conclude by iterating this procedure. $\square$

Proof of Theorem 4. Consider $\mathbb{P}_{\gamma,\boldsymbol{X}} \in \mathcal{R}_{L,4}(\{\boldsymbol{M}_{k,m}\}_{k=1}^{p})$ such that $\mathbb{P}\left(\mathbb{P}_{\gamma,\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\gamma} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. Consider, for $l = 1, \dots, p$, $U_l = \prod_{k=1}^{l} V_k \subset \mathbb{S}_{\boldsymbol{X}}$ such that $V_k \subset \mathbb{S}_{\boldsymbol{X}_k}$ contains an accumulation point. For all $(t, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{S}_{\boldsymbol{X}}$, we have $\varphi_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) =$

$\varphi_{\boldsymbol{\gamma}}(t,t\boldsymbol{x}) = \varphi^*_{\boldsymbol{\gamma}}(t,t\boldsymbol{x})$. Take $t\neq 0$ and $\boldsymbol{u}\in U_{p-1}$, the function $g_p : z\in\mathbb{R}\mapsto \varphi_{\boldsymbol{\gamma}}(t,t\boldsymbol{u}_{[p-1]},z)-\varphi^*_{\boldsymbol{\gamma}}(t,t\boldsymbol{u}_{[p-1]},z)$ is zero on tV_p by the assumptions. Denote by

$$\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}(\cdot) = \int_{\mathbb{R}^p} e^{it\left(a+\boldsymbol{b}_{[p-1]}^\top \boldsymbol{u}_{[p-1]}\right)} d\mathbb{P}_{\boldsymbol{\gamma}}\left(a,\boldsymbol{b}_{[p-1]},\cdot\right).$$

Using that for f positive and measurable, $\int_{\mathbb{R}} f(z)\, d|\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}|(z)\leq \mathbb{E}\left[f\left(\boldsymbol{\beta}_p\right)\right]$ (see Theorem 6.2 in Rudin, 1973) and $R_{L,4}$ (ii) yields that $s_{|\cdot|,|\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}|}(m)\leq \boldsymbol{M}_{p,m}$. More, using that $\{s_{|\cdot|,|\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}|}(m)\}$ is log-convex and less than 1 at 0 and Proposition 1 ensures that g_p, which is the Fourier transform of $\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}(\cdot)$, belongs to the class $C^{\{\boldsymbol{M}_{p,m}\}}(\mathbb{R},1)\cap \{\mathcal{F}[\mu],\ \mu\in\mathfrak{M}^*_c(\mathbb{R})\}$. Then, $R_{L,4}$ (ii) yields $\sum_{m\in\mathbb{N}_0} s_{|\cdot|,|\mu_{p,t,\boldsymbol{u},\boldsymbol{x}}|}(m)^{-1/(2m)}=\infty$. Thus, Proposition 1 implies that the latter class is quasi-analytic and using that tV_p is a set of uniqueness of this class, g_p is zero on $\mathbb{R}$. As a result, $\boldsymbol{x}\in\mathbb{R}^p\mapsto\varphi_{\boldsymbol{\gamma}}(t,t\boldsymbol{x})-\varphi^*_{\boldsymbol{\gamma}}(t,t\boldsymbol{x})$ is zero on $U_{p-1}\times\mathbb{R}$.

Take $\boldsymbol{u}\in U_{p-2}$, for all $x\in\mathbb{R}$, the function $g_{p-1}: z\in\mathbb{R}\mapsto\varphi_{\boldsymbol{\gamma}}(t,t\boldsymbol{u}_{[k-1]},z,x)-\varphi^*_{\boldsymbol{\gamma}}(t,t\boldsymbol{u}_{[k-1]},z,x)$ is zero on tV_{p-1}. Using similar arguments as above, Proposition 1 and $R_{L,4}$ (ii) ensure that the Fourier transform of

$$\mu_{p-1,t,\boldsymbol{u},\boldsymbol{x}}(\cdot) = \int_{\mathbb{R}^p} e^{it\left(a+\boldsymbol{b}_{[k-1]}^\top \boldsymbol{u}_{[k-1]}+\boldsymbol{b}_{-[k]}^\top\boldsymbol{x}\right)} d\mathbb{P}_{\boldsymbol{\gamma}}\left(a,\boldsymbol{b}_{[k-1]},\cdot,\boldsymbol{b}_{-[k]}\right)$$

belongs to $C^{\{\boldsymbol{M}_{p,m}\}}(\mathbb{R},1)\cap\{\mathcal{F}[\mu],\ \mu\in\mathfrak{M}^*_c(\mathbb{R})\}$, which is a quasi-analytic class. Using that tV_{p-1} is set of uniqueness of this class yields that g_{p-1} is zero everywhere on the domain of definition of the functions of the class $C^{\{\boldsymbol{M}_{p-1,m}\}}(\mathbb{R},1)$ and thus on $\mathbb{R}$. We conclude by iterating this argument and using continuity for $t=0$. □

Proof of Theorem 5. First, we consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\in\mathcal{R}_{NL}(\{\boldsymbol{M}_{k,m}\}_{k\in\mathbb{Z}})$ with R_{NL} (i) (b) such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right)=\mathbb{P}\left(\mathbb{P}^*_{\boldsymbol{\gamma}}\otimes\mathbb{P}_{\boldsymbol{X}}\right)$ and show that $\mathbb{P}_{\boldsymbol{\gamma}}=\mathbb{P}^*_{\boldsymbol{\gamma}}$. (2.13) is a particular case of (2.12), where $\boldsymbol{\gamma}:=(\alpha,\boldsymbol{\beta}^\top)^\top$ is of the form $\boldsymbol{\gamma}(\boldsymbol{\Theta})$ and $\mathbb{S}_{(\boldsymbol{f}_{-j_0}(X),\dots,\boldsymbol{f}_{j_0}(X))}$ is degenerate, possibly discrete if $\mathbb{S}_X$ is discrete but infinite. Using

$$L_x:\quad t\mapsto\mathcal{F}[\mathbb{P}^*_{\boldsymbol{\gamma}}-\mathbb{P}_{\boldsymbol{\gamma}}](t(\boldsymbol{f}_{-j_0}(x),\dots,\boldsymbol{f}_{j_0}(x))).$$

then, for all $t \in \mathbb{R}$ and $x \in \mathbb{S}_X$, we have $L_x(t) = 0$. Thus, taking k derivatives with respect to t we obtain

$$P(\boldsymbol{f}_{-j_0}(x), \dots, \boldsymbol{f}_{j_0}(x)) = 0, \quad \forall x \in \mathbb{S}_X,$$

where

$$P(Z_1, \dots, Z_{2j_0+1}) = \sum_{\boldsymbol{j} \in \mathbb{N}_0^{2j_0+1}:\ |\boldsymbol{j}|_1 = k} \begin{pmatrix} k \\ \boldsymbol{j}_1, \dots, \boldsymbol{j}_{2j_0+1} \end{pmatrix} c_{\boldsymbol{j}}\ Z_1^{\boldsymbol{j}_1} \dots Z_{2j_0+1}^{\boldsymbol{j}_{2j_0+1}}.$$

and $c_{\boldsymbol{j}} := \int_{\mathbb{R}^{2j_0+1}} \boldsymbol{g}^{\boldsymbol{j}} d(\mathbb{P}^*_{\boldsymbol{\gamma}} - \mathbb{P}_{\boldsymbol{\gamma}})(\boldsymbol{g})$. Using that (2.16) is not contained in any projective hypersurface in $\mathbb{R}^{2j_0+1}$ and R_{NL} (i) (b) we obtain that then $P = 0$ for all $k \in \mathbb{N}_0$. Thus, $\mathbb{P}_{\boldsymbol{\gamma}}$ and $\mathbb{P}^*_{\boldsymbol{\gamma}}$ have the same moments. Using R_{NL} (ii), which ensures that they are determined by their moments using Theorem 2.3 in De Jeu (2003), this yields that $\mathbb{P}_{\boldsymbol{\gamma}} = \mathbb{P}^*_{\boldsymbol{\gamma}}$.

Second, consider R_{NL} (i) (b). Denote by $(\boldsymbol{g}_j)_{j\in\mathbb{Z}}$ the unique biorthogonal system associated to $(\boldsymbol{f}_j)_{j\in\mathbb{Z}}$. First, we recover the expectations of the $(\boldsymbol{\gamma}_j(\boldsymbol{\theta}))_{j\in\mathbb{Z}}$. We use that, under R_{NL} (i) (a), $\mathbb{E}\left[\boldsymbol{\gamma}_j(\boldsymbol{\Theta})\right] = \mathbb{E}\left[Y\boldsymbol{g}_j(X)/f_{X|\mathcal{X}}(X)\mathbb{1}\{X \in \mathcal{X}\}\right]$, where $f_{X|\mathcal{X}}$ the truncated density of $\boldsymbol{X}$ given $\boldsymbol{X} \in \mathcal{X}$. This brings the model back to a finite sum by substracting on both sides of (2.14). Second, we recover j_0. Consider J in $\mathbb{N}_0$, using the proof of the case R_{NL} (i) (b), if

$$Y = \sum_{|j|\le J} \boldsymbol{\gamma}_j(\boldsymbol{\Theta})\boldsymbol{f}_j(x),$$

then the distribution of $\left(\boldsymbol{\gamma}_{j,J}(\boldsymbol{\Theta})\right)_{|j|\le J}$ is identified. Then, j_0 is the smallest index J such that the following equation holds

$$\forall t \in \mathbb{R}, \quad \epsilon(t, X) := \mathbb{E}\left[e^{itY} - e^{it\sum_{|j|\le J}\boldsymbol{\gamma}_{j,J}(\boldsymbol{\Theta})\boldsymbol{f}_j(X)}\middle| X\right] = 0,$$

which yields the result. $\square$

Proof of Theorem 6. Take $0 < \epsilon \le 1/2$ and consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}_{BC,2}(\epsilon)$ such that

$\mathbb{P}(\mathbb{P}_{\gamma,\boldsymbol{X}}) = \mathbb{P}\left(\mathbb{P}^*_{\gamma} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. For $\boldsymbol{s} \in \mathbb{S}^p$ and $f \in L^1(\mathbb{S}^p)$, define $\mathcal{T}f(\boldsymbol{s}) := \int_{\mathbb{S}^p} \mathbb{1}\{\boldsymbol{u}^\top \boldsymbol{s} \geq 0\} f(\boldsymbol{u}) d\sigma(\boldsymbol{u}) - 1/2$. Recall that, from Gautier and Kitamura (2013),

$$\mathcal{T}f(\boldsymbol{s}) = \mathcal{T}f^-(\boldsymbol{s}) = (\mathcal{T}f)^-(\boldsymbol{s}) \tag{2.23}$$

and, from the restrictions $\mathcal{R}_{BC,0}$,

$$\forall \boldsymbol{s} \in \mathbb{S}_{\boldsymbol{S}},\ \mathcal{T}f_{\gamma}(\boldsymbol{s}) + \frac{1}{2} = \mathbb{E}[Y|\boldsymbol{S} = \boldsymbol{s}]. \tag{2.24}$$

One has in $\mathcal{S}'$ (see Gautier and Kitamura, 2013), for $f \in L^q(\mathbb{S}^p)$,

$$\mathcal{T}f = \sum_{m \in \mathbb{N}_0} \lambda_{2m+1,p} Q_{2m+1,p} f, \tag{2.25}$$

where $\lambda(2m+1,p) = (-1)^m 2\pi^{p/2} 1 \cdot 3 \cdots (2m-1)/(\Gamma(p/2)p(p+2)\cdots(p+2m))$ for all $m \in \mathbb{N}_0$. Hence, we have, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$,

$$\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \sum_{m \in \mathbb{N}_0} \lambda_{2m+1,p} Q_{2m+1,p} f_{\gamma}\left(\frac{(1,\boldsymbol{x})}{\sqrt{1+|\boldsymbol{x}|_2^2}}\right) + \frac{1}{2}.$$

Denote the dot product and Lie norm by $\boldsymbol{z}^2 = \sum_{k=1}^p \boldsymbol{z}_k^2$ and $L(\boldsymbol{z}) = \sqrt{|\boldsymbol{z}|_2^2 + \sqrt{|\boldsymbol{z}|_2^4 - |\boldsymbol{z}^2|^2}}$ respectively. Denote also by $G(\boldsymbol{z}) = \sum_{m \in \mathbb{N}_0} \lambda_{2m+1,p} Q_{2m+1,p} f_{\gamma}(\boldsymbol{z}) + 1/2$ for all $\boldsymbol{z} \in \mathbb{C}^{p+1}$ and $F(\boldsymbol{z}) = (f_1(\boldsymbol{z}), \ldots, f_{p+1}(\boldsymbol{z}))^\top$ for all $\boldsymbol{z} \in \mathbb{C}^p$, where:

$$f_1(\boldsymbol{z}) = \frac{1}{\sqrt{1+\boldsymbol{z}^2}},\ f_2(\boldsymbol{z}) = \frac{\boldsymbol{z}_1}{\sqrt{1+\boldsymbol{z}^2}},\ \ldots\ ,\ f_{p+1}(\boldsymbol{z}) = \frac{\boldsymbol{z}_p}{\sqrt{1+\boldsymbol{z}^2}}.$$

We have $\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = G \circ F(\boldsymbol{x})$ and we need to prove $G \circ F \in \mathcal{A}(\mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}})$. For this, we check the conditions of Theorem 1.2.3 in Rudin (1980). First, we have $F \in \mathcal{A}(\mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}})$. Indeed, for all $\boldsymbol{z} \in \mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}}$, $1 + \boldsymbol{z}^2 \in (\mathbb{C} \setminus (-\infty, 0])^{p+1}$. Now, for all $\boldsymbol{z} \in \mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}}$, we have

$$L(F(\boldsymbol{z}))^2 = \frac{L(1,\boldsymbol{z})^2}{\left|\sqrt{1+\boldsymbol{z}^2}\right|^2} \leq \frac{1 + (|\mathrm{Re}(\boldsymbol{z})|_2 + |\mathrm{Im}(\boldsymbol{z})|_2)^2}{1 + |\mathrm{Re}(\boldsymbol{z})|_2^2 - |\mathrm{Im}(\boldsymbol{z})|_2^2} \leq 1 + 2\epsilon.$$

Moreover, using Lemma 3.23 in Morimoto (1962) for the first display and using the Young inequalities (see, *e.g.*, Gautier and Kitamura, 2013) for the third display, we have, for all $\boldsymbol{z} \in \mathbb{C}^{p+1}$ such that $L(\boldsymbol{z}) \leq \sqrt{1+2\epsilon}$,

$$\begin{aligned}
\left|\lambda_{2m+1,p} Q_{2m+1,p} f_{\boldsymbol{\gamma}}(\boldsymbol{z})\right| &\leq L(\boldsymbol{z})^{2m+1} \left\|\lambda_{2m+1,p} Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^\infty(\mathbb{S}^p)} \\
&\leq L(\boldsymbol{z})^{2m+1} \left\|\mathcal{T}\left[Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right]\right\|_{L^\infty(\mathbb{S}^p)} \\
&\leq L(\boldsymbol{z})^{2m+1} \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)} \\
&\leq (1+2\epsilon)^{m+1/2} \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)}.
\end{aligned}$$

Now, because $\overline{\lim}_{m\to\infty} \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)}^{1/m} \leq q/(1+2\epsilon)$ with $q<1$, there exists $m_0 \in \mathbb{N}$ such that, for all $m \geq m_0$, $(1+2\epsilon)^{m+1/2} \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)} \leq \sqrt{1+2\epsilon}$. Hence, there exists $C_\epsilon < \infty$ such that, for all $m \in \mathbb{N}_0$, $(1+2\epsilon)^{m+1/2} \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)} \leq C_\epsilon$. As a result, for all $\boldsymbol{z} \in \mathbb{C}^{p+1}$ such that $L(\boldsymbol{z}) < \sqrt{1+2\epsilon}$, we have $\sup_{m\in\mathbb{N}_0} \left|\lambda_{2m+1,p} Q_{2m+1,p} f_{\boldsymbol{\gamma}}(\boldsymbol{z})\right| < \infty$. Using the fact that $\overline{\boldsymbol{z}} \mapsto Q_{2m+1,p} f_{\boldsymbol{\gamma}}(\overline{\boldsymbol{z}})$ are homogeneous harmonic polynomials and Theorem 1.5.6 in Rudin (1980) yields $G \circ F \in \mathcal{A}(\mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}})$. Moreover, for all $\boldsymbol{z} \in \mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}}$, we have

$$\left|G \circ F(\boldsymbol{z})\right| \leq \sqrt{1+2\epsilon} \sum_{m\in\mathbb{N}_0} \left((1+2\epsilon) \left\|Q_{2m+1,p} f_{\boldsymbol{\gamma}}\right\|_{L^1(\mathbb{S}^p)}^{1/m}\right)^m + \frac{1}{2},$$

and the upper bound is a convergent series using $R_{BC,2}$ (ii). Hence, $G \circ F \in \mathcal{A}(\mathbb{R}^p + i\epsilon\mathbb{B}^p_{\mathbb{R}})$ is bounded. □

Proof of Theorem 7. Consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}_K(\{M_m\})$ such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. For $t_0 > 0$ small enough, on $(-t_0, t_0)$, φ_δ, φ^*_δ, $\varphi_{\boldsymbol{\epsilon}_1}$, $\varphi^*_{\boldsymbol{\epsilon}_1}$, $\varphi_{\boldsymbol{\epsilon}_2}$, $\varphi^*_{\boldsymbol{\epsilon}_2}$ do not vanish, hence there exist nonvanishing continuous functions p_δ, $p_{\boldsymbol{\epsilon}_1}$, and $p_{\boldsymbol{\epsilon}_2}$ such that $\varphi_\delta = p_\delta \varphi^*_\delta$, $\varphi_{\boldsymbol{\epsilon}_1} = p_{\boldsymbol{\epsilon}_1} \varphi^*_{\boldsymbol{\epsilon}_1}$, $\varphi_{\boldsymbol{\epsilon}_2} = p_{\boldsymbol{\epsilon}_2} \varphi^*_{\boldsymbol{\epsilon}_2}$. By the restriction $\mathcal{R}_K(\{M_m\})$, we have $p_{\boldsymbol{\epsilon}_1} \in C^\infty(-t_0, t_0)$. We have

$$p_\delta(\boldsymbol{t}_1 + \boldsymbol{t}_2) p_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_1) p_{\boldsymbol{\epsilon}_2}(\boldsymbol{t}_2) = 1 \quad \text{for all } -t_0 < \boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_1 + \boldsymbol{t}_2 < t_0, \tag{2.26}$$

hence

$$p_\delta(t)p_{\boldsymbol{\epsilon}_1}(t) = 1 \quad \text{and} \quad p_\delta(t)p_{\boldsymbol{\epsilon}_2}(t) = 1 \quad \text{for all } -t_0 < t < t_0. \tag{2.27}$$

Injecting (2.27) into (2.26) we obtain, for all $-t_0 < \boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_1 + \boldsymbol{t}_2 < t_0$, $p_\delta(\boldsymbol{t}_1 + \boldsymbol{t}_2) = p_\delta(\boldsymbol{t}_1)p_\delta(\boldsymbol{t}_2)$, which, using again (2.27), yields $p_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_1 + \boldsymbol{t}_2) = p_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_1)p_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_2)$. Hence, we have $p'_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_1 + \boldsymbol{t}_2) = p'_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_1)p_{\boldsymbol{\epsilon}_1}(\boldsymbol{t}_2)$, which at $\boldsymbol{t}_1 = 0$ and $\boldsymbol{t}_2 = t$ yields

$$p'_{\boldsymbol{\epsilon}_1}(t) = p'_{\boldsymbol{\epsilon}_1}(0)p_{\boldsymbol{\epsilon}_1}(t) \quad \text{for all } -t_0 < t < t_0, \tag{2.28}$$

where $p'_{\boldsymbol{\epsilon}_1}(0) = \varphi'_{\boldsymbol{\epsilon}_1}(0) - \left(\varphi^*_{\boldsymbol{\epsilon}_1}\right)'(0) = i(\mathbb{E}_{\mathbb{P}_{\boldsymbol{\epsilon}_1}}[\boldsymbol{\epsilon}_1] - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\epsilon}_1}}[\boldsymbol{\epsilon}_1]) \in i\mathbb{R}$ which we denote by $p'_{\boldsymbol{\epsilon}_1}(0) := ib$. Thus, we obtain, for all $t_0 < t < t_0$, $p_{\boldsymbol{\epsilon}_1}(t) = \exp(ibt)$. Moreover, $t \mapsto \varphi_{\boldsymbol{\epsilon}_1}(t) - \exp(ibt)\varphi^*_{\boldsymbol{\epsilon}_1}(t) \in C^{\{M_m\}}(\mathbb{R})$ and $C^{\{M_m\}}(\mathbb{R})$ is a $\{0\}-$quasi-analytic class, hence the function is identically 0 on $\mathbb{R}$ because it is 0 on $(-t_0, t_0)$ (see Example QA.2 in the supplemental Appendix). Thus $\mathbb{P}_{\boldsymbol{\epsilon}_1}$ and $\mathbb{P}^*_{\boldsymbol{\epsilon}_1}$ might only differ by their averages but it is assumed that they are both mean 0, hence $\mathbb{P}_{\boldsymbol{\epsilon}_1} = \mathbb{P}^*_{\boldsymbol{\epsilon}_1}$.

Now, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{Y}_1}(t) = \varphi_\delta(t)\varphi^*_{\boldsymbol{\epsilon}_1}(t) = \varphi^*_\delta(t)\varphi^*_{\boldsymbol{\epsilon}_1}(t)$, hence, because the zeros of $\varphi^*_{\boldsymbol{\epsilon}_1}$ are isolated (see Lemma 4.8 in Belisle et al., 1997) and φ_δ and φ^*_δ are continuous, we obtain $\varphi_\delta = \varphi^*_\delta$.

Similarly, because, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{Y}_2 - \boldsymbol{Y}_1}(t) = \varphi^*_{\boldsymbol{\epsilon}_2}(t)\varphi^*_{\boldsymbol{\epsilon}_1}(-t) = \varphi_{\boldsymbol{\epsilon}_2}(t)\varphi^*_{\boldsymbol{\epsilon}_1}(-t)$, the zeros of $\varphi^*_{\boldsymbol{\epsilon}_1}$ are isolated and $\varphi_{\boldsymbol{\epsilon}_2}$ and $\varphi^*_{\boldsymbol{\epsilon}_2}$ are continuous, we obtain $\varphi_{\boldsymbol{\epsilon}_2} = \varphi^*_{\boldsymbol{\epsilon}_2}$, hence the result. $\square$

Proof of Theorem 8. Consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}_{SI}(\{M_m\}, \epsilon)$ such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. Denote by $\mathcal{X}_1 := \{(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{S}_{\boldsymbol{X}_1,\boldsymbol{X}_2} : \boldsymbol{x}_1 = \boldsymbol{x}_2\}$. Using R_{SI} (ii), there exists $r > 0$ such that $r(\mathbf{1}_c, \mathbf{1}_c) \in \mathcal{X}_1$, hence, for all $\boldsymbol{t} \in \mathbb{R}^2$, using $\delta := f(\boldsymbol{\gamma}^\top(r\mathbf{1}_c))$, we have

$$\varphi_{\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{t}|r(\mathbf{1}_c, \mathbf{1}_c)) = \varphi_\delta(\boldsymbol{t}_1 + \boldsymbol{t}_2)\varphi_{\boldsymbol{\eta}_1}(\boldsymbol{t}_1)\varphi_{\boldsymbol{\eta}_2}(\boldsymbol{t}_2) = \varphi^*_\delta(\boldsymbol{t}_1 + \boldsymbol{t}_2)\varphi^*_{\boldsymbol{\eta}_1}(\boldsymbol{t}_1)\varphi^*_{\boldsymbol{\eta}_2}(\boldsymbol{t}_2). \tag{2.29}$$

Then, following the same steps as in the proof of Theorem 7 yields $\varphi_{\boldsymbol{\eta}_1} = \varphi^*_{\boldsymbol{\eta}_1}$. Now, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{Y}_1|\boldsymbol{X}_1}(t|r\mathbf{1}_c) = \varphi_\delta(t)\varphi^*_{\boldsymbol{\eta}_1}(t) = \varphi^*_\delta(t)\varphi^*_{\boldsymbol{\eta}_1}(t)$, hence, because the zeros of $\varphi^*_{\boldsymbol{\eta}_1}$ are isolated (see Lemma 4.8 in Belisle et al., 1997) and φ_δ and φ^*_δ are continuous, we obtain $\varphi_\delta = \varphi_\delta^*$. Similarly, because, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{Y}_2-\boldsymbol{Y}_1|\boldsymbol{X}_1,\boldsymbol{X}_2}(t|(r\mathbf{1}_c, r\mathbf{1}_c)) = \varphi_{\boldsymbol{\eta}_2}^*(t)\varphi_{\boldsymbol{\eta}_1}^*(-t) = \varphi_{\boldsymbol{\eta}_2}(t)\varphi_{\boldsymbol{\eta}_1}^*(-t)$, the zeros of $\varphi_{\boldsymbol{\eta}_1}^*$ are isolated and $\varphi_{\boldsymbol{\eta}_2}$ and $\varphi_{\boldsymbol{\eta}_2}^*$ are continuous, we obtain $\varphi_{\boldsymbol{\eta}_2} = \varphi_{\boldsymbol{\eta}_2}^*$.

Thus, we obtain, for all $\boldsymbol{t} \in \mathbb{R}^2$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{S}_{\boldsymbol{X}_1,\boldsymbol{X}_2}$,

$$\varphi_{f(\boldsymbol{\gamma}^\top \boldsymbol{x}_1), f(\boldsymbol{\gamma}^\top \boldsymbol{x}_2)}(\boldsymbol{t}) = \varphi^*_{f(\boldsymbol{\gamma}^\top \boldsymbol{x}_1), f(\boldsymbol{\gamma}^\top \boldsymbol{x}_2)}(\boldsymbol{t}). \tag{2.30}$$

This amount to study identification in $\boldsymbol{Z}_t = f(\boldsymbol{\gamma}^\top \boldsymbol{X}_t)$, $t = 1, 2$ and $\boldsymbol{Z}_2 \geq \boldsymbol{Z}_1$ is equivalent to $\boldsymbol{\gamma}^\top (\boldsymbol{X}_2 - \boldsymbol{X}_1) \geq 0$. This is the binary choice model, hence Theorem 6 yields the result. □

Proofs of Theorem 9. We use $\mathcal{X}_0 := \{\boldsymbol{x} \in \mathbb{R}^T : \prod_{j=1}^T \boldsymbol{x}_j \neq 0, \prod_{m \neq j}(\boldsymbol{x}_m - \boldsymbol{x}_j) \neq 0\}$ and, for all $(\boldsymbol{v}, \boldsymbol{x}) \in \mathbb{R}^T \times \mathcal{X}_0$,

$$\Theta : \ (\boldsymbol{v}, \boldsymbol{x}) \mapsto \sum_{k=1}^T \boldsymbol{v}_k \left(\sum_{j=1}^T b_{jk}(\boldsymbol{x}) \boldsymbol{x}_j^T \right), \tag{2.31}$$

$$\begin{cases} b_{jk}(\boldsymbol{x}) := \dfrac{(-1)^{k+1}}{\prod\limits_{\substack{1 \leq m \leq T \\ m \neq j}} (\boldsymbol{x}_m - \boldsymbol{x}_j)} \sum\limits_{\substack{1 \leq i_1 < \dots < i_{T-k} \leq T \\ i_1, \dots, i_{T-k} \neq j}} \boldsymbol{x}_{i_1} \boldsymbol{x}_{i_2} \dots \boldsymbol{x}_{i_{T-k}} & \text{for all } k \neq T \\ b_{jk}(\boldsymbol{x}) := \dfrac{(-1)^{T+1}}{\prod\limits_{\substack{1 \leq m \leq T \\ m \neq j}} (\boldsymbol{x}_m - \boldsymbol{x}_j)} & \text{for } \ k = T. \end{cases}$$

Consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}_{LP,T}(\{M_{m,j}\}_{j=1,2})$ such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma}}^* \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. Using the same arguments as in the proof of Theorem 11, we obtain $\varphi_{\boldsymbol{\epsilon}_j} = \varphi_{\boldsymbol{\epsilon}_j}^*$ for all $j = 2, \dots, T$ and that for all $\boldsymbol{v} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathcal{X}_0$,

$$\varphi_{\alpha,\boldsymbol{\beta}}\left(\boldsymbol{v}, \Theta(\boldsymbol{v}, \boldsymbol{x})\right) = \varphi^*_{\alpha,\boldsymbol{\beta}}\left(\boldsymbol{v}, \Theta(\boldsymbol{v}, \boldsymbol{x})\right). \tag{2.32}$$

holds. Then, using arguments from the proof of Theorem 4, R_{LP} (ii) and Proposition 1 ensure that, for all $\boldsymbol{v} \in \mathbb{R}^T$, $z \in \mathbb{R} \mapsto (\varphi_{\alpha,\boldsymbol{\beta}} - \varphi^*_{\alpha,\boldsymbol{\beta}})(\boldsymbol{v}, z)$ can be extended uniquely to $\mathcal{C}^{\{M_{m,2}\}}(\mathbb{R}) \cap \{\mathcal{F}[\mu],\ \mu \in \mathfrak{M}_c^*(\mathbb{R})\}$ which is a quasi-analytic class. Finally, for all $\boldsymbol{v} \in \mathbb{R}^T \setminus \mathcal{V}$ where $\mathcal{V}$ is such that $\mathbb{R}^T \setminus \mathcal{V}$ is dense in $\mathbb{R}^T$, R_{LP} (i) and that $\boldsymbol{x} \mapsto \Theta(\boldsymbol{v}, \boldsymbol{x})$ is continuous on $\mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0$ yield that $U_{T,\boldsymbol{v}} = \{\Theta(\boldsymbol{v}, \boldsymbol{u}),\ \forall \boldsymbol{u} \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0\}$ contains a point of accumulation. We obtain that, for all $\boldsymbol{v} \in \mathbb{R}^T \setminus \mathcal{V}$ and $z \in \mathbb{R}$, $\varphi_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, z) = \varphi^*_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, z)$ hence $\mathbb{P}_{\alpha,\boldsymbol{\beta}} = \mathbb{P}^*_{\alpha,\boldsymbol{\beta}}$ by continuity for all $\boldsymbol{v} \in \mathbb{R}^T$. $\square$

2.5 Supplemental Appendix

2.5.1 Notations and preliminaries

$\mathbb{C}_+ = \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$, $\mathbb{K}_d[X_1, \ldots, X_p]$ is the subset of $\mathbb{K}[X_1, \ldots, X_p]$ of polynomials of degree at most d (by convention $\mathbb{K}_\infty[X_1, \ldots, X_p] = \mathbb{K}[X_1, \ldots, X_p]$) in the ring $\mathbb{K}$, $\mathbb{K}(X_1, \ldots, X_p)$ is the corresponding field of fractions, and we use the same notation for polynomial or fractional functions, $a \vee b$ is the maximum between a and b, $|E|$ the cardinal of E or the length of the interval if E is an interval. The interior of the set $\mathcal{S}$ is $\overset{\circ}{\mathcal{S}}$ and the closure is $\overline{\mathcal{S}}$. The sequences $\{1_m\}$ and $\{M_m^c\}$ are log-convex. $\{M_m^c\}$ is the *convex regularization of* $\{M_m\}$ *by means of the logarithm* (*i.e.*, the largest convex minorant of $m \mapsto \log(M_m)$).

Lemma 1 (Lemma 3.3 (2) in De Jeu (2004))**.** *Let* $(a_m)_{m \in \mathbb{N}_0}$ *be a nonnegative nonincreasing sequence of real numbers. If* $k, l \in \mathbb{N}$*, then we have:* $\sum_{m \in \mathbb{N}_0} a_{km} = \infty$ *if and only if* $\sum_{m \in \mathbb{N}_0} a_{lm} = \infty$.

By the Hölder inequality, if $\|\cdot\|$ is a norm and $\mu \in \mathfrak{M}^*(\mathbb{R}^p)$, $\left(\int_{\mathbb{R}^p} \|\boldsymbol{x}\|^m d\mu(\boldsymbol{x})\right)_{m \in \mathbb{N}_0}$ is log-convex and $M_0 = \mu(\mathbb{R}^p)$. The support of a complex measure μ is denoted by $\operatorname{supp}(\mu)$. If $\mathcal{S} \subseteq \mathbb{R}^p$, $C^{\{\boldsymbol{M}_m\}}(\mathcal{S})$ is a vector space. If $\mathcal{S} = \mathbb{R}$ and $\underline{\lim}_{m \to \infty} M_m^{1/m} = 0$ then $C^{\{M_m\}}(\mathcal{S}) = C^{\{0_m\}}(\mathcal{S})$ and if $0 < \underline{\lim}_{m \to \infty} M_m^{1/m} < \infty$ then $C^{\{M_m\}}(\mathcal{S}) = C^{\{1_m\}}(\mathcal{S})$, else, if the terms in the sequence are positive, $C^{\{M_m\}}(\mathcal{S}) = C^{\{M_m^c\}}(\mathcal{S})$. Also, recall (see Rudin, 1973) that, if $M_m = 1$ and $(M_m)_{m \in \mathbb{N}_0}$ is log-convex, then $C^{\{M_m\}}(\mathbb{R})$ is an algebra with respect to the multiplication.

2.5.2 Examples of sets of uniqueness

We give here examples of sets of uniqueness which are used in Section 2.3.

Sets of uniqueness of homogeneous polynomials

Let us start by considering homogeneous polynomials of degree d. U is a set of uniqueness of homogeneous polynomials in $\mathbb{K}_d[X_1, \ldots, X_{p+1}]$, where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, if there exists $A \in GL(\mathbb{K}^{p+1})$ such that $\{(1, \boldsymbol{u}^\top)^\top : \boldsymbol{u} \in \widetilde{U}\} \subseteq AU$ and $\widetilde{U}$ is a set of uniqueness of $\mathbb{K}_d[X_1, \ldots, X_p]$. Due to the use of $A \in GL(\mathbb{K}^{p+1})$, $\widetilde{U}$ can be defined in any system of coordinates. Because polynomials remain polynomials when composed by affine transformations, 0 plays no role. Polynomials are analytic, hence the sets of uniqueness in sections 2.5.2 and 2.5.2 are sets of uniqueness of polynomials. Let us present two examples.

Example P.1.

(i) $p = 1$ and $\widetilde{U}$ contains at least $d + 1$ points;

(ii) $p \geq 2$ and $\widetilde{U} = U_p$, where U_p is defined recursively via U_1, which contains at least $d+1$ points, and, for all $j = 2, \ldots, p$, $U_j = \bigcup_{\boldsymbol{u} \in U_{j-1}} \{(\boldsymbol{u}^\top, v)^\top,\ v \in V_j(\boldsymbol{u})\}$, where $V_j(\boldsymbol{u})$ contains at least $d + 1$ points.

The set in Example P.1 is a set of uniqueness because $P \in \mathbb{K}_d[X_1, \ldots, X_p]$ can be written as $P(X_1, \ldots, X_p) = \sum_{k=0}^{d} Q_k(X_1, \ldots, X_{p-1}) X_p^k$, where $Q_k \in \mathbb{K}_d[X_1, \ldots, X_{p-1}]$. Specifically, $\widetilde{U} = \prod_{j=1}^{p} V_j$, where, for all $j = 1, \ldots, p$, $V_j \subseteq \mathbb{R}$ has at least $d+1$ points is a set of uniqueness of homogeneous polynomials of degree d. The more general formulation in Example P.1 allows to show that the infinite fan $\{\boldsymbol{x} \in \mathbb{R}^2 :\ \exists n \in \mathbb{N},\ \boldsymbol{x}_2 = n\boldsymbol{x}_1\}$ and infinite staircase $\{\boldsymbol{x} \in \mathbb{R}^2 :\ \boldsymbol{x}_2 = \lceil \boldsymbol{x}_1 \rceil\}$ (see Bochnak et al., 1998), among others, are sets of uniqueness as well. Recall the following definition.

Definition 6. *Let A be a ring and k a subring.*

1. *$f \in A$ is* transcendental over *k if, for all $P \in k[X]$, $P(f) = 0$ implies that $P = 0$.*

2. *$S \subseteq A$ is* algebraically independent over *k if, for all $p \in \mathbb{N}_0$ and $(f_j)_{j=1}^{p} \in S^p$, $P \in k[X_1, \ldots, X_p]$, $P(f_1, \ldots, f_p) = 0$ implies that $P = 0$.*

Basic elements and examples of algebraically independent functions are given in Waldschmidt (2000).

Example P.2. There exist $f_j \in \mathcal{F}(\mathcal{S})$ for all $j = 1, \ldots, p$, where $\mathcal{S} \subseteq \mathbb{C}^q$ for $q \in \mathbb{N}$ and $\mathcal{F}(\mathcal{S})$ is an algebra with respect to multiplication, which form an algebraically independent set over $\mathbb{C}$ and a set of uniqueness $U_q \subset \mathcal{S}$ of $\mathcal{F}(\mathcal{S})$ (see sections 2.5.2 and 2.5.2) such that $\{(f_1(\boldsymbol{u}), \ldots, f_p(\boldsymbol{u}))^\top,\ \boldsymbol{u} \in U_q\} \subseteq \widetilde{U}$.

Example P.2 allows p to be large and $\widetilde{U}$ discrete. The next result gives a family of algebraically independent functions which allows to uniquely decompose square-integrable functions. Recall the following definition (see Young, 2001).

Definition 7. *A* Riesz basis $(f_j)_{j\in\mathbb{N}_0}$ *of a separable Hilbert space is the image of an orthonormal basis by a bounded invertible operator.*

Proposition 3. *Let* $x_0 \in \mathbb{R}_+$, $\lambda_j = j + 1/(4r^{|j|})$, *where, for all* $j \in \mathbb{Z} \setminus \{0\}$, $f_j(z) = e^{i\pi\lambda_j z/x_0}$, $f_0 = 1$, *and* $r \in (1, \infty)$ *transcendental over* $\mathbb{Z}$. *We have:*

(P6.1) $(f_j)_{j\in\mathbb{Z}\setminus\{0\}}$ *is an algebraically independent family of functions over* $\mathbb{C}(X)$*;*

(P6.2) $(f_j)_{j\in\mathbb{Z}}$ *is a* Riesz basis *of* $L^2(-x_0, x_0)$.

Proof of Proposition 3. Start by proving (P6.1). Take $J \subseteq (\mathbb{Z}\setminus\{0\})$ finite , $J_s \subseteq J$ the positive indices j in J such that $-j \in J$.
Let $(b_j)_{j\in J} \in \mathbb{N}_0^{|J|}$ such that $\sum_{j\in J} b_j\lambda_j = 0$. We have

$$\sum_{j\in J} b_j\lambda_j = \sum_{j\in J_s} \frac{b_j + b_{-j}}{4r^j} + \sum_{j\in J\setminus J_s} \frac{b_j}{4r^{|j|}} + \sum_{j\in J} jb_j,$$

hence, $P(r) = 0$, where $j_0 = \max_{j\in J} |j|$ and

$$P(X) = \sum_{j\in J_s} (b_j + b_{-j})X^{j_0-j} + \sum_{j\in J\setminus J_s} b_j X^{j_0-|j|} + 4X^{j_0} \sum_{j\in J} jb_j \in \mathbb{Z}[X].$$

Because r is transcendental over $\mathbb{Z}$ and $0 \notin J$, for all $j \in J \setminus J_s$, $b_j = 0$, and, for all $j \in J_s$, $b_j + b_{-j} = 0$ thus $b_j = b_{-j} = 0$ because b_j and b_{-j} are nonnegative. Hence, for all $j \in J$, $b_j = 0$.
Now, take $P \in \mathbb{C}(X)[X_1, \ldots, X_p]$, where $p = |J|$, which is zero when evaluated at

$(f_j)_{j\in J}$. Hence, we have $\sum_k c_k(z)\exp(z\sum_{j\in J} b_{k,j}\lambda_j) = 0$, where the sum over k is finite, all $b_{k,j}$ belong to $\mathbb{N}_0$, and c_k are rational functions. All exponentials in this sum are distinct by the above computations and we conclude that all c_ks are zero by taking limits in $\mathbb{C}$. This yields the result.

(P6.2) follows from Kadec's 1/4-Theorem (see, *e.g.*, Theorem 14 page 42 in Young, 2001) because $\sup_{j\in\mathbb{Z}}|\lambda_j - j| < 1/4$. □

Sets of uniqueness of analytic classes

Take $\mathcal{S} \subseteq \mathbb{R}$ nonempty, open, and connected. Due to the Weierstrass theorem (see, *e.g.*, Theorem 15.11 in Rudin, 1973), every $U \subseteq \mathcal{S}$ without an accumulation point in $\mathcal{S}$ is the set of zeros of a function in $\mathcal{A}(\mathcal{S})$. We give examples of real discrete sets of uniqueness by working with strict subclasses $\mathcal{A}_0(\mathcal{S})$. Denote by $M(x)$ the *trace function* of $\{M_m\}$, where, for all $m \in \mathbb{N}$, $M_m^c := \exp\left(\sup_{x\geq 0}(mx - M(x))\right)$ and $M(x) := \sup_{m\in\mathbb{N}}(mx - \log(M_m))$ (see Mandelbrojt, 1952).

Example A.1. ((2) in Section 1 of Hirschman, 1950) $U \subseteq \mathbb{R}$ is a set of uniqueness of $C^{\{m!\}}(\mathbb{R}, b)$ if

$$\overline{\lim}_{t\to\infty}\frac{\log\left(|U\cap(-t,t)|\right)}{t} > \frac{\pi b}{2}. \tag{2.33}$$

Recall that from Theorem 19.9 in Rudin (1973), $C^{\{m!\}}(\mathbb{R})$ consists of the functions f such that there exists $\rho > 0$ such that f can be extended uniquely to $\mathcal{H}^\infty\left(\{z \in \mathbb{C} : |\mathrm{Im}(z)| < \rho\}\right)$.

Example A.2. Let $m \in \mathbb{R}_+^{[0,\infty)}$, $m \to \infty$, and

$$\mathcal{A}_0(\mathbb{C}) := \left\{ f \in \mathcal{A}(\mathbb{C}) : \ f(0) = 0, \ \exists C > 0 : \ \forall r \geq 0, \ \max_{z:|z|=r}|f(z)| \leq Cm(r) \right\}.$$

U is a set of uniqueness of such a class $\mathcal{A}_0(\mathbb{C})$ if, for $\epsilon > 0$, we have

$$\exists \alpha > 1 : \ \underline{\lim}_{t\to\infty}\frac{\log(\alpha)}{\log(m(\alpha t))}\left|U\cap((-t,-\epsilon]\cup[\epsilon,t))\right| > 1. \tag{2.34}$$

A detailed argument proving this characterization is provided in the Appendix 2.5.5. The case where $m(r) = e^{\rho r}$ occurs when f is the Fourier transform of the difference of two complex measures such that $|\mu| \leq 1$ and supported in $[-\rho, \rho]$ (see (P3.3)) in which case $U = \mathbb{Z}h$ is a set of uniqueness if $\rho h/2 < 1$.

Sets of uniqueness of quasi-analytic classes

The following sets U are sets of uniqueness of the following V-quasi-analytic classes:
Example QA.1. $V \subseteq \mathring{U}$. Indeed, a function which is zero on U has all its partial derivatives or Laplacians (for functions on $\mathcal{S} \subseteq \mathbb{S}^p$) equal to zero at every point in V.
Example QA.2. (Lemma 4.8 in Belisle et al., 1997) $\mathcal{S}$ is an interval, $\mathring{\mathcal{S}} \neq \emptyset$, $V = \{\underline{\boldsymbol{x}}\} \subseteq U$ and U contains an accumulation point.
Example QA.3. (Theorem 4b in Hirschman, 1950) Let $M_m = \nu(m)^m m!$, where $x \in [0, \infty) \mapsto \nu(x)$ is increasing and continuously differentiable such that $\nu(0) = 1$ and

$$\lim_{x\to\infty} \frac{x\nu'(x)}{\nu(x)} = 0.$$

U is a set of uniqueness of $C^{\{M_m\}}(\mathbb{R}, b)$ (which contains non-analytic functions) if

$$\overline{\lim}_{t\to\infty} \frac{1}{t} \int_1^{|U\cap(-t,t)|} M(\log(r)) \frac{dr}{r^2} > \frac{\pi b}{2}. \tag{2.35}$$

If, for all $n \in \mathbb{N}$, $\nu(m) = 1$ for $0 \leq m < \exp^{*n}(1)$ and else $\nu(m) = (\log(m) \log^{*2}(m) \cdot \log^{*n}(m))^m$ (n^{th} logarithmic class, see Corollary 4c in Hirschman, 1950), U is a set of uniqueness of $C^{\{M_m\}}(\mathbb{R}, b)$ if

$$\overline{\lim}_{t\to\infty} \frac{\log^{*n+1}(|U \cap (-t,t)|)}{t} > \frac{\pi b}{2}.$$

Examples QA.1 and QA.2 are also sets of uniqueness of analytic classes on $\mathcal{S} \supseteq \mathbb{R}$ and we refer to Section 2.5.2 for more examples.

2.5.3 Quasi-analytic and analytic classes of Fourier transform of measures

Some of our identification strategies use Fourier transform of measures belonging to V-quasi-analytic and analytic classes. We give here integrability conditions that are sufficient and use them in Section 2.3.

Proposition 4. *Let $\mu \in \mathfrak{M}_c(\mathbb{R})$, $f = \mathcal{F}[\mu]$ can be extended to a complex analytic function that we also denote by f in the following cases:*

($P3.1$) *If there exists a function h such that $\int_{\mathbb{R}} e^{h(x)} d|\mu|(x) \le R < \infty$, then $f \in \mathcal{A}(\mathcal{S})$ for all open set $\mathcal{S} \subseteq \bigcap_{x \in \mathrm{supp}(\mu)} \{z \in \mathbb{C} : \ -Im(z)x \le h(x)\}$ and is bounded by R;*

($P3.2$) *If there exists an increasing function m such that, $m(0) = 0$ and, for all $r \in \mathbb{R}_+$, $\int_{\mathbb{R}} e^{r|x|} d|\mu|(x) \le m(r)$, then f belongs to*

$$\mathcal{A}_0(\mathbb{C}) := \left\{ f \in \mathcal{A}(\mathbb{C}) : \ f(0) = 0, \ \forall r \ge 0, \ \max_{z:|z|=r} |f(z)| \le m(r) \right\};$$

($P3.3$) *If, for $\rho \in \mathbb{R}_+$, $\mathrm{supp}(\mu) \subseteq \rho\overline{\mathbb{B}_{\mathbb{R}}}$, then f is an entire function and*

$$\exists C \in \mathbb{R}_+ : \ \forall z \in \mathbb{C}, \ |f(z)| \le C e^{\rho|\mathrm{Im}(z)|_2}, \tag{2.36}$$

where $C = |\mu|(\mathbb{R})$. Conversely, if f is an entire function in $\mathbb{C}$ and satisfies (2.36), then f is the extension of $\mathcal{F}[\mu]$ for some $\mu \in \mathfrak{M}_c(\mathbb{R})$ such that $\mathrm{supp}(\mu) \subseteq \rho\overline{\mathbb{B}_{\mathbb{R}}}$.

The above cases are ordered so that the assumptions on μ are increasingly demanding. (P3.1) can be applied when $\int_{\mathbb{R}} e^{\rho|x|} d|\mu|(x) \le R$ and $\mathcal{S} \subseteq \{z \in \mathbb{C} : \ |\mathrm{Im}(z)|_\infty < \rho\}$. In the above cases, $z \mapsto f(iz)$ and $z \mapsto f(-iz)$ are extensions of, respectively, the Laplace transform and moment generating function of μ. Proposition 4 and the injectivity of the Fourier transform imply that $\mu^* = \mu$ if $\mathcal{F}[\mu^*](u) = \mathcal{F}[\mu](u)$ for all u in a set of uniqueness U of the class of functions containing $\mathcal{F}[\mu^* - \mu]$. Also, in cases (P3.2) and (P3.3), f is characterized by $(f^m(0))_{m\in\mathbb{N}_0}$, hence μ is determinate in $\mathfrak{M}_c(\mathbb{R})$ by injectivity of the Fourier transform.

Proof of Proposition 4. Consider (P4.1). Use that there exists a complex Borel function g with $|g| = 1$ such that $d\mu = gd|\mu|$ and rewrite all integrals as integrals with respect to $|\mu|$. For all $x \in \text{supp}(\mu)$, $z \in \mathcal{S} \mapsto e^{izx}$ is holomorphic and, for all $z \in \mathcal{S}$, $x \in \mathbb{R} \mapsto e^{izx}$ is $|\mu|$-integrable. For all compact $K \subset \mathcal{S}$, for all $z \in K$, we have, for all $x \in \text{supp}(\mu)$, $|e^{izx}| \leq e^{h(x)}$ which is $|\mu|-$ integrable. The rest follows by the same argument as those of the proof of theorem p91 of Tauvel (2006) for complex variables. (P4.2) is obtained similarly.
(P4.3) is a corollary of the Paley-Wiener-Schwartz theorem using that complex measures are the distributions of order 0. □

2.5.4 The Denjoy-Carleman theorem

Theorem 10 reminds the link between condition (2.10) and quasi-analyticity. We work more generaly here with the convex regularisation of the sequences, which gives the following conditions, for $\{M_m\} \in \mathbb{R}^{\mathbb{N}_0}$,

(C.1) (Carleman's condition) $\sum_{m\in\mathbb{N}} 1/\beta_m = \infty$, where $\beta_m = \inf_{k\geq m} M_k^{1/k}$;

(C.2) $\sum_{m\in\mathbb{N}} 1/(M_m^c)^{1/m} = \infty$;

(C.3) (Mandelbrojt's condition) $\sum_{m\in\mathbb{N}} M_{m-1}^c/M_m^c = \infty$.

Theorem 10. *Let $\mathcal{S} \subseteq \mathbb{R}^p$ be a nonempty connected open set, $\underline{\boldsymbol{x}} \in \mathcal{S}$, then:*

T10.(*i*) *When $\{\boldsymbol{M}_m\} \in ((0,\infty]^p)^{\mathbb{N}_0}$, $C^{\{\boldsymbol{M}_m\}}(\mathcal{S})$ is $\{\underline{\boldsymbol{x}}\}-$quasi-analytic if the p coordinates sequences of $\{\boldsymbol{M}_m\}$ satisfy (C.1).*

T10.(*ii*) *When $\{M_m\}$ has positive elements, (C.1)-(C.3) are equivalent.*

T10.(*iii*) *When $\mathcal{S} = \mathbb{R}$ and $\{M_m\} \in \mathbb{R}_+^{\mathbb{N}_0}$, (C.1)-(C.3) are equivalent to the $\{\underline{\boldsymbol{x}}\}-$quasi-analyticity of $C^{\{M_m\}}(\mathcal{S})$.*

T10.(i) is Theorem B.1 in De Jeu (2004) extended to connected sets by classical arguments. Theorem T10.(ii) is given in Mandelbrojt (1952). T10.(iii) is the Denjoy-Carleman theorem (see, *e.g.*, Theorem 19.11 in Rudin, 1973). Due to Lemma 1, we can replace the index of the general term in the series in (C.1)-(C.3) by lm for all $l \in \mathbb{N}$.

2.5.5 Arguments for the characterization of set of uniqueness in Example A.2

The characterization in Example A.2 is a consequence of the fact that, if $f \in \mathcal{A}_0(\mathbb{C})$, then there exists $k \in \mathbb{N}$ such that $f(z) = z^k g(z)$ and $g(0) \neq 0$. Jensen formula applied to g (see, *e.g.* 15.18 in Rudin, 1973), with $\max_{z:|z|=r} |g(z)| = \max_{z:|z|=r} |f(z)|/z^k \leq Cm(r)/r^k$, ensures that

$$|g(0)| \prod_{k=1}^{n_g(\alpha r)} \frac{\alpha r}{\omega_k} \leq \frac{Cm(\alpha r)}{\alpha^k r^k},$$

where $\omega_1, \dots, \omega_{n_g(\alpha r)}$ are the zeros of g in $B(0, \alpha r)$ ranked according to their multiplicity and $n_g(\alpha r)$ is the number of zeros of g with multiplicity in $B(0, \alpha r)$. Thus, we have

$$\frac{Cm(\alpha r)}{\alpha^k r^k} \geq |g(0)| \prod_{k=1}^{n_g(r)} \frac{\alpha r}{\omega_k} \geq |g(0)| \, \alpha^{n_g(r)}$$

which yields that $n_f(r) = k + n_g(r) \leq k + \log(Cm(\alpha r)(\alpha r)^{-k}/\,|g(0)|)/\log(\alpha)$. Thus, denoting by $\overline{n}_f(r)$ the number of zeros of f with multiplicity in $B(0, r) \setminus \{\mathbf{0}_c\}$, we have

$$\frac{\overline{n}_f(r)\log(\alpha)}{\log(m(\alpha r))} \leq 1 + \frac{\log(C(\alpha r)^{-k}/\,|g(0)|)}{\log(m(\alpha r))}$$

then $\underline{\lim}_{r\to\infty} \overline{n}_f(r)\log(\alpha)/\log(m(\alpha r)) \leq 1$. This yields that U satisfying (2.34) is a set of uniqueness of $\mathcal{A}_0(\mathbb{C})$.

2.5.6 $\mathcal{A}(\mathcal{S})$ is a small subset of $C^\infty(\mathcal{S})$

Proposition 5 can be found in Schmets and Valdivia (1991), it is of the same spirit as the statement that the complement of nowhere analytic functions in $C^\infty(\mathcal{S})$ is meager with empty interior, which we detail below.

Proposition 5. *If a condition (C.1)-(C.3) fails, then $C^{\{M_m\}}(\mathcal{S})$ contains a vector space of dimension $2^{\aleph_0}$ such that for all nonzero element f there does not exists $x \in \mathcal{S}$, a neighborhood $\mathcal{N}_x$ of x, and a sequence $\{M_{x,m}\}$ such that f belongs to a quasi-analytic class $C^{\{M_{x,m}\}}(\mathcal{N}_x)$.*

Assume, for simplicity, that $\mathcal{S} \subset \mathbb{R}^p$ is compact and as usual $C^\infty(\mathcal{S})$ is equipped with the distance $d(f,g) := \sum_{\boldsymbol{m}\in\mathbb{N}_0^p} \min\left(1/\mathbf{2}_c^{\boldsymbol{m}}, \left\|f^{(\boldsymbol{m})} - g^{(\boldsymbol{m})}\right\|_{L^\infty(\mathcal{S})}\right)$. By the Cauchy-Hadamard theorem (see Shabat, 1992), $f \in C^\infty(\mathcal{S})$ is analytic at $\underline{\boldsymbol{x}}$ if and only if there exist $b, c \in \mathbb{N}$ such that f belongs to

$$T(\underline{\boldsymbol{x}}, b, c) := \left\{f \in C^\infty(\mathcal{S}) : \ \forall \boldsymbol{m} \in \mathbb{N}_0^p, \ \left|f^{(\boldsymbol{m})}(\underline{\boldsymbol{x}})\right| \le c b^{|\boldsymbol{m}|}\boldsymbol{m}!\right\},$$

thus $\mathcal{A}(\mathcal{S}) = \bigcup_{\underline{\boldsymbol{x}}\in\mathbb{Q}^p\cap\mathcal{S}, b,c\in\mathbb{N}} T(\underline{\boldsymbol{x}}, b, c)$. This is a countable union of closed sets, hence closed. Moreover, $T(\underline{\boldsymbol{x}}, b, c)$ has empty interior. Indeed (see Salzmann and Zeller, 1955), given f in $T(\underline{\boldsymbol{x}}, b, c)$, for all $\epsilon > 0$, $m \in \mathbb{N}$ such that $\sum_{j=m}^{\infty} 2^{-j} < \epsilon$, $c < (\epsilon b^m/(2m!))^{1/(2m)}$, the functions $f_\epsilon : \ \boldsymbol{x} \in \mathcal{S} \mapsto f(\boldsymbol{x}) + \epsilon \cos\left(c(\boldsymbol{x}_1 - \underline{\boldsymbol{x}}_1)\right)/(2c^m)$ are such that $d(f_\epsilon, f) < \epsilon$ and $\left|f_\epsilon^{(2m)}(\underline{\boldsymbol{x}}) - f^{(2m)}(\underline{\boldsymbol{x}})\right| > b^{2m}(2m)!$, hence $f_\epsilon \notin T(\underline{\boldsymbol{x}}, b, c)$. Due to Baire's theorem, the meager set $\mathcal{A}(\mathcal{S})$ of the complete metric space $C^\infty(\mathcal{S})$ has an empty interior. With the arguments in Cater (1984), the complement of nowhere analytic functions in $C^\infty(\mathcal{S})$, hence containing $\mathcal{A}(\mathcal{S})$, can be shown to be meager with empty interior.

2.5.7 Complementary results

On the linear panel data model where regressors are monomials of a baseline scalar regressor

We give here results about identification under restrictions on the tails or the sign of the marginals of $\mathbb{P}^*_\gamma$. We remind $\mathcal{X}_0 := \{\boldsymbol{x} \in \mathbb{R}^T : \ \prod_{j=1}^T \boldsymbol{x}_j \neq 0, \ \prod_{m\neq j}(\boldsymbol{x}_m - \boldsymbol{x}_j) \neq 0\}$.

Assumption 11. *Given $\{M_m\}$, $\{\boldsymbol{M}_{k,m}\}_{k=1}^{T+1}$ log-convex sequences which satisfy $M_0 =$*

1 *and* $\boldsymbol{M}_{k,0} = 1$*, define*

$$\mathcal{R}_{LP,T}\left(\{M_m\}, \{\boldsymbol{M}_{k,m}\}_{k=1}^{T+1}\right) := \mathcal{R}_{LP,0}(\{M_m\}) \cap \left\{\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathfrak{M}_1\left(\mathbb{R}^{3T+1}\right) : \ R_{LP} \ (i)\text{-}(ii)\right\},$$

R_{LP} *(i)* *There exists a set* $V \subset \mathbb{R}^T$ *with non-empty interior such that for all* $\boldsymbol{v} \in V$*,* $U_{T,\boldsymbol{v}} = \{\Theta(\boldsymbol{v}, \boldsymbol{u}), \ \forall \boldsymbol{u} \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0\}$*, where* Θ *is defined in* (2.31)*, are sets of uniqueness of* $C^{\{\boldsymbol{M}_{T+1,m}\}}(\mathbb{R})$*;*

R_{LP} *(ii)* *For all* $k = 1, \dots, T+1$ *and* $m \in \mathbb{N}_0$*,* $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\gamma}}}[|\boldsymbol{\gamma}_k^\top|^m] \le \boldsymbol{M}_{k,m}$ *and, if* $\mathbb{S}_{\boldsymbol{\gamma}_k} \subseteq \mathbb{R}_+$*,* $\{\boldsymbol{M}_{k,m}\}$ *satisfies* (2.11) *and else, if* $\mathbb{S}_{\boldsymbol{\gamma}_k} \not\subseteq \mathbb{R}_+$*,* (2.10) *otherwise.*

Theorem 11. *Let* $\{M_m\}$*,* $\{\boldsymbol{M}_{k,m}\}_{k=1}^{T+1}$ *be log-convex sequences which satisfy* $M_0 = 1$ *and* $\boldsymbol{M}_{k,0} = 1$*. Then,* $\mathbb{P}^*_{\boldsymbol{\gamma}}$ *is identified under* $\mathcal{R}_{LP,T}(\{M_m\}, \{\boldsymbol{M}_{k,m}\}_{k=1}^{T+1})$*.*

Proofs of Theorem 11. Let us start with the sketch of the proof. We first show that under $\mathcal{R}_{LP,0}(\{M_m\})$ if $\{M_m\}$ is a log-convex sequence which satisfies $M_0 = 1$ and (C.3), then, for all $\boldsymbol{t} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$,

$$\varphi^*_{\alpha,\boldsymbol{\beta}}\left(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \dots, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j^T\right) \tag{2.37}$$

is identified. Then, for all $\boldsymbol{x} \in \mathcal{X}_0$, we use a change of variable that relates $\boldsymbol{t} \in \mathbb{R}^T$ to $\boldsymbol{v} \in \mathbb{R}^T$ such that $(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \dots, \sum_{j=1}^{T-1} \boldsymbol{t}_j \boldsymbol{x}_j^{T-1}) = \boldsymbol{v}$. This change of variable allows to choose the values of the $T-1$ first variables in (2.37) independently from each other. This can be written as $\boldsymbol{t} = \left(V^{-1}(\boldsymbol{x})\right)^\top \boldsymbol{v}$, where

$$V(\boldsymbol{x}) := \begin{pmatrix} 1 & \boldsymbol{x}_1 & \boldsymbol{x}_1^2 & \dots & \boldsymbol{x}_1^{T-1} \\ \vdots & & & \dots & \vdots \\ 1 & \boldsymbol{x}_T & \boldsymbol{x}_T^2 & \dots & \boldsymbol{x}_T^{T-1} \end{pmatrix}$$

is the Vandermonde matrix. We use $D(\boldsymbol{x})$, the diagonal matrix which entries are the coordinates of $\boldsymbol{x}$. It is a classical result that, for all $\boldsymbol{x} \in \mathcal{X}_0$, $D(\boldsymbol{x})$ and $V(\boldsymbol{x})$ are invertible hence $\widetilde{V}(\boldsymbol{x}) := \left(D(\boldsymbol{x})V(\boldsymbol{x})\right)^\top$ is invertible. Then, for all $\boldsymbol{v} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathcal{X}_0$, we can express $\boldsymbol{t}$ as a function of $\boldsymbol{v}$ and $\boldsymbol{x}$, and obtain for the last variable in

(2.37)

$$\sum_{j=1}^{T} \boldsymbol{t}_j \boldsymbol{x}_j^T = \boldsymbol{t}^\top \boldsymbol{x}^T = \left(\left(V(\boldsymbol{x})^{-1}\right)^\top \boldsymbol{v}\right)^\top \boldsymbol{x}^T = \left(D(\boldsymbol{x})\widetilde{V}(\boldsymbol{x})^{-1}\boldsymbol{v}\right)^\top \boldsymbol{x}^T = \Theta(\boldsymbol{v}, \boldsymbol{x}). \quad (2.38)$$

Thus, for all $\boldsymbol{v} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0$, $\varphi^*_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, \Theta(\boldsymbol{v}, \boldsymbol{x}))$ is identified. Assuming that, for all $\boldsymbol{v} \in \mathbb{R}^T \setminus \mathcal{V}$, where $\mathbb{R}^T \setminus \mathcal{V}$ is dense in $\mathbb{R}^T$, $x \in \mathbb{R} \mapsto \varphi_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, x)$ belongs to V-quasi-analytic classes for which $U_{\boldsymbol{v}}$ is a set of uniqueness yields identification of $\varphi^*_{\alpha,\boldsymbol{\beta}}$.

Consider $\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}} \in \mathcal{R}_{LP,T}(\{M_m\}, \{\boldsymbol{M}_{k,m}\}_{k=1}^{T+1})$ such that $\mathbb{P}\left(\mathbb{P}_{\boldsymbol{\gamma},\boldsymbol{X}}\right) = \mathbb{P}\left(\mathbb{P}^*_{\boldsymbol{\gamma}} \otimes \mathbb{P}_{\boldsymbol{X}}\right)$. Using $R_{LP,0}$ (ii), $r > 0$ and δ defined in (2.20), and (2.20), and following the same steps as in the proof of Theorem 7 yields $\mathbb{P}_{\boldsymbol{\epsilon}_1} = \mathbb{P}^*_{\boldsymbol{\epsilon}_1}$.

Now, for all $t \in \mathbb{R}$, $\varphi_{\boldsymbol{Y}_1|\boldsymbol{X}_1}(t|r) = \varphi_\delta(t)\varphi^*_{\boldsymbol{\epsilon}_1}(t) = \varphi^*_\delta(t)\varphi^*_{\boldsymbol{\epsilon}_1}(t)$, hence, because the zeros of $\varphi^*_{\boldsymbol{\epsilon}_1}$ are isolated (see Lemma 4.8 in Belisle et al., 1997) and φ_δ and φ^*_δ are continuous, we obtain $\varphi_\delta = \varphi^*_\delta$. Similarly, because, for all $t \in \mathbb{R}$ and $j = 2, \ldots, T$, $\varphi_{\boldsymbol{Y}_j - \boldsymbol{Y}_1|\boldsymbol{X}_1,\boldsymbol{X}_j}(t|(r,r)) = \varphi^*_{\boldsymbol{\epsilon}_j}(t)\varphi^*_{\boldsymbol{\epsilon}_1}(-t) = \varphi_{\boldsymbol{\epsilon}_j}(t)\varphi^*_{\boldsymbol{\epsilon}_1}(-t)$, the zeros of $\varphi^*_{\boldsymbol{\epsilon}_1}$ are isolated and $\varphi_{\boldsymbol{\epsilon}_j}$ and $\varphi^*_{\boldsymbol{\epsilon}_j}$ are continuous, we obtain $\varphi_{\boldsymbol{\epsilon}_j} = \varphi^*_{\boldsymbol{\epsilon}_j}$ for all $j = 2, \ldots, T$.
Using that, for all $\boldsymbol{t} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$, we have

$$\begin{aligned}
&\varphi_{\alpha,\boldsymbol{\beta}}\left(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \ldots, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j^T\right) \prod_{j=1}^T \varphi_{\boldsymbol{\epsilon}_j}(\boldsymbol{t}_j) \\
&= \varphi^*_{\alpha,\boldsymbol{\beta}}\left(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \ldots, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j^T\right) \prod_{j=1}^T \varphi_{\boldsymbol{\epsilon}_j}(\boldsymbol{t}_j)
\end{aligned}$$

that, for all $j = 1, \ldots, T$, the zeros of $\varphi_{\boldsymbol{\epsilon}_j}$ are isolated, and that $\varphi_{\alpha,\boldsymbol{\beta}}$ and $\varphi^*_{\alpha,\boldsymbol{\beta}}$ are continuous, yield, for all $\boldsymbol{t} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$,

$$\varphi_{\alpha,\boldsymbol{\beta}}\left(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \ldots, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j^T\right) = \varphi^*_{\alpha,\boldsymbol{\beta}}\left(\sum_{j=1}^T \boldsymbol{t}_j, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j, \ldots, \sum_{j=1}^T \boldsymbol{t}_j \boldsymbol{x}_j^T\right).$$

Now, we make use of the change of variables $\boldsymbol{t} = (V^{-1}(\boldsymbol{x}))^\top \boldsymbol{v}$. Thus, using (2.38), (2.32) holds, for all $\boldsymbol{v} \in \mathbb{R}^T$ and $\boldsymbol{x} \in \mathcal{X}_0$. Using using arguments from the proof of Theorem 4, with R_{LP} (ii), and Proposition 1 ensure that, for all $\boldsymbol{v} \in V$, $z \in \mathbb{R} \mapsto$

$(\varphi_{\alpha,\boldsymbol{\beta}} - \varphi^*_{\alpha,\boldsymbol{\beta}})(\boldsymbol{v}, z)$ can be extended uniquely to $\mathcal{C}^{\{\boldsymbol{M}_{T+1,m}\}}(\mathbb{R}) \cap \{\mathcal{F}[\mu],\ \mu \in \mathfrak{M}_c^*(\mathbb{R})\}$ which is a quasi-analytic class, and is zero on a set of uniqueness of this class, thus on $\mathbb{R}$. Then, R_{LP} (i) yields, for all $(\boldsymbol{v}, z) \in V \times \mathbb{R}$, $\varphi_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, z) = \varphi^*_{\alpha,\boldsymbol{\beta}}(\boldsymbol{v}, z)$. Take $\boldsymbol{v} \in V$ and $t \in \mathbb{R}$, R_{LP} (ii) and Proposition 1 ensure that, $z \in \mathbb{R} \mapsto (\varphi_{\alpha,\boldsymbol{\beta}} - \varphi^*_{\alpha,\boldsymbol{\beta}})(\boldsymbol{v}_{[T-1]}, z, t)$ can be extended uniquely to $\mathcal{C}^{\{\boldsymbol{M}_{T,m}\}}(\mathbb{R}) \cap \{\mathcal{F}[\mu],\ \mu \in \mathfrak{M}_c^*(\mathbb{R})\}$ which is a quasi-analytic class, and is zero on a set of uniqueness of this class, thus on $\mathbb{R}$. Iterating this procedure we obtain that $\mathbb{P}_{\alpha,\boldsymbol{\beta}} = \mathbb{P}^*_{\alpha,\boldsymbol{\beta}}$. □

Condition R_{LP} (i) excludes the case where $\boldsymbol{X}_1 = \boldsymbol{X}_2$ (which would yield empty $U_{T,\boldsymbol{v}}$), which does not provide enough variation in (2.19) to recover uniquely the joint distribution of the coefficients.

Proposition 6. *The following conditions are sufficient for R_{LP} when $T = 2$:*

1. *There exists $a \in \mathbb{S}_{\boldsymbol{X}_2}$ such that $\{\boldsymbol{x}_1 \in \mathbb{R},\ \boldsymbol{x} = (\boldsymbol{x}_1, a) \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0\}$ satisfies Example QA.2 or QA.3 in Section 2.5.2;*

2. *There exists $a \in \mathbb{S}_{\boldsymbol{X}_2}$ and $\rho, R > 0$ such that $\{\boldsymbol{x}_1 \in \mathbb{R},\ \boldsymbol{x} = (\boldsymbol{x}_1, a) \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0\}$ satisfies Example A.2 in Section 2.5.2 and $\mathbb{E}\left[e^{\rho|\boldsymbol{\beta}_2|}\right] \le R$.*

Proof of Proposition 6. Consider $\mathcal{V} = \{\boldsymbol{v} :\ \boldsymbol{v}_2 = \boldsymbol{v}_1 a\}$, $U_{T,\boldsymbol{v}} := \{\boldsymbol{x}_1(\boldsymbol{v}_2 - \boldsymbol{v}_1 a) + a\boldsymbol{v}_2 :\ \boldsymbol{x} = (\boldsymbol{x}_1, a) \in \mathbb{S}_{\boldsymbol{X}} \cap \mathcal{X}_0\}$. Start by proving (1). Then, for all $\boldsymbol{v} \in \mathbb{R}^2 \setminus \mathcal{V}$ such that $(\boldsymbol{v}_2 - \boldsymbol{v}_1 a) \ge 1$, $U_{T,\boldsymbol{v}}$ also satisfies QA.2 and QA.3, hence is a set of uniqueness of $C^{\{\boldsymbol{M}_{T+1,m}\}}(\mathbb{R})$.

We now prove (2). For all $\boldsymbol{v} \in \mathbb{R}^2 \setminus \mathcal{V}$ such that $(\boldsymbol{v}_2 - \boldsymbol{v}_1 a) \ge 1$, $U_{T,\boldsymbol{v}}$ also satisfies A.1 and is a set of uniqueness of $\mathcal{H}^\infty(\{\boldsymbol{z} \in \mathbb{C} :\ |\mathrm{Im}(\boldsymbol{z})| < \rho\})$ which belongs to $C^{\{m!\}}(\mathbb{R}^p, 1/\rho)$. Using Proposition 4 and $\mathbb{E}\left[e^{\rho|\boldsymbol{\beta}_2|}\right] \le R$, we have that $z \mapsto \varphi_{\boldsymbol{\gamma}'}(\boldsymbol{v}, z)$ belongs to this class. □

Chapter 3

Adaptive Estimation in the Linear Random Coefficients Model when Regressors have Limited Variation

Joint with Eric Gautier, Toulouse School of Economics.

Preprint available here.

Associated R package (RandomCoefficients) and vignette available here.

Abstract

We consider a linear model where the coefficients - intercept and slopes - are random and independent from the regressors. The law of the coefficients is nonparametric. Without further restriction, nonparametric identification requires the regressors to have a support which is the whole space. This is hardly ever the case in practice. It is possible to handle regressors with limited variation when the coefficients can have a compact support. This is not compatible with unbounded error terms as usual in regression models. In this paper, the regressors can have a support which is a proper subset but the slopes do not have heavy-tails. Lower bounds on the minimax risk for the estimation of the joint density of the random coefficients are obtained for a wide range of smoothness. Some allow for polynomial and nearly parametric rates of convergence. We present a minimax optimal estimator and a data-driven rule for adaptive estimation. A R package is available to implement this estimator.

Keywords: Adaptation, Ill-posed Inverse Problem, Minimax, Random Coefficients.

3.1 Introduction

Inferring causal effects from a data set is of great importance for applied researchers. This paper assumes that the explanatory variables are determined outside the model (*e.g.*, a treatment is randomly assigned) and addresses the question of the heterogeneity of the effects. We can argue that the effects are heterogeneous in nearly all applications of the linear regression. In such a case, the coefficients of the linear regression capture average effects. For example, it is well understood that the effect of the income of the parents or the class size on pupils' achievements differ across pupils. A second example consists of models of consumer demand such as Engel curves. These are models of the effect of the total budget an household can spend on consumption goods on the budget share spent on a particular one (*e.g.*, food). It is also well known that there is a great deal of heterogeneity in consumer demand (see, *e.g.*, Hoderlein et al. (2010)). Understanding the nature of the heterogeneity of the effects can have useful policy implications. For example, it could be worthwhile to implement less costly targeted measures. It is thus of interest to recover the distribution of heterogenous effects. Based on this model, we can compute prediction intervals for an outcome (see Beran (1995)), welfare measures, and counterfactual effects.

The linear regression with random coefficients is a continuous mixture of linear regressions. Maintaining parametric assumptions on the mixture density is open to criticism because these assumptions can drive the results (see Heckman and Singer (1984)). For this reason, this paper considers a nonparametric setup. Unfortunately, most of the estimation theory for this model has relied on assumptions which are almost never satisfied. This is probably the reason why, up to now, applied researchers have preferred to use the quantile regression to account for heterogenous effects. There, the conditional quantiles of an outcome given the regressors are linear in the regressors. When the conditional quantiles are strictly increasing then the quantile regression defines the same data generating process as a linear random coefficients model where the coefficients are functions of a scalar uniform distribution. However, it can be hard to argue for such degeneracy of the random coefficients distribution or for the linearity of the conditional quantiles. The unobserved scalar uniform variable can be interpreted as a ranking variable. In the education and demand examples, one can argue that such ranking can be based on a more complex multidimensional vector of unobserved attributes. Restricting unobserved heterogeneity to a scalar can have undesirable implications such as monotonicity in the literature on policy evaluation (see Gautier and Hoderlein (2015)).

Formally, for a random variable α and random vectors $\boldsymbol{X}$ and $\boldsymbol{\beta}$ of dimension p, the linear random coefficients model is

$$Y = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}, \tag{3.1}$$

$$(\alpha, \boldsymbol{\beta}^\top) \text{ and } \boldsymbol{X} \text{ are independent.} \tag{3.2}$$

The researcher has at her disposal n observations $(Y_i, \boldsymbol{X}_i^\top)_{i=1}^n$ of $(Y, \boldsymbol{X}^\top)$ but does not observe the realizations $(\alpha_i, \boldsymbol{\beta}_i^\top)_{i=1}^n$ of $(\alpha, \boldsymbol{\beta}^\top)$. α subsumes the intercept and error term and the vector of slope coefficients $\boldsymbol{\beta}$ is heterogeneous (*i.e.*, varies across i). $(\alpha, \boldsymbol{\beta}^\top)$ corresponds to multidimensional unobserved heterogeneity and $\boldsymbol{X}$ to observed heterogeneity. Other random coefficients models have been analyzed recently in econometrics (see, *e.g.*, Breunig and Hoderlein (2018); Gautier and Kitamura (2013); Hoderlein et al. (2017); Masten (2017) and references therein). This paper focuses on the most simple model but addresses important practical issues.

Estimation of the density of the random coefficients $f_{\alpha,\boldsymbol{\beta}}$ has similarities with tomography problems involving the Radon transform (see Beran et al. (1996*b*); Beran and Millar (1994); Hoderlein et al. (2010)). Indeed, the density of $Y/\sqrt{1+|\boldsymbol{X}|_2^2}$ given $\boldsymbol{S} = (1, \boldsymbol{X}^\top)^\top/\sqrt{1+|\boldsymbol{X}|_2^2}$, where $|\cdot|_2$ is the Euclidian norm, at point u given $\boldsymbol{s}$ is the integral of $f_{\alpha,\boldsymbol{\beta}}$ on the affine hyperplane defined via the pair $(u, \boldsymbol{s})$. It could be tempting to analyze the related tomography problem with an additive Gaussian white noise. But the random coefficients model (3.1)-(3.2) has it specificities. Treating it requires: (1) that $(\alpha, \boldsymbol{\beta}^\top)$ has a noncompact support to allow for usual unbounded errors, (2) to allow the dimension to be larger than in tomography due to more than one or two regressors, and (3) the directions to have an unknown but estimable density.

To obtain rates of convergence, Hoderlein et al. (2010) assumes the density of $\boldsymbol{S}$ is bounded from below. When $p = 1$, this holds when $\boldsymbol{X}$ has tails at least as fat as the ones of the Cauchy distribution. Recently, Dunker et al. (2017) motivates testing large features of the density by the possible slow rates of convergence of density estimation and Holzmann and Meister (2019) obtains rates of convergence for density estimation with less heavy tails on $\boldsymbol{X}$. But assuming the support of $\boldsymbol{X}$ is $\mathbb{R}^p$ is unrealistic for nearly all applications. In the motivating examples, the income of the parents, the class size, and the total budget have limited variation.

Limited angle tomography considers the case where $p = 1$, $\boldsymbol{S}$ has a support which is a known cap (*i.e.*, the support of the angle is an interval), and the object has support in a ball. Frikel (2013) proposes a soft-thresholded curvelet regularization for the problem with an additive bounded noise but does not obtain results for the statistical problem (*e.g.*, consistency). Importantly, Hohmann and Holzmann (2015) shows that the rate of the minimax risk in Sobolev type ellipsoids relative to the right-singular functions of the Radon transform is logarithmic. It shows that projection estimators are adaptive. It gives the analogy with a random coefficients model where $p = 1$, $(\alpha, \boldsymbol{\beta}^\top)$ has support in the unit ball, for some known densities of the regressors. It concludes that a lot remains to be done to handle $p > 1$ and estimable densities of the regressors. The random coefficients model when the support of $\boldsymbol{X}$ can be a proper (*i.e.* strict) subset is considered in Beran and Millar (1994). When $p = 1$ and $(\alpha, \boldsymbol{\beta}^\top)$ has compact support, it is shown that a minimum distance estimator is consistent. Section 2 in the online appendix of Hoderlein et al. (2017) proposes another consistent estimator in this case.

This paper is directly applicable to (3.1)-(3.2). It allows for the essential feature that α can have a noncompact support, that the researcher does not need to have knowledge on the support of $\boldsymbol{\beta}$ if compact, and that the latter can also be noncompact. It also allows for estimable density of the regressors and arbitrary p. We assume the marginals of $\boldsymbol{\beta}$ (but not of α) do not have heavy tails. This allows for many parametric families which are used in mixture modelling, while leaving unspecified the parametric family. We do not rely on the Radon transform but on the truncated Fourier transform (see, *e.g.*, Alibaud et al. (2009)). Due to (3.2), the conditional characteristic function of Y given $\boldsymbol{X} = \boldsymbol{x}$ at t is the Fourier transform of $f_{\alpha,\boldsymbol{\beta}}$ at $(t, t\boldsymbol{x}^\top)^\top$. Hence, the family of conditional characteristic functions indexed by $\boldsymbol{x}$ in the support of $\boldsymbol{X}$ gives access to the Fourier transform of $f_{\alpha,\boldsymbol{\beta}}$ on a double cone of axis $(1, 0, \ldots, 0) \in \mathbb{R}^{p+1}$ and apex 0. When $\alpha = 0$ and the supports of $\boldsymbol{\beta}$ and $\boldsymbol{X}$ are compact with nonempty interior, this is the problem of out-of-band extrapolation or super-resolution (see, *e.g.*, Bertero and Boccacci (1998)). Because we do not restrict α and the support of $\boldsymbol{\beta}$ can be noncompact, we generalize this approach.

A related problem is extrapolation. It is used in Meister (2007) to perform deconvolution of compactly supported densities, allowing the characteristic function of the error to vanish. This paper does not use extrapolation nor assume densities are analytic. Rather, the operator of the inverse problem is a composition of two operators based on partial Fourier transforms. One involves a truncated Fourier transform and we make use of properties of the singular value decomposition.

Unlike Beran and Millar (1994); Hoderlein et al. (2017), we go beyond consistency and provide a full analysis of the general case. Similar to Gautier and Le Pennec (2018); Hohmann and Holzmann (2015); Holzmann and Meister (2019), we study minimax optimality. However, we obtain lower bounds under a wide variety of assumptions. We show that polynomial and nearly parametric rates can be attained. Hence, we can lose little in terms of rates of convergence from going from a parametric model to a nonparametric one. This contrasts with the pessimistic logarithmic rates in Hohmann and Holzmann (2015) (also mentioned in Hoderlein et al. (2017)) and the message to avoid estimating densities in Dunker et al. (2017). We present an estimator involving: series based estimation of the partial Fourier transform of the density with respect to the first variable, interpolation around zero, and inversion of the partial Fourier transform. The orthonormal system is a tensor products of the Prolate Spheroidal Wave Functions (henceforth PSWF, see Osipov et al. (2013)) when the law of $\boldsymbol{\beta}$ has a support included in a known bounded set. Else, it is composed of singular functions of an operator first studied in Morrison (1962); Widom (1964). The relevant results on these systems are given in the appendix. They can also be used in a wide range of applications such as for stable analytic continuation by Hilbert space techniques (see Gaillac and Gautier (2019*a*)). We use a Goldenshluger-Lepski type method to obtain data-driven estimators. We consider estimation of the marginal $f_{\boldsymbol{\beta}}$ in Appendix 3.7.3.

The adaptive estimator is implemented in the R package RandomCoefficients. In contrast with the EM algorithm for parametric continuous mixtures of regressions, the estimator has the advantage of being robust to misspecification of the parametric family. It also avoids possible non convergence issues of the EM algorithm. The proposed estimator relies on the computation of a SVD which we obtain once and for all by a numerically efficient method, on two simple sums, and a one dimensional Fourier transform carried by FFT. Additional practical and computational details are available in Gaillac and Gautier (2019*d*).

The paper is organized as follows. Section 3.2 gives preliminaries. It introduces the baseline assumption on $f_{\alpha,\boldsymbol{\beta}}$ and the variation of the regressors. It frames the recovery of $f_{\alpha,\boldsymbol{\beta}}$ as an inverse problem involving a composition at the basis of the estimation procedure. The section also introduces a related Gaussian white noise model and the main elements on interpolation, which is key to obtain an optimal estimator. Finally, it presents the sets of smooth and integrable functions for the minimax analysis and the risk. Section 3.3 provides the lower bounds. Section 3.4 describes the estimator and its rates of convergence assuming knowledge of the parameters of the sets of functions. Section 3.4.3 provides a data-driven estimator and presents its nearly minimax rates of convergence. Section 3.5 concludes with details on the numerical implementation and simulations illustrating the finite sample performances of the data-driven estimator.

3.2 Preliminaries

3.2.1 Notations

The notations $\cdot$, $\cdot_1$, $\cdot_2$, $\star$ are used to denote variables in a function. $a \wedge b$ (resp. $a \vee b$) is used for the minimum (resp. maximum) between a and b, $(\cdot)_+$ for $0 \vee \cdot$, and $\mathbb{1}\{A\}$ for the indicator function of set A. $\mathbb{N}$ and $\mathbb{N}_0$ stand for the positive and nonnegative integers. Bold letters are used for vectors. For all $r \in \mathbb{R}$, $\underline{\boldsymbol{r}}$ is the vector, which dimension will be clear from the text, where each entry is r. For $x \geq 1$ we denote by $\ln_2(x) = \ln(\ln(x))$. $\mathcal{W}$ is the inverse of $x \in [0, \infty) \mapsto xe^x$. $|\cdot|_q$ for $q \in [1, \infty]$ stands for the ℓ_q norm of a vector or sequence. For all $\boldsymbol{\beta} \in \mathbb{C}^d$, $(f_m)_{m \in \mathbb{N}_0}$ functions with values in $\mathbb{C}$, and $\boldsymbol{m} \in \mathbb{N}_0^d$, denote by $\boldsymbol{\beta}^{\boldsymbol{m}} = \prod_{k=1}^d \boldsymbol{\beta}_k^{\boldsymbol{m}_k}$, $|\boldsymbol{\beta}|^{\boldsymbol{m}} = \prod_{k=1}^d |\boldsymbol{\beta}_k|^{\boldsymbol{m}_k}$, and $f_{\boldsymbol{m}} = \prod_{k=1}^d f_{\boldsymbol{m}_k}$. For a function f of real variables, $\mathrm{supp}(f)$ denotes its support. The inverse of a mapping f, when it exists, is denoted by f^I. We denote the interior of $\mathcal{S} \subseteq \mathbb{R}^d$ by $\mathring{\mathcal{S}}$.

When $\mathcal{S}$ is a Borel set and h a nonnegative function from $\mathcal{S}$ to $[0, \infty]$, $L^2(h)$ is the space of complex-valued square integrable functions equipped with $\langle f, g \rangle_{L^2(h)} = \int_{\mathcal{S}} f(\boldsymbol{x})\overline{g}(\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x}$. This is denoted by $L^2(\mathcal{S})$ when $h = 1$. For a Borel set $\mathcal{S} \subseteq \mathbb{R}^d$, we denote by $i_{\mathcal{S}} = \mathbb{1}\{\mathcal{S}\} + \infty\ \mathbb{1}\{\mathcal{S}^c\}$. We have $L^2(i_{\mathcal{S}}) = \{f \in L^2(\mathbb{R}^d) :\ \mathrm{supp}(f) \subseteq \mathcal{S}\}$ and $\langle f, g \rangle_{L^2(i_{\mathcal{S}})} = \int_{\mathcal{S}} f(\boldsymbol{x})\overline{g}(\boldsymbol{x})d\boldsymbol{x}$. Denote by $\mathcal{D}$ the set of densities and by $\otimes$ the product of functions (*e.g.*, $h^{\otimes d}(\boldsymbol{b}) = \prod_{j=1}^d h(\boldsymbol{b}_j)$) or measures.

The Fourier transform of $f \in L^1\left(\mathbb{R}^d\right)$ is $\mathcal{F}[f](\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{b}^\top \boldsymbol{x}} f(\boldsymbol{b})d\boldsymbol{b}$ and $\mathcal{F}[f]$ is also the Fourier transform in $L^2\left(\mathbb{R}^d\right)$. For all $c > 0$, denote the Paley-Wiener space by $PW(c) := \{f \in L^2(\mathbb{R}) : \mathrm{supp}\left(\mathcal{F}[f]\right) \subseteq [-c, c]\}$, by $\mathcal{P}_c$ the projector from $L^2(\mathbb{R})$ to $PW(c)$ ($\mathcal{P}_c[f] = \mathcal{F}^I\left[\mathbb{1}\{[-c, c]\}\mathcal{F}[f]\right]$). For all $f \in L^1\left(\mathbb{R}^d\right)$, $\mathcal{F}_{1\mathrm{st}}[f](t, \cdot_2)$ denotes the partial Fourier transform of f with respect to the first variable.

Now on, W is a function which can be either $i_{[-R,R]}$ or $\cosh(\cdot/R)$ and which depends on a given $R > 0$. The truncated Fourier transform $\mathcal{F}_c$ is defined by

$$\forall c \neq 0, \quad \mathcal{F}_c : \begin{array}{ccc} L^2\left(W^{\otimes d}\right) & \rightarrow & L^2\left([-1,1]^d\right) \\ f & \rightarrow & \mathcal{F}\left[f\right](c\ \cdot). \end{array} \tag{3.3}$$

For a random vector $\boldsymbol{X}$, $\mathbb{P}_{\boldsymbol{X}}$ is its law, $f_{\boldsymbol{X}}$ its density, $f_{\boldsymbol{X}|\mathcal{X}}$ the truncated density of $\boldsymbol{X}$ given $\boldsymbol{X} \in \mathcal{X}$, $\mathbb{S}_{\boldsymbol{X}}$ its support, and $f_{Y|\boldsymbol{X}=\boldsymbol{x}}$ the conditional density.

3.2.2 Baseline assumption

Assumption 1 below imposes restrictions on the integrability of the density of the random coefficients and on the variation of the regressors ensuring nonparametric identification.

Assumption 1. *(H1.1) $f_{\boldsymbol{X}}$ and $f_{\alpha,\boldsymbol{\beta}}$ exist;*

(H1.2) There exist $w \geq 1$ and $R > 0$ such that $f_{\alpha,\boldsymbol{\beta}} \in L^2\left(w \otimes W^{\otimes p}\right)$, where W is either $i_{[-R,R]}$ or $\cosh(\cdot/R)$;

(H1.3) There exists $x_0 > 0$ and $\mathcal{X} = [-x_0, x_0]^p \subseteq \mathbb{S}_{\boldsymbol{X}}$.

The assumption (H1.2) is important to rely on Hilbert-space methods. It is satisfied when

$$\mathbb{E}\left[w(\alpha)\prod_{k=1}^{p} e^{|\boldsymbol{\beta}_k|/R}\right] < \infty \quad \text{and} \quad \|f_{\alpha,\boldsymbol{\beta}}\|_{L^\infty(\mathbb{R}^{p+1})} < \infty$$

hold. The first condition is a condition on the tails of the density of the random coefficients. The weights W are such that $L^2\left(w \otimes W^{\otimes p}\right) \subseteq L^2(w \otimes (e^{|\cdot|/R})^{\otimes p})$. When $W = i_{[-R,R]}$, $f_{\alpha,\boldsymbol{\beta}} \in L^2(w \otimes W^{\otimes p})$ implies that $\mathbb{S}_{\boldsymbol{\beta}} \subseteq [-R,R]^p$. If $w^{-1} \in L^1(\mathbb{R})$, (H1.2) implies that the slopes of $\boldsymbol{\beta}$ do not have heavy tails. This means that their tails are not heavier than that of the Laplace distribution (*i.e.*, the Laplace transform of their absolute value is finite near 0). Indeed, we have, for all $\epsilon \in (0,1)$ and $k = 1, \dots, p$, for $\lambda = (1-\epsilon)/(2R)$, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[e^{\lambda|\boldsymbol{\beta}_k|}\right] \leq \|f_{\alpha,\boldsymbol{\beta}}\|_{L^2(w\otimes W^{\otimes p})} \left\|w^{-1}\right\|_{L^1(\mathbb{R})}^{1/2} (2R/\epsilon)^{p/2} < \infty.$$

(H1.2) allows for slopes which marginal distributions are Gaussian, inverse Gaussian, logistic, and all distributions with compact support, among others. But it rules out the lognormal distribution. The case $w = 1$ corresponds to mild integrability in the first variable. The condition that the support of the regressors contains an hypercube $\mathcal{X} = [-x_0, x_0]^p$ in (H1.3) is not restrictive because $Y = \alpha + \boldsymbol{\beta}^\top \underline{\boldsymbol{x}} + \boldsymbol{\beta}^\top \boldsymbol{M}^{-1}\boldsymbol{M}(\boldsymbol{X} - \underline{\boldsymbol{x}})$, we can take $\underline{\boldsymbol{x}}$, $\boldsymbol{M}$ an invertible $p \times p$ matrix, and x_0 such that $\mathcal{X} \subseteq \mathbb{S}_{\boldsymbol{M}(\boldsymbol{X}-\underline{\boldsymbol{x}})}$, and there is a one-to-one mapping between $f_{\alpha+\boldsymbol{\beta}^\top\underline{\boldsymbol{x}},(\boldsymbol{M}^{-1})^\top\boldsymbol{\beta}}$ and $f_{\alpha,\boldsymbol{\beta}}$.

3.2.3 Inverse problem in Hilbert spaces

Estimation of $f_{\alpha,\boldsymbol{\beta}}$ corresponds to solving a statistical ill-posed inverse problem. Indeed, we can relate $f_{\alpha,\boldsymbol{\beta}}$ to the conditional characteristic function $f_{Y|\boldsymbol{X}=x}$ known over a subset of $\mathbb{R}^{p+1}$, namely a double cone. This can be formalized via the introduction of the operator: for all $t \in \mathbb{R}$ and $\boldsymbol{u} \in [-1,1]^p$,

$$\mathcal{K} f_{\alpha,\boldsymbol{\beta}}(t,\boldsymbol{u}) = \mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\boldsymbol{u}}\right](t) x_0 |t|^{p/2}, \tag{3.4}$$

where

$$\begin{array}{llll}\mathcal{K}: & L^2\left(w \otimes W^{\otimes p}\right) & \to & L^2(\mathbb{R}\times[-1,1]^p) \\ & f & \to & (t,\boldsymbol{u}) \mapsto \mathcal{F}\left[f\right](t, x_0 t\boldsymbol{u}) x_0 |t|^{p/2}.\end{array} \tag{3.5}$$

Proposition 1. *$L^2\left(w \otimes W^{\otimes p}\right)$ is continuously embedded into $L^2(\mathbb{R}^{p+1})$. Moreover, $\mathcal{K}$ is injective and continuous, and not compact if $w = 1$.*

Thus, when $w = 1$, the SVD of $\mathcal{K}$ does not exist. This makes it difficult to prove rates of convergence even for estimators which do not rely explicitly on the SVD such as the Tikhonov and Landweber method (Gerchberg algorithm in out-of-band extrapolation, see Bertero and Boccacci (1998)). Rather than working with $\mathcal{K}$ directly, we use that $\mathcal{K}$ is the composition of operators which are easier to analyze

$$\forall\, t \in \mathbb{R},\ \mathcal{K}[f](t,\star) = \mathcal{F}_{tx_0}\left[\mathcal{F}_{1\mathrm{st}}\left[f\right](t,\cdot_2)\right](\star) x_0 |t|^{p/2} \text{ in } L^2([-1,1]^p). \tag{3.6}$$

For all $f \in L^2\left(w \otimes W^{\otimes p}\right)$ and $t \in \mathbb{R}$, $\mathcal{F}_{1\mathrm{st}}\left[f\right](t,\cdot_2)$ belongs to $L^2(W^{\otimes p})$ and, for

$c \neq 0$, $\mathcal{F}_c: \ L^2(W^{\otimes p}) \to L^2([-1,1]^p)$ admits a SVD, where both orthonormal systems are complete. This is a tensor product of the SVD when $p = 1$ that we denote by $(\sigma_m^{W,c}, \varphi_m^{W,c}, g_m^{W,c})_{m\in\mathbb{N}_0}$, where $(\sigma_m^{W,c})_{m\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0}$ is in decreasing order repeated according to multiplicity. Note that (3.6) allows to rely on the property of the unidimensional truncated Fourier transform to analyze the problem of the multidimensional truncated Fourier transform on a double cone. This is a natural way to exploit the geometry of the problem.

Proposition 2. *For all $c \neq 0$, $(g_m^{W,c})_{m\in\mathbb{N}_0}$ and $(\varphi_m^{W,c})_{m\in\mathbb{N}_0}$ are bases of, respectively, $L^2([-1,1])$ and $L^2(W)$.*

The SVD when $W = i_{[-1,1]}$ is well studied. The singular functions $(g_m^{i_{[-1,1]},c})_{m\in\mathbb{N}_0}$ are the PSWF. They can be extended as entire functions in $L^2(\mathbb{R})$ and form a complete orthogonal system of $PW(c)$ for which we use the same notation. They are useful to carry interpolation and extrapolation (see, *e.g.*, Lindberg (2012)) with Hilbertian techniques. Nonasymptotic upper bounds on the singular values show that the latter decay exponentially with m faster than $e^{-m\ln(4(m+3/2)/e|c|)}$ (see, *e.g.*, Lemma B.4. in Appendix 3.7 and Bonami and Karoui (2014*b*)).

The weight $\cosh(\cdot/R)$ allows for larger classes than $PW(c)$, hence noncompact $\mathbb{S}_{\boldsymbol{\beta}}$. The singular functions $(g_m^{\cosh(\cdot/R),c})_{m\in\mathbb{N}_0}$ allow to carry extrapolation of nonbandlimited functions (see Gaillac and Gautier (2019*a*)). This is useful even if $\mathbb{S}_{\boldsymbol{\beta}}$ is compact when the researcher does not know a superset containing $\mathbb{S}_{\boldsymbol{\beta}}$. Extending the work of Widom (1964), Gaillac and Gautier (2019*a*) also proves nonasymptotic lower and upper bounds on the singular values, which show that, for $c < 1$, the latter decay with m faster than $e^{-m\ln(1/|c|)}/\sqrt{2m+1}$ (see Theorem 7 in Gaillac and Gautier (2019*a*)). Useful results for the present paper on the corresponding SVD are in Appendix 3.7. More properties and a numerical algorithm to compute it are in Gaillac and Gautier (2019*a*).

Nonasymptotic upper bounds on $\|g_m^{W,c}\|_{L^\infty(-1,1)}$ are proved in Bonami and Karoui (2014*b*) for the PSWF and in Gaillac and Gautier (2019*a*) when $W = \cosh(\cdot/R)$. They are recalled in Proposition 5. They take the form $\|g_m^{W,c}\|_{L^\infty(-1,1)} \leq H(c)\sqrt{m+1/2}$

for a given $H(c)$. We use the explicit dependence of H in c for all $m \in \mathbb{N}_0$ to prove the results in Section 3.4.3.

3.2.4 Related Gaussian white noise model

The next idealized model is related to (3.1)-(3.2) when $f_{\boldsymbol{X}}$ is known:

$$dZ(t) = \mathcal{K}\left[f\right](t, \cdot_2)dt + \frac{1}{\sqrt{n}}dG(t), \quad t \in \mathbb{R}, \tag{3.7}$$

where f plays the role of $f_{\alpha,\boldsymbol{\beta}}$ and $(G(t))_{t\in\mathbb{R}}$ is a complex two-sided cylindrical Gaussian process on $L^2([-1,1]^p)$ (see Da Prato and Zabczyk (2014)). Z and G are function-valued processes and those functions can take complex values. The partial derivative in the sense of distributions with respect to time of G is the space time white noise in $L^2(\mathbb{R} \times [-1,1]^p)$. G is the partial derivative in the sense of distributions obtained by differentiating once with respect to each space variable the Brownian sheet.

By taking the inner product of both sides of (3.7) with $g_{\boldsymbol{m}}^{W,x_0t}$ for all $\boldsymbol{m} \in \mathbb{N}_0^p$, we get the system of independent equations

$$Z_{\boldsymbol{m}}(t) = \int_0^t \sigma_{\boldsymbol{m}}^{W,x_0s} b_{\boldsymbol{m}}(s)ds + \frac{1}{\sqrt{n}}B_{\boldsymbol{m}}(t), \quad t \in \mathbb{R}, \tag{3.8}$$

where $Z_{\boldsymbol{m}} := \langle Z(\star), g_{\boldsymbol{m}}^{W,x_0\star}\rangle_{L^2([-1,1]^p)}$, $B_{\boldsymbol{m}}(t) = B_{\boldsymbol{m}}^{\mathfrak{R}}(t) + iB_{\boldsymbol{m}}^{\mathfrak{I}}(t)$, and $(B_{\boldsymbol{m}}^{\mathfrak{R}}(t))_{t\in\mathbb{R}}$ and $(B_{\boldsymbol{m}}^{\mathfrak{I}}(t))_{t\in\mathbb{R}}$ are independent two-sided Brownian motions.

It is common in statistics of inverse problems to analyze an idealized Gaussian white noise model (see, *e.g.*, Hohmann and Holzmann (2015)). This paper sometimes refer to model (3.8) to develop intuitions. Some of the minimax lower bounds are proved only in this context for simplicity. Asymptotic equivalence with white noise models has been proved in some cases such as regression models (see, *e.g.*, Jähnisch and Nussbaum (2003); Reiß (2008)). These results do not apply in the current random coefficients model because, among other things, it is an inverse problem. Asymptotic equivalence results for some inverse problems have been proved (see, *e.g.*, Meister (2011) for the functional linear regression). Proving such an equivalence in our context is out of the scope of this paper.

3.2.5 Interpolation

This paper relies on interpolation when the variance of an initial estimator $\widehat{f}^0(t)$ of $f(t)$ is large when t is close to 0 but $\|f-\widehat{f}^0\|^2_{L^2(\mathbb{R}\setminus(-\epsilon,\epsilon))}$ is small. In (3.8), because $\sigma_{\boldsymbol{m}}^{W,x_0t}$ is small when $|\boldsymbol{m}|_q$ is large or t is small (see Lemma 12), the estimator of Section 3.4.1 truncates large values of $|\boldsymbol{m}|_q$ and does not rely on small values of $|t|$ but uses interpolation. Specifically, we work with

$$\widehat{f}(t) = \widehat{f}^0(t)\mathbb{1}\{|t|\geq\epsilon\} + \mathcal{I}_{\underline{a},\epsilon}[\widehat{f}^0](t)\mathbb{1}\{|t|<\epsilon\},$$

where, for all $\underline{a},\epsilon>0$, $\mathcal{I}_{\underline{a},\epsilon}$ is the interpolation operator on $L^2(\mathbb{R})$ with domain $PW(\underline{a})$,

$$\mathcal{I}_{\underline{a},\epsilon}[f] := \sum_{m\in\mathbb{N}_0} \frac{\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\left\langle f, g_m^{i_{[-1,1]},\underline{a}\epsilon}(\cdot/\epsilon)\right\rangle_{L^2(\mathbb{R}\setminus(-\epsilon,\epsilon))}}{\left(1-\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\right)\epsilon} g_m^{i_{[-1,1]},\underline{a}\epsilon}\left(\frac{\cdot}{\epsilon}\right), \tag{3.9}$$

and, for all $m\in\mathbb{N}_0$ and $c\neq 0$, $\rho_m^{W,c} := 2\pi(\sigma_m^{W,c})^2/|c|$. Then, (3.11) below yields

$$\begin{aligned}\left\|f-\widehat{f}\right\|^2_{L^2(\mathbb{R})} \leq &(1+2C_0(\underline{a}\epsilon))\left\|f-\widehat{f}^0\right\|^2_{L^2(\mathbb{R}\setminus(-\epsilon,\epsilon))} \\ &+ 2(1+C_0(\underline{a}\epsilon))\left\|f-\mathcal{P}_{\underline{a}}[f]\right\|^2_{L^2(\mathbb{R})},\end{aligned} \tag{3.10}$$

where $C_0 := 4\cdot/(\pi(1-\rho_0^{i_{[-1,1]},\cdot})^2)$, which bounds the error made on $\mathbb{R}$ by the interpolated estimator $\widehat{f}$ by the sum of the error of the initial estimator $\widehat{f}^0$ on $\mathbb{R}\setminus(-\epsilon,\epsilon)$ and a term related to the distance of f to its projection on $PW(\underline{a})$.

Proposition 3. *For all $\underline{a},\epsilon>0$, $\mathcal{I}_{\underline{a},\epsilon}(L^2(\mathbb{R}))\subseteq L^2([-\epsilon,\epsilon])$ and, for all $g\in PW(\underline{a})$, $\mathcal{I}_{\underline{a},\epsilon}[g]=g$ in $L^2(\mathbb{R})$, and, for all $f,h\in L^2(\mathbb{R})$,*

$$\begin{aligned}\left\|f-\mathcal{I}_{\underline{a},\epsilon}[h]\right\|^2_{L^2([-\epsilon,\epsilon])} \leq &2(1+C_0(\underline{a}\epsilon))\left\|f-\mathcal{P}_{\underline{a}}[f]\right\|^2_{L^2(\mathbb{R})} \\ &+ 2C_0(\underline{a}\epsilon)\left\|f-h\right\|^2_{L^2(\mathbb{R}\setminus(-\epsilon,\epsilon))}.\end{aligned} \tag{3.11}$$

If $f \in PW(\underline{a})$, $\mathcal{I}_{\underline{a},\epsilon}[f]$ only relies on $f \mathbb{1}\{\mathbb{R} \setminus (-\epsilon, \epsilon)\}$ and $\mathcal{I}_{\underline{a},\epsilon}[f] = f$ on $\mathbb{R} \setminus (-\epsilon, \epsilon)$, so (3.9) provides an analytic formula to carry interpolation on $[-\epsilon, \epsilon]$ of functions in $PW(\underline{a})$. Else, (3.11) provides an upper bound on the error made by approximating f by $\mathcal{I}_{\underline{a},\epsilon}[h]$ on $[-\epsilon, \epsilon]$ when h approximates f outside $[-\epsilon, \epsilon]$. When $\text{supp}(\mathcal{F}[f])$ is compact, $\underline{a}$ is taken such that $\text{supp}(\mathcal{F}[f]) \subseteq [-\underline{a}, \underline{a}]$. Else, $\underline{a}$ goes to infinity so the second term in (3.10) goes to 0. ϵ is taken such that $\underline{a}\epsilon$ is constant because, due to (3.87) in Osipov et al. (2013), $\lim_{t\to\infty} C_0(t) = \infty$ and (3.11) and (3.10) become useless. We set $C = 2(1 + C_0(\underline{a}\epsilon))$.

3.2.6 Sets of smooth and integrable functions

Define, for $q \in \{1, \infty\}$,

$$b_{\boldsymbol{m}}(t) := \left\langle \mathcal{F}_{1\text{st}}[f](t, \cdot_2), \varphi_{\boldsymbol{m}}^{W,x_0t} \right\rangle_{L^2(W^{\otimes p})}, \theta_{q,k}(t) := \left(\sum_{\boldsymbol{m} \in \mathbb{N}_0^p:\ |\boldsymbol{m}|_q = k} |b_{\boldsymbol{m}}(t)|^2 \right)^{1/2},$$

and, for all $(\phi(t))_{t\geq 0}$ and $(\omega_m)_{m\in\mathbb{N}_0}$ increasing, $\phi(0) = \omega_0 = 1$, $l, M > 0$, $t \in \mathbb{R}$, $\boldsymbol{m} \in \mathbb{N}_0^p$, $k \in \mathbb{N}_0$, $\mathcal{I}_{w,W}(M) := \{f :\ \|f\|_{L^2(w\otimes W^{\otimes p})} \leq M\}$, and

$$\mathcal{H}_{w,W}^{q,\phi,\omega}(l, M) := \left\{ f : \sum_{k\in\mathbb{N}_0} \int_{\mathbb{R}} \phi^2(|t|)\theta_{q,k}^2(t)dt \bigvee \sum_{k\in\mathbb{N}_0} \omega_k^2 \|\theta_{q,k}\|_{L^2(\mathbb{R})}^2 \leq 2\pi l^2 \right\} \bigcap \mathcal{I}_{w,W}(M).$$

We use the notation $\mathcal{H}_{w,W}^{q,\phi,\omega}(l)$ when we require $\|f\|_{L^2(w\otimes W^{\otimes p})} < \infty$ rather than $\|f\|_{L^2(w\otimes W^{\otimes p})} \leq M$. The set $\mathcal{I}_{w,W}(M)$ imposes the integrability discussed in the beginning fo the section. The first set in the definition of $\mathcal{H}_{w,W}^{q,\phi,\omega}(l, M)$ defines the notion of smoothness analyzed in this paper. It involves a maximum, thus two inequalities: the first for smoothness in the first variable and the second for smoothness in the other variables. The asymmetry in the treatment of the first and remaining variables is due to the fact that only the random slopes are multiplied by regressors which have limited variation and we make integrability assumptions in the first variable which are as mild as possible. The smoothness classes in the analysis of the Radon transform usually involve nonstandard weight functions well suited to the operator.

In contrast, the ones in this paper are not too hard to interpret. To gain intuition on the inequalities, note that, by Plancherel and Parseval's identities,

$$
\begin{aligned}
2\pi \|f\|^2_{L^2(1\otimes W^{\otimes p})} &= \int_{\mathbb{R}} \|\mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2)\|^2_{L^2(W^{\otimes p})} dt \\
&= \int_{\mathbb{R}} \sum_{\boldsymbol{m}\in\mathbb{N}_0^p} |b_{\boldsymbol{m}}(t)|^2 dt \\
&= \sum_{k\in\mathbb{N}_0} \|\theta_{q,k}\|^2_{L^2(\mathbb{R})}.
\end{aligned}
\tag{3.12}
$$

The multi-index $\boldsymbol{m}$ is a type of frequency associated to the space variable $\boldsymbol{b}$ and k indices a level of all frequencies with same norm of the multi-index. The fact that $\boldsymbol{m}$ is discrete is easy to understand when $W = i_{[-R,R]}$ and the function has compact support in $\boldsymbol{b}$. It is similar to Fourier series. The choice of the functions $(\varphi_{\boldsymbol{m}}^{W,x_0t})_{\boldsymbol{m}\in\mathbb{N}_0^p}$ is well suited to the decomposition of the operator $\mathcal{K}$. The frequency t is related to the space variable a. It is continuous because we make weaker assumptions on the tails of α.

The first inequality can be written as

$$
\int_{\mathbb{R}} \phi^2(|t|) \|\mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2)\|^2_{L^2(W^{\otimes p})} dt \le 2\pi l^2.
$$

Thus, when $\phi = 1\vee|\cdot|^s$ with $s>0$, it is the usual Sobolev smoothness in $L^2(\mathbb{R};L^2(W^{\otimes p}))$.

The second inequality is obtained by using weighted sums rather than (3.12). The weights are indexed by a frequency level $|\boldsymbol{m}|_q$. There is no weight function for t. Hence, it can be viewed as a smoothness assumption in $\boldsymbol{b}$ only. It can be written: there exists a density $\overline{\phi}$ on $\mathbb{R}$ such that

$$
\forall t\in\mathbb{R},\ \sum_{\boldsymbol{m}\in\mathbb{N}_0^p} \omega^2_{|\boldsymbol{m}|_q} |b_{\boldsymbol{m}}(t)|^2 \le \overline{\phi}(t) 2\pi l^2. \tag{3.13}
$$

The different choices of sequences $(\omega_k)_{k\in\mathbb{N}_0}$ that we shall consider correspond to different source conditions for fixed t expressed in terms of the operator $\mathcal{J}_{W,x_0t}$ from

$L^2\left(W^{\otimes p}\right)$ to $L^2\left([-1,1]^p\right)$, for $f \in L^2\left(W^{\otimes p}\right)$, by

$$\mathcal{J}_{W,x_0t}[f] := \sum_{\boldsymbol{m}\in\mathbb{N}_0^p} \frac{1}{|\boldsymbol{m}|_q} \left\langle \varphi_{\boldsymbol{m}}^{W,x_0t}, f\right\rangle_{L^2(W^{\otimes p})} g_{\boldsymbol{m}}^{W,x_0t}.$$

When $q=1$, the unbounded operator $\mathcal{J}_{W,x_0t}^{-1}$ can be viewed as a differential operator. When $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, with $\sigma > 0$, then, $\mathcal{F}_{1\mathrm{st}}\left[f\right](t,\cdot_2)$ satisfies (3.13) if and only if it belongs to

$$\left\{g \in L^2\left(W^{\otimes p}\right):\ g = (\mathcal{J}_{W,x_0t}^{*}\mathcal{J}_{W,x_0t})^\sigma v,\ \|v\|^2_{L^2(W^{\otimes p})} \le \overline{\phi}(t)2\pi l^2\right\}. \tag{3.14}$$

When, for almost every a, $\boldsymbol{b} \mapsto f(a,\boldsymbol{b})$ has compact support, we show in Appendix 3.7.5 that it is possible to relate the smoothness defined via (3.14) to the Sobolev smoothness defined using Fourier series. There, smoothness corresponds to the function having bounded sum of squared L^2 norm of partial derivatives of degree σ. When ω are exponentials, the smoothness defined using Fourier series implies that all partial derivatives are square integrable. It corresponds to supersmooth classes (see, *e.g.*, Cavalier (2000)). As it is common in the literature (see Alquier et al. (2011*b*); Carrasco et al. (2007)), due to the different rate of decay of the singular values of $\mathcal{F}_{x_0t}$ in Section 3.2.3, we consider slightly different supersmooth classes when $W = i_{[-R,R]}$ and $W = \cosh(\cdot/R)$. It is $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k\ln(k+1)}\right)_{k\in\mathbb{N}_0}$ when $W = i_{[-R,R]}$ and $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k}\right)_{k\in\mathbb{N}_0}$ when $W = \cosh(\cdot/R)$. This case is similar to nonparametric deconvolution where the density of the noise and the density of interest have Fourier transforms which both decay like $e^{-\kappa|x|^r}$ when $|x| \to \infty$ with same $r \ge 1$ but potentially different $\kappa > 0$ (see, *e.g.*, Lacour (2006); Tsybakov (2000)). When $W = i_{[-R,R]}$, we also consider the case where $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa(k\ln(k+1))^r})_{k\in\mathbb{N}_0}$ and $r > 1$. In nonparametric deconvolution, this case corresponds to the case where the Fourier transforms of the noise and density of interest decay respectively as $e^{-\kappa_1|x|^s}$ and $e^{-\kappa|x|^r}$ for $|x| \to \infty$, where $\kappa_1, \kappa > 0$ and $r > s$ (see case 3 in Theorem 3.1 in Lacour (2006)). The two values of q (1 or ∞) that we consider matter for the rates of convergence for supersmooth functions.

3.2.7 Risk

The risk is the mean integrated squared error (MISE)

$$\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}, f_{\alpha,\boldsymbol{\beta}}\right) := \mathbb{E}\left[\left\|\widehat{f}_{\alpha,\boldsymbol{\beta}} - f_{\alpha,\boldsymbol{\beta}}\right\|_{L^2(1\otimes W^{\otimes p})}^2 \middle| \mathcal{G}_{n_0}\right].$$

It is $\mathbb{E}[\|\widehat{f}_{\alpha,\boldsymbol{\beta}} - f_{\alpha,\boldsymbol{\beta}}\|_{L^2(\mathbb{R}^{p+1})}^2|\mathcal{G}_{n_0}]$ when $W = i_{[-R,R]}$ and $\mathrm{supp}(\widehat{f}_{\alpha,\boldsymbol{\beta}}) \subseteq \mathbb{R}\times[-R,R]^p$, else it is

$$\mathbb{E}\left[\left\|\widehat{f}_{\alpha,\boldsymbol{\beta}} - f_{\alpha,\boldsymbol{\beta}}\right\|_{L^2(\mathbb{R}^{p+1})}^2 \middle| \mathcal{G}_{n_0}\right] \le \|1/W\|_{L^\infty(\mathbb{R})}^p \mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}, f_{\alpha,\boldsymbol{\beta}}\right). \tag{3.15}$$

3.3 Lower bounds

The lower bounds involve a function r (for rate) and take the form

$$\exists \nu > 0 :\ \underline{\lim}_{n\to\infty} \inf_{\widehat{f}_{\alpha,\boldsymbol{\beta}}} \sup_{f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l)\cap\mathcal{D}} \frac{\mathbb{E}\left[\left\|\widehat{f}_{\alpha,\boldsymbol{\beta}} - f_{\alpha,\boldsymbol{\beta}}\right\|_{L^2(\mathbb{R}^{p+1})}^2\right]}{r(n)^2} \ge \nu. \tag{3.16}$$

When we replace $f_{\alpha,\boldsymbol{\beta}}$ by f, $\widehat{f}_{\alpha,\boldsymbol{\beta}}$ by $\widehat{f}$, remove $\mathcal{D}$ from the nonparametric class, and consider model (3.8), we refer to (3.16'). We use $k_q := \mathbb{1}\{q=1\} + p\mathbb{1}\{q=\infty\}$ and $k_q' = p+1-k_q$. We consider polynomial and exponential weights $(\omega_k)_{k\in\mathbb{N}_0}$ which yield respectively the smooth and supersmooth cases described in Section 3.2.6.

Theorem 1. *Let $q\in\{1,\infty\}$, ϕ increasing on $[0,\infty)$, $0<l,s,\kappa<\infty$, and w such that $\int_1^\infty w(a)/a^4<\infty$. When $W=i_{[-R,R]}$,*

- *(T1.1a) $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, ϕ is such that $\lim_{\tau\to\infty}\int_0^\infty \phi(t)^2 e^{-2\tau t}dt = 0$, $f_{\boldsymbol{X}}$ is known, and $\|f_{\boldsymbol{X}}\|_{L^\infty(\mathcal{X})} < \infty$, $\mathbb{S}_{\boldsymbol{X}} = \mathcal{X}$, then (3.16) holds with $r(n) = (\ln(n)/\ln_2(n))^{-(2+k_q/2)\vee\sigma}$,*

- *(T1.1b) we consider model (3.8), $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k\ln(k+1)}\right)_{k\in\mathbb{N}_0}$, then (3.16') holds with $r(n) = n^{-\kappa/(2\kappa+2k_q)}/\ln(n)$.*

When $W=\cosh(\cdot/R)$, we consider model (3.8), *(T1.2a)* $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, *for all* $\overline{\sigma} > 1/2$, *then (3.16') holds with* $r(n) = \ln(n/\ln(n))^{-\overline{\sigma}\vee\sigma}$

(T1.2b) $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k}\right)_{k\in\mathbb{N}_0}$, *then (3.16') holds with* $r(n) = n^{-\kappa/(2\kappa+2k_q)}$.

By (3.15), (T1.2a), and (T1.2b), we obtain lower bounds involving $\mathcal{R}_{n_0}^W$. For smooth functions when $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, we obtain logarithmic lower bounds. This is consistent with the rates obtained in other severely ill-posed inverse problems with smooth functions of interest (see, *e.g.*, Hohmann and Holzmann (2015); Lacour (2006); Mair and Ruymgaart (1996)). Note that in (T1.1a) and (T1.2a), $r(n)$ does not depend on the dimension p (see, *e.g.*, Bissantz et al. (2007)). Recall that, to match the rate of decay of the singular values, cases (T1.1b) and (T1.2b) consider supersmooth functions characterized by $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k\ln(k+1)}\right)_{k\in\mathbb{N}_0}$ when $W = i_{[-R,R]}$ and by $(\omega_k)_{k\in\mathbb{N}_0} = \left(e^{\kappa k}\right)_{k\in\mathbb{N}_0}$ when $W = \cosh(\cdot/R)$. Such a situation corresponds to "2exp-severely ill-posed problems" (see, *e.g.*, Bissantz et al. (2007); Cavalier et al. (2004); Tsybakov (2000)), where the eigenvalues of the operator decay exponentially fast to zero and the weights are exponentials of the same form. Importantly, (T1.1b) and (T1.2b) will be supplemented by matching upper bounds showing that, for sufficiently smooth classes of functions, polynomial rates can be attained, for this severely ill-posed inverse problem. The lower bound (T1.2b) is, up to the logarithmic term, the one in in the first case of Theorem 3.1 in Lacour (2006) for nonparametric deconvolution when the density of the noise and the density of interest have Fourier transforms decaying respectively like $e^{-k_q|x|}$ and $e^{-\kappa|x|}$ when $|x| \to \infty$. The links to classical nonparametric estimation bounds are not direct because the operator is a composition and sometimes does not have a SVD.

The proof of the lower bounds follow the usual reduction scheme where one obtains a lower bound on the left-hand side of (3.16) (or (3.16')) by replacing the supremum by a maximum over two well-chosen densities $f_{\alpha,\boldsymbol{\beta}} \in \mathcal{H}_{w,W}^{q,\phi,\omega}(l) \cap \mathcal{D}$ (respectively $f \in \mathcal{H}_{w,W}^{q,\phi,\omega}(l)$). The standard approach is to choose the latter two as linear combinations of the singular functions of the operator $\mathcal{K}$. However, there are two main complications. First, by Proposition 1, a SVD sometimes does not exist so we rely on the decomposition of $\mathcal{K}$ as a composition of two operators. Second, to deal with the composition, we make use nonasymptotic (over $\boldsymbol{m}$) bounds involving the SVD of $\mathcal{F}_c$ with an explicit dependence on c. The proof of Theorem 1 thus relies on results in harmonic analysis that we prove in Appendix 3.7 for the PSWF and in Gaillac and Gautier (2019a) when $W = \cosh(\cdot/R)$.

3.4 Estimation

This section uses the additional following simplifying assumption.

Assumption 2. (*H2.1*) *We have at our disposal i.i.d* $(Y_i, \boldsymbol{X}_i)_{i=1}^n$ *and an estimator* $\widehat{f}_{\boldsymbol{X}|\mathcal{X}}$ *based on* $\mathcal{G}_{n_0} = (\boldsymbol{X}_i)_{i=-n_0+1}^0$ *independent of* $(Y_i, \boldsymbol{X}_i)_{i=1}^n$*;*

(*H2.2*) $\mathcal{E}$ *is a set of densities on* $\mathcal{X}$ *defined in (H1.3) such that, for* $c_{\boldsymbol{X}}, C_{\boldsymbol{X}} \in (0,\infty)$*, for all* $f \in \mathcal{E}$*,* $\|f\|_{L^\infty(\mathcal{X})} \le C_{\boldsymbol{X}}$ *and* $\|1/f\|_{L^\infty(\mathcal{X})} \le c_{\boldsymbol{X}}$*, and, for* $(v(n_0,\mathcal{E}))_{n_0\in\mathbb{N}} \in (0,1)^{\mathbb{N}}$ *which tends to 0, we have*

$$\frac{1}{v(n_0,\mathcal{E})} \sup_{f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}} \left\| \widehat{f}_{\boldsymbol{X}|\mathcal{X}} - f_{\boldsymbol{X}|\mathcal{X}} \right\|_{L^\infty(\mathcal{X})}^2 = O_p(1)\,.$$

We assume (H2.1) because the estimator involves estimators of $f_{\boldsymbol{X}|\mathcal{X}}$ in denominators. Alternative solutions exist when $p = 1$ (see, *e.g.*, Hoderlein et al. (2017); Holzmann and Meister (2019); Kerkyacharian and Picard (2004)). The availability of the preliminary sample $\mathcal{G}_{n_0}$ in (H2.1) is not really an assumption but makes it explicit that the theory below relies on sample splitting. We do this for the simplicity of the proofs and relaxing it could be done in future work. In practice we do not use sample splitting in Section 3.5. These simulation results indicate that a practically oriented researcher does not need to implement sample splitting. By choosing well $\mathcal{X} \subseteq \mathbb{S}_{\boldsymbol{X}}$ the assumption $\|f_{\boldsymbol{X}|\mathcal{X}}\|_{L^\infty(\mathcal{X})} \le C_{\boldsymbol{X}}$ and $\|1/f_{\boldsymbol{X}|\mathcal{X}}\|_{L^\infty(\mathcal{X})} \le c_{\boldsymbol{X}}$ in (H2.2) will be satisfied. By doing so, $\widehat{f}_{\boldsymbol{X}|\mathcal{X}}$ effectively uses a subsample of the preliminary sample so $\mathcal{X}$ should not be too small.

3.4.1 Estimator

Based on the decomposition (3.6), the regularized inverse of $\mathcal{K}$ is obtained in three steps. First, for $|t| \le T$, we use an approximation of $\mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot)$ using the regularized inverse of the truncated Fourier operator $\mathcal{F}_{tx_0}$. It involves spectral cut-off. Second, because the singular values of $\mathcal{F}_{tx_0}$ go to 0 as t goes to 0, using the SVD for estimation is problematic and we would rely on few to none coefficients. Rather, we rely on the interpolation strategy of Section 3.2.5 for $t \in [-\epsilon, \epsilon]$, where $0 < \epsilon < 1 < T$. Third, we use a regularized inverse of the partial Fourier transform with respect to the first variable to recover $f_{\alpha,\boldsymbol{\beta}}$.

Let us now precise these three steps. For all $q \in \{1, \infty\}$, $0 < \epsilon < 1 < T$, $\underline{N} \in \mathbb{R}^{\mathbb{R}}$, $N(t) = \lfloor \underline{N}(t) \rfloor$ for $\epsilon \le |t| \le T$, $N(t) = N(\epsilon)$ for $|t| \le \epsilon$ and $N(t) = N(T)$ for $|t| > T$, a regularized inverse is obtained by:

(S.1) for all $t \neq 0$, obtain a first approximation of $F_1(t,\cdot) := \mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot)$

$$F_1^{q,N,T,0}(t,\cdot_2) := \mathbb{1}\{|t| \le T\} \sum_{|\boldsymbol{m}|_q \le N(t)} \frac{c_{\boldsymbol{m}}(t)}{\sigma_{\boldsymbol{m}}^{W,x_0 t}} \varphi_{\boldsymbol{m}}^{W,x_0 t}, \tag{3.17}$$

$$c_{\boldsymbol{m}}(t) := \left\langle \mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\cdot}\right](t), g_{\boldsymbol{m}}^{W,x_0 t} \right\rangle_{L^2([-1,1]^p)}, \tag{3.18}$$

(S.2) for all $t \in [-\epsilon, \epsilon]$, we use the interpolation

$$F_1^{q,N,T,\epsilon}(t,\cdot) := F_1^{q,N,T,0}(t,\cdot)\mathbb{1}\{|t| \ge \epsilon\} + \mathcal{I}_{\underline{a},\epsilon}\left[F_1^{q,N,T,0}(\star,\cdot)\right](t)\mathbb{1}\{|t| < \epsilon\},$$

(S.3) $f_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon}(\cdot_1,\cdot_2) := \mathcal{F}_{1\text{st}}^{I}\left[F_1^{q,N,T,\epsilon}(\star,\cdot_2)\right](\cdot_1)$.

Let us comment the choice of the estimator. For the regularized inverse of the truncated Fourier operator $\mathcal{F}_{tx_0}$ in step (S.1), we choose spectral cut-off instead other regularization methods such as Tikhonov regularization or Landweber iteration (see, *e.g.*, Section 1.2.2 in Alquier et al. (2011*b*)). We do this because the SVD is fast to compute using numerical schemes developed recently (see Section 3.5). Moreover, the rates of spectral cut-off do not suffer from limitations due to qualification. Step (S.2) is the interpolation step. The interpolation is a nonstandard step, but it is essential to obtain the polynomial rates for supersmooth densities. The regularized inverse of the partial Fourier transform $\mathcal{F}_{1\mathrm{st}}$ is performed using the indicator $\mathbb{1}\{|t| \leq T\}$ in (S.3), which is a standard regularization device when inverting the Fourier transform which consists in removing high frequencies (see, *e.g.*, Cavalier (2000); Comte and Lacour (2011)). We could however use different smoothing kernels.

The regularized inverse depends on R defined in Assumption 1 which should be known. This is not a smoothing parameter. When the $\mathbb{S}_{\boldsymbol{\beta}}$ is assumed to be compact then it is needed that $\mathbb{S}_{\boldsymbol{\beta}} \subseteq [-R, R]^p$, not $\mathbb{S}_{\boldsymbol{\beta}} = [-R, R]^p$. The choice of the parameter $\underline{a}$ in the interpolation step (S.2) will be discussed later. The factor $x_0|t|^{p/2}$ in the definition of $\mathcal{K}$ in (3.5) is used to show the continuity of $\mathcal{K}$ in Proposition 1. Because it also appears on the right-hand side of (3.4), it does not enter the regularized inverse.

To deal with the statistical problem, we replace $c_{\boldsymbol{m}}$ by

$$\widehat{c}_{\boldsymbol{m}} := \frac{1}{n} \sum_{j=1}^{n} \frac{e^{i\star Y_j}}{x_0^p \widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j)} \overline{g_{\boldsymbol{m}}^{W,x_0\star}} \left(\frac{\boldsymbol{X}_j}{x_0} \right) \mathbb{1}\left\{ \boldsymbol{X}_j \in \mathcal{X} \right\}, \tag{3.19}$$

where $\widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j) := \widehat{f}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j) \vee \sqrt{\delta(n_0)}$ and $\delta(n_0)$ is a trimming factor converging to zero. This yields the estimators $\widehat{F}_1^{q,N,T,0}$, $\widehat{F}_1^{q,N,T,\epsilon}$, and $\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon}$. Because inverting the truncated Fourier operator $\mathcal{F}_{tx_0}$ is more ill-posed near 0 (see Lemma 12 and Theorem 7 in Gaillac and Gautier (2019*a*)), $\widehat{F}_1^{q,N,T,0}$ has a large variance for $t \in [-\epsilon, \epsilon]$. Hence we use interpolation (see Section 3.2.5).

We use $(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon})_+$ as a final estimator of $f_{\alpha,\boldsymbol{\beta}}$ which has a smaller risk than $\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon}$ (see Gautier and Kitamura (2013); Tsybakov (2008)). We use $n_e = n \wedge \lfloor \delta(n_0)/v(n_0, \mathcal{E}) \rfloor$ for the sample size required for an ideal estimator where $f_{\boldsymbol{X}|\mathcal{X}}$ is known to achieve the rate of the plug-in estimator.

3.4.2 Upper bounds

The upper bounds below take the form

$$\sup_{f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}^{q,\phi,\omega}_{w,W}(l,M)\cap\mathcal{D},\ f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}} \frac{\mathcal{R}^{W}_{n_0}\left(\widehat{f}^{q,N,T,\epsilon}_{\alpha,\boldsymbol{\beta}}, f_{\alpha,\boldsymbol{\beta}}\right)}{r(n_e)^2} = O_p(1). \tag{3.20}$$

With the restriction $f_{\alpha,\boldsymbol{\beta}} \in \mathcal{H}^{q,\phi,\omega}_{w,W}(l)\cap\mathcal{D}$, we refer to (3.20').

Choice of the parameters. In this section $\underline{N}$, hence N, is a constant independent of t. $\underline{N}$ is a function of t only in Section 3.4.3. Denote, for $u>0$, by $K_{\underline{a}}(u) := \underline{a}\mathbb{1}\{w \neq i_{[-\underline{a},\underline{a}]}\} + u\mathbb{1}\{w = i_{[-\underline{a},\underline{a}]}\}$. We use $T=\phi^I(\omega_{\underline{N}})$, $\underline{a}=w^I(\omega^2_{\underline{N}})$ when we do not know that $\mathbb{S}_\alpha \subseteq [-\underline{a},\underline{a}]$ and cannot take $w=i_{[-\underline{a},\underline{a}]}$, and

1. when $W=i_{[-R,R]}$, $\underline{N}$ is solution to

$$2k_q\left(\underline{N}+\frac{k'_q}{2}\right)\ln\left(\underline{N}K_{\underline{a}}(1)\right)+\ln(\omega^2_{\underline{N}})+(p-1)\ln\left(\underline{N}\right)=\ln(n_e) \tag{3.21}$$

   and $\epsilon = 7e\pi/(Rx_0K_{\underline{a}}(1))$;

2. when $W=\cosh(\cdot/R)$, $\underline{N}$ is solution to

$$2k_q\left(\underline{N}+\frac{k'_q}{2}\right)\ln\left(K_{\underline{a}}(e)\right)+\ln(\omega^2_{\underline{N}})+\frac{p-1}{q}\ln(\underline{N})=\ln(n_e) \tag{3.22}$$

   and $\epsilon = 7e^2\pi/(2Rx_0K_{\underline{a}}(14e^2))$.

(3.21)-(3.22) describe the choices of $\underline{N}$ which realize the bias-variance trade-off. Under these choices of tuning parameters, theorems 2 and 3 provide the convergence rates $r(n_e)$ in (3.20) or (3.20'). Results are presented in increasing order of restrictions on ϕ, then on $(\omega_k)_{k\in\mathbb{N}_0}$, and on w.

Theorem 2. *Let $W=i_{[-R,R]}$, $\mathbb{S}_{\boldsymbol{\beta}}\subseteq[-R,R]^p$, $q\in\{1,\infty\}$, and $l,M,s,\sigma,\kappa,\mu,\gamma,\nu>0$. Consider $\phi=1\vee|\cdot|^s$,*

(T2.1) $(\omega_k)_{k\in\mathbb{N}_0}=(k^\sigma)_{k\in\mathbb{N}_0}$, and $w=1\vee|\cdot|^\mu$, then (3.20) holds with $r(n_e)=(\ln(n_e)/\ln_2(n_e))^{-\sigma}$,

(T2.2) $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa k \ln(k+1)})_{k\in\mathbb{N}_0}$, *with* $s \geq \kappa(p+1)/(2k_q(\nu\mathbb{1}\{W \neq i_{[-\underline{a},\underline{a}]}\}+1))$, *and*

(T2.2a) $w^I(e^{2\kappa|\cdot|\ln(|\cdot|+1)}) = \cdot^{\nu}$, *then (3.20) holds with* $r(n_e) = n_e^{-\kappa/(2\kappa+2(\nu+1)k_q)}\ln(n_e)^{\Lambda(\nu)}$,

(T2.2b) $\underline{a}$ *such that* $\mathbb{S}_\alpha \subseteq [-\underline{a},\underline{a}]$, $w = i_{[-\underline{a},\underline{a}]}$, *then (3.20') holds with* $r(n_e) = n_e^{-\kappa/(2\kappa+2k_q)}\ln(n_e)^{\Lambda(0)}$,

where $\Lambda := ((2+\cdot)p-1)(1-(\kappa(p+1)/\left(2s\left(k_q(\cdot+1)+\kappa\right)\right))/2$.

(T2.3) Consider $\phi = e^{\gamma|\cdot|}$, $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa(k\ln(k+1))^r})_{k\in\mathbb{N}_0}$, w *such that* $w^I(e^{2\kappa(|\cdot|\ln(|\cdot|+1))^r}) = \cdot^{\nu}$, *and* $r > 1$, *then (3.20) holds with* $r(n_e) = \sqrt{\varphi\left(n_e\right)/n_e}$, *where*

$$\varphi := \exp\left(-\sum_{k=1}^{k_0}(-1)^k d_k \ln(\cdot)^{(1/r-1)k+1}\right)\bigvee \ln(\cdot)^{p+1+(p-1)/r},$$

$d_0 = 2\kappa(1+(k_q(1+\nu)+(2+\nu)p-1)/(\ln(2+1/p)(1+1/p))^{r-1})$, $k_0 := \lfloor r/(r-1)\rfloor$, *and for* $k \in \{1,\dots,k_0\}$,

$$d_k := \left(\frac{k_q(1+\nu)(2\kappa)^{1-1/r}\mathbb{1}\{k \equiv 0(mod\ 2)\}}{\kappa(1+\ln(2)/\ln(1+1/p))^r}\right)^k + \left((k_q(1+\nu)+(2+\nu)p-1)\frac{\mathbb{1}\{k \equiv 1(mod\ 2)\}}{\kappa d_0^{1/r-1}}\right)^k.$$

Theorem 1 shows the rate in (T2.1) is optimal when $f_{\boldsymbol{X}}$ is known and $\mathbb{S}_{\boldsymbol{X}} = \mathcal{X}$. It is the same as in Meister (2007) for deconvolution with a known characteristic function of the noise on an interval when the signal has compact support. The rates in Theorem 2 are independent of p as common for severely ill-posed problems (see Chen and Reiss (2011); Gaillac and Gautier (2019*a*)). The rates in (T2.2) and (T2.3) are polynomial and nearly parametric even if the problem is severely ill-posed. This means that we can lose little by going from a possibly misspecified parametric model to a nonparametric one. The rate in (T2.3) is similar to the rate obtained in Lacour (2006) in the deconvolution problem with supersmooth density of interest and noise density, when the former is smoother than the later. The next theorem relaxes the condition $\mathbb{S}_{\boldsymbol{\beta}}$ is compact.

Theorem 3. *Let* $W = \cosh(\cdot/R)$. *For all* $q \in \{1,\infty\}$, $l, M, s, \sigma, \kappa, \mu > 0$, $\phi = 1 \vee |\cdot|^s$, *(T3.1)* *if* $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$*, and* $w = 1 \vee |\cdot|^\mu$*, then* (3.20) *holds with* $r(n_e) = (\ln(n_e)/\ln_2(n_e))^{-\sigma}$,

(T3.2) *if* $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa k})_{k\in\mathbb{N}_0}$*,* $\underline{a}$ *such that* $\mathbb{S}_\alpha \subseteq [-\underline{a}, \underline{a}]$*, and* $w = i_{[-\underline{a},\underline{a}]}$*, then (3.20') holds with* $r(n_e) = n_e^{-\kappa/(2\kappa+2k_q)} \ln(n_e)^{(p-1)\kappa/(2q(\kappa+k_q))}$.

When $1/v(n_0, \mathcal{E}) \geq n$ and $p = 1$, the rate in (T3.2) matches the lower bound (T1.2b) for model (3.8). Again, the results (T2.2), (T2.3), and (T3.2) are related to those for "2exp-severely ill-posed problems", and Tsybakov (2000) obtains the same rates up to logarithmic factor as in (T3.2) when $1/v(n_0, \mathcal{E}) \geq n$.

3.4.3 Data-driven estimator

The estimator depends on two parameters and an optimal estimator can be obtained when they are chosen depending on unknowns such as s, σ or κ, etc. To make it practical, we show that a nearly minimax estimator can be obtained in a data-driven way by a type of Goldenshluger-Lepski method (see, *e.g.*, Goldenshluger and Lepski (2008, 2014); Lacour and Massart (2016)).

To gain insight on the data-driven choices $\widehat{N}$ and $\widehat{T}$ of N and T, let us sketch the proof when $\widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}} = f_{\boldsymbol{X}|\mathcal{X}}$ (hence we simply write $\mathcal{R}^W$). Consider $W = i_{[-R,R]}$. Let $\epsilon \in (0,1)$ and $q \in \{1, \infty\}$. We use the following upper bound on the risk: for all $f_{\alpha,\boldsymbol{\beta}} \in \mathcal{H}^{q,\phi,\omega}_{w,W}(l, M)$,

$$\mathcal{R}^W\left(\widehat{f}^{q,\widehat{N},\widehat{T},\epsilon}_{\alpha,\boldsymbol{\beta}}, f_{\alpha,\boldsymbol{\beta}}\right) \leq \frac{C}{2\pi}\int_{\epsilon\leq|t|} \mathbb{E}\left[\mathcal{L}^W_q\left(t, \widehat{N}(t), \widehat{T}\right)\right] dt + CM^2\widetilde{w}(\underline{a}), \tag{3.23}$$

where $\widetilde{w} := \mathbb{1}\{w \neq i_{[-\underline{a},\underline{a}]}\}/w$, and, for all $t \in \mathbb{R}$, $N \in \mathbb{N}_0^{\mathbb{R}}$, and $T' \in [0, \infty)$,

$$\mathcal{L}^W_q(t, N, T') := \left\| \left(\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right] - \widehat{F}^{q,N,T',0}_1\right)(t, \cdot_2)\right\|^2_{L^2(W^{\otimes p})}.$$

This is proved in the appendix using the Plancherel identity and (3.58). The latter is derived from Proposition 3. It is important that the first integral in (3.23) is restricted to $\{t \in \mathbb{R} : \epsilon \leq |t|\}$ by the considerations in Section 3.2.5. The upper bound in (3.23) when $\widehat{N}$ and $\widehat{T}$ are nonrandom is the one we use to obtain theorems 2 and 3. The aim of the selection rules is to obtain an upper bound on the right-hand side of (3.23) with a similar quantity but with arbitrary nonrandom $\widehat{N}$ and $\widehat{T}$.

Let us start with intuitions on the selection rule for $\widehat{N}$. We define a maximum $N^W_{\max,q}$ for the values that $\widehat{N}(t)$ can take. We set, for $N_0 \in \{0, \dots, N^W_{\max,q}\}$,

$$B_1(t, N_0) := \max_{N_0 \le N' \le N^W_{\max,q}} \left(\left\| \widehat{F}_1^{q,N',\infty,0}(t,\cdot) - \widehat{F}_1^{q,N_0,\infty,0}(t,\cdot) \right\|^2_{L^2(W^{\otimes p})} - \Sigma(t, N') \right)_+ ,$$

where $\Sigma(t, N')$ is a penalty term, called the "majorant" (see Goldenshluger and Lepski (2014)), which expression is given in equation (3.26) below. To explain the role of $B_1(t, N_0)$ and $\Sigma(t, N')$ in the choice of $\widehat{N}(t)$, consider $|t| \ge \epsilon$ and $N', N_0 \in \{0, \dots, N^W_{\max,q}\}$, $N' \ge N_0$. The majorant is chosen to compensate for the fluctuation of the squared $L^2(W^{\otimes p})$ norm of the statistic $b \mapsto \widehat{F}_1^{q,N',\infty,0}(t,b) - \widehat{F}_1^{q,N_0,\infty,0}(t,b)$. The term

$$\left(\left\| \widehat{F}_1^{q,N',\infty,0}(t,\cdot) - \widehat{F}_1^{q,N_0,\infty,0}(t,\cdot) \right\|^2_{L^2(W^{\otimes p})} - \Sigma(t, N') \right)_+$$

is a proxy for $\|F_1^{q,N',\infty,0}(t,\cdot) - F_1^{q,N_0,\infty,0}(t,\cdot)\|^2_{L^2(W^{\otimes p})}$. The maximum of the later is $\|\mathcal{F}_{1\text{st}}[f_{\alpha,\boldsymbol{\beta}}](t,\cdot) - F_1^{q,N_0,\infty,0}(t,\cdot)\|^2_{L^2(W^{\otimes p})}$. It is the square norm of the bias due to the cut-off at N_0. The majorant is an upper bound on the weighted integral of the variance of $\widehat{F}_1^{q,N_0,\infty,0}(t,\cdot)$. As a result, $\widehat{N}$ is selected as

$$\forall t \in \mathbb{R} \setminus (-\epsilon, \epsilon), \quad \widehat{N}(t) \in \operatorname*{argmin}_{0 \le N \le N^W_{\max,q}} \left(B_1(t, N) + c_1 \Sigma(t, N) \right), \tag{3.24}$$

where $c_1 = 1 + 1/(2+\sqrt{5})^2$.

In the same spirit, define a grid for the possible values of $\widehat{T}$,

$$\mathcal{T}_n := \{2^k : k = 1, \dots, K_{\max}\},$$

where $K_{\max} := \lfloor \zeta_0 \ln(n)/\ln(2) \rfloor$ and $\zeta_0 = 1/(1 + 4p(1 + \mathbb{1}\{W = \cosh(\cdot/R)\}))$. The choice of $\widehat{T}$ relies on, for all $T \in \mathbb{R}$ and $N \in \mathbb{N}_0^{\mathbb{R}}$,

$$B_2\left(T,N\right) := \max_{T' \in \mathcal{T}_n, T' \geq T} \left(\left\| \widehat{F}^{q,N,T',0} - \widehat{F}^{q,N,T,0} \right\|^2_{L^2(1 \otimes W^{\otimes p})} - \Sigma_2\left(T',N\right) \right)_+ ,$$

and on the majorant

$$\Sigma_2(T,N) := \int_{\epsilon \leq |t| \leq T} \Sigma(t, N(t)) dt.$$

Then, $\widehat{T}$ is selected by

$$\widehat{T} \in \underset{T \in \mathcal{T}_n}{\operatorname{argmin}} \left(B_2\left(T, \widehat{N}\right) + \Sigma_2\left(T, \widehat{N}\right) \right). \tag{3.25}$$

The expression of the majorant is as follows: for all $N \in \mathbb{N}_0^{\mathbb{R}}$, $N_0 \in \mathbb{N}_0$, and $t \neq 0$,

$$\begin{aligned} \Sigma\left(t, N_0\right) &:= 8(2+\sqrt{5})(1+2p_n)\frac{c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p \nu_q^W(x_0 t, N_0), \qquad (3.26) \\ \nu_q^W(t,N_0) &:= (N_0+1)^{k_q} Q_q^W(N_0)\left(1 \bigvee \frac{7e\pi(N_0+1)}{R\,|t|}\right)^{2k_q N_0 + p}, \\ Q_q^W(N_0) &:= \mathbb{1}\{q=\infty\}(2^p \mathbb{1}\{W = i_{[-R,R]}\} \\ &\qquad + \mathbb{1}\{W = \cosh(\cdot/R)\}) + \frac{(N_0+p-1)^{p-1}\mathbb{1}\{q=1\}}{(p-1)!}, \end{aligned}$$

and $p_n := 3 \bigvee 6(1+\zeta_0)\ln(n)$. The term $\nu_q^W(t, N_0)$ is an upper bound on the sum of the inverse of the square of the singular values up to N_0 (see Lemma 10). We also take $N_{\max,q}^W = \lfloor \underline{N_{\max,q}^W} \rfloor$, where $\underline{N_{\max,q}^W}$ satisfies $2k_q(\underline{N_{\max,q}^W} + k_q'/2)\ln(7e\pi\underline{N_{\max,q}^W}/(Rx_0\epsilon)) = \ln(n)$.

Let us explain how the definition (3.25) of $\widehat{T}$ yields an upper bound on the right-hand side of (3.23) by a quantity where $\widehat{T}$ can be replaced by an arbitrary T. By arguments in the proof of Lemma 5 for the first inequality and (3.25) for the second, we have, for all $T \in \mathcal{T}_n$,

$$\begin{aligned}
&\int_{\epsilon\le|t|} \mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),\widehat{T}\right)\right]dt\\
&\le \frac{2+\sqrt{5}}{\sqrt{5}}\int_{\epsilon\le|t|} \mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),T\right)\right]dt\\
&\quad+(2+\sqrt{5})\left(\mathbb{E}\left[B_2\left(\widehat{T},\widehat{N}\right)+\Sigma_2\left(\widehat{T},\widehat{N}\right)\right]+\mathbb{E}\left[B_2\left(T,\widehat{N}\right)+\Sigma_2\left(T,\widehat{N}\right)\right]\right)\\
&\le \frac{2+\sqrt{5}}{\sqrt{5}}\int_{\epsilon\le|t|} \mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),T\right)\right]dt+2(2+\sqrt{5})\mathbb{E}\left[B_2\left(T,\widehat{N}\right)+\Sigma_2\left(T,\widehat{N}\right)\right].
\end{aligned}$$

We then rely on an upper bound on $\mathbb{E}\left[B_2\left(T,\widehat{N}\right)\right]$ in the second term on the right-hand side. It involves a term proportional to the first one :

$$\mathbb{E}\left[B_2\left(T,\widehat{N}\right)\right]\le\left(1+\frac{2}{\sqrt{5}}\right)\int_{\epsilon\le|t|}\mathbb{E}\left[\mathcal{L}_q^W(t,\widehat{N}(t),T)\right]dt+O\left(\frac{1}{n}\right).$$

The $O(1/n)$ term is independent of T and $\widehat{N}$ and is obtained using a Talagrand's inequality. Similarly, (3.24) allows to obtain yet another upper bound which replaces $\widehat{N}$ by an arbitrary nonrandom N. We conclude because the final upper bound (3.95) has a similar form as the upper bound (3.56) appearing when we deal with nonrandom $\widehat{N}$ and $\widehat{T}$ in theorems 2 and 3.

The upper bounds below take the form

$$\sup_{\substack{f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l,M)\cap\mathcal{D}\\ f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}}}\frac{\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,\widehat{N},\widehat{T},\epsilon},f_{\alpha,\boldsymbol{\beta}}\right)}{r(n)^2}=\underset{\substack{v(n_0,\mathcal{E})/\delta(n_0)\le n^{-2}\ln(n)^{-p}\\ n_e\ge3}}{O_p}(1),\tag{3.27}$$

where the above O_p symbol means that the left-hand side doubly-indexed sequence of random variables, denoted for simplicity by $X_{n_0,n}$, is such that, for all $\epsilon>0$, there exists M such that $\mathbb{P}(|X_{n_0,n}|\ge M)\le\epsilon$ for all $(n_0,n)\in\mathbb{N}_0^2$ satisfying the condition underneath the O_p symbol. The results in this section are for $v(n_0,\mathcal{E})/\delta(n_0)\le n^{-2}\ln(n)^{-p}$, in which case $n_e=n$. We refer to (3.27') when we use the restriction $f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l)\cap\mathcal{D}$.

Theorem 4. *Let* $W=i_{[-R,R]}$, $\mathbb{S}_{\boldsymbol{\beta}}\subseteq[-R,R]^p$. *For all* $l,M,s,\sigma>0$, $q\in\{1,\infty\}$,

$\phi = 1 \vee |\cdot|^s$, *if*

(T4.1) $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, $\underline{a} = 1/\epsilon$, $w = 1 \vee |\cdot|$, *and* $\epsilon = 7e\pi/(Rx_0 \ln(n))$, *then (3.27) holds with* $r(n) = (\ln(n)/\ln_2(n))^{-\sigma}$*;*

(T4.2) $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa k\ln(1+k)})_{k\in\mathbb{N}_0}$, $\underline{a}$ *such that* $\mathbb{S}_\alpha \subseteq [-\underline{a}, \underline{a}]$, $w = i_{[-\underline{a},\underline{a}]}$, $\epsilon = 7e\pi/(Rx_0)$, *and* $s > (2p+1/2) \vee (\kappa(p+1)/(2k_q))$, *then (3.27') holds with* $r(n) = n^{-\kappa/(2\kappa+2k_q)}\ln(n)^{1/2+\Lambda(0)}$ *and* Λ *defined in (T2.2).*

The rate in (T4.2) matches, up to a logarithmic factor, the lower bound in Theorem (T1.1b) for model (3.8). (T4.2) relies on $\mathbb{S}_\alpha \subseteq [-\underline{a}, \underline{a}]$ because, else, the choice $\underline{a} = w^I(\omega^2_{\underline{N}})$ in Section 3.4.2 depends on the parameters of the smoothness class. However, it is possible to check that we can obtain the rate in (T2.2a) up to a $\sqrt{\ln(n)}$ factor when $\nu = 1$ for a choice of $\underline{a}$ independent of the parameters of the smoothness class.

When $W = \cosh(\cdot/R)$, we keep the rules (3.24) and (3.25) for $\widehat{N}$ and $\widehat{T}$ with the majorant Σ with parameters

$$\begin{aligned}
\nu_q^W(t, N_0) =& 2^{k_q}\left(\frac{\pi^2}{56}\right)^p Q_q^W(N_0)\left(\frac{7e^2\pi}{2R|t|}\right)^{2k_qN_0+p} \mathbb{1}\left\{|t| \le \frac{\pi}{4R}\right\} \\
&+ 2^p\left(\frac{2eR\,|t|}{\pi}\right)^{k_q} Q_q^W(N_0)\exp\left(\frac{\pi k_q(N_0+k_q')}{2R\,|t|}\right)\mathbb{1}\left\{|t| > \frac{\pi}{4R}\right\}, \\
\underline{N_{\max,q}^W} =& \frac{\ln(n)}{2k_q}\mathbb{1}\left\{\epsilon = \frac{\pi}{4Rx_0}\right\} + \frac{\ln(n)}{2k_q\ln\left(7e^2\pi/(2Rx_0\epsilon)\right)}\mathbb{1}\left\{\epsilon < \frac{\pi}{4Rx_0}\right\}.
\end{aligned}$$

Theorem 5. *Let* $W = \cosh(\cdot/R)$*. For all* $l, M, s, \sigma > 0$, $q \in \{1, \infty\}$, $\phi = 1 \vee |\cdot|^s$, *if*

(T5.1) $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$, $\underline{a} = 1/\epsilon$, $w = 1 \vee |\cdot|$, *and* $\epsilon = 7e^2\pi/(2Rx_0\ln(n))$, *then (3.27) holds with* $r(n) = (\ln(n)/\ln_2(n))^{-\sigma}$*;*

(T5.2) $(\omega_k)_{k\in\mathbb{N}_0} = (e^{\kappa k})_{k\in\mathbb{N}_0}$, $\underline{a}$ *such that* $\mathbb{S}_\alpha \subseteq [-\underline{a}, \underline{a}]$, $w = i_{[-\underline{a},\underline{a}]}$, $\epsilon = \pi/(4Rx_0)$, *and* $s > 4p+1/2$, *then (3.27') holds with* $r(n) = n^{-\kappa/(2\kappa+2k_q)}\ln(n)^{1/2+(p-1)\kappa/(2q(\kappa+k_q))}$.

The risk is different for the lower and upper bounds in theorems 1 and 5. However, by (3.15), we obtain the same rate up to logarithmic factors for the risk involving the weight $\cosh(\cdot/R)$. Theorem 1 and (T5.2) with $q = \infty$ show that $\widehat{f}^{q,\widehat{N},\widehat{T},\epsilon}_{\alpha,\boldsymbol{\beta}}$ is adaptive.

3.5 Simulations

Let $p=1$, $q=\infty$, and $(\alpha,\beta)^\top=\xi_1 D+\xi_2(1-D)$ with $\mathbb{P}(D=1)=\mathbb{P}(D=0)=0.5$. The law of X is a truncated normal based on a normal of mean 0 and variance 2.5 and truncated to $\mathcal{X}$ with $x_0=1.5$. The laws of ξ_1 and ξ_2 are either: (Case 1) truncated normals based on normals with means $\mu_1=\left(\begin{smallmatrix}-2\\3\end{smallmatrix}\right)$ and $\mu_2=\left(\begin{smallmatrix}3\\0\end{smallmatrix}\right)$, same covariance $\left(\begin{smallmatrix}2&1\\1&2\end{smallmatrix}\right)$, and truncated to $[-6,6]^{p+1}$ or (Case 2) nontruncated. The estimators in Hoderlein et al. (2010); Holzmann and Meister (2019) cannot be used in this context. Case (1) could be treated with Beran and Millar (1994) in this particular case where $p=1$. This requires choosing many parameters and Beran and Millar (1994) only provides a consistency result. Case (2) allows for unbounded errors as in usual linear regression models (thus is very different from tomography problems) and no other nonparametric method is available up to our knowledge.

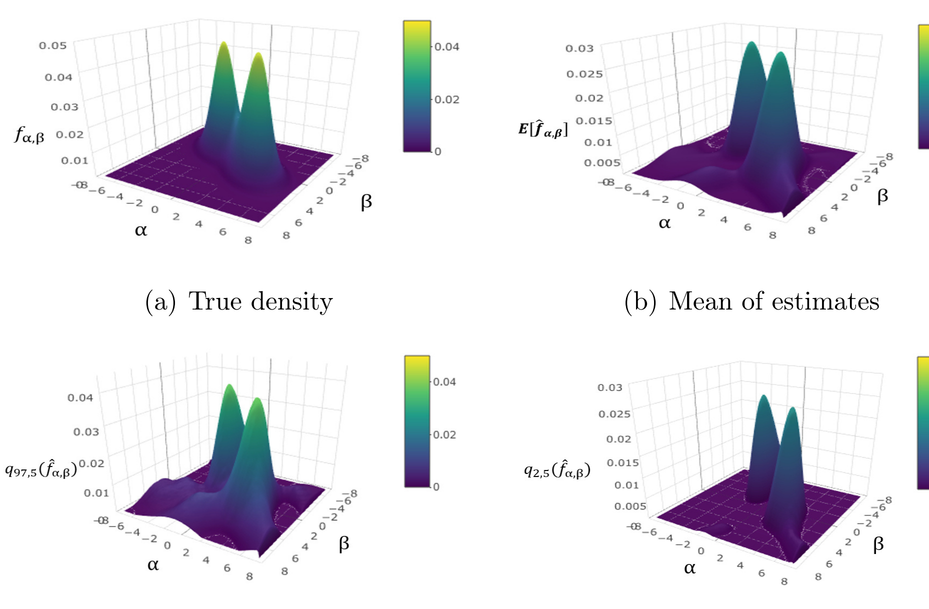
Table 3.1 compares $\mathbb{E}[\|\widehat{f}_{\alpha,\beta}^{\infty,\widehat{N},\widehat{T},\epsilon}-f_{\alpha,\beta}\|^2_{L^2([-7.5,7.5]^2)}]$ and $\min_{T\in\mathcal{T}_n,N\in\mathcal{N}_{n,H}}\mathbb{E}[\|\widehat{f}_{\alpha,\beta}^{\infty,N,T,\epsilon}-f_{\alpha,\beta}\|^2_{L^2([-7.5,7.5]^2)}]$ (risk of the oracle) for cases 1 and 2. The Monte-Carlo experiment uses 1000 simulations.

	$W=i_{[-7.5,7.5]}$, Case 1			$W=\cosh(\cdot/7.5)$, Case 2		
MISE	$n=300$	$n=500$	$n=1000$	$n=300$	$n=500$	$n=1000$
data-driven	0.092	0.086	0.083	0.089	0.087	0.085
oracle	0.091	0.086	0.082	0.088	0.087	0.085

Table 3.1: Risk

Figure 6-4 (resp. Figure 6-3) displays summaries of the law of the estimator for $W=i_{[-7.5,7.5]}$ (resp. $W=\cosh(\cdot/7.5)$) in Case 1 (resp. Case 2) and $n=1000$. As standard in the literature (see, *e.g.*, Comte et al. (2013); Dion (2014)), the multiplicative constant appearing in Σ is in practice calibrated from a simulation study. $\widehat{f}_{X|X\in\mathcal{X}}$ is obtained with the same data to illustrate that sample splitting is unnecessary in practice and only used for the mathematical argument. For $\widehat{f}_{X|X\in\mathcal{X}}$ we use a Gaussian kernel density estimator using the R package ks and the multivariate plug-in bandwidth selector of Wand and Jones (1994). ϵ is chosen as in (T4.1) and (T5.1) respectively for Case 1 and Case 2. The estimator requires the SVD of $\mathcal{F}_c$. By

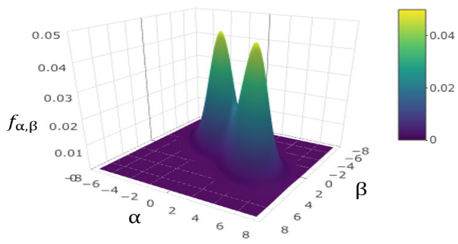
Proposition 4, we have $g_m^{W(\cdot/R),c} = g_m^{W,Rc}$ for all $m \in \mathbb{N}_0$. When $W = i_{[-1,1]}$, the first coefficients of the decomposition on the Legendre polynomials are obtained by solving for the eigenvectors of two tridiagonal symmetric Toeplitz matrices (see Section 2.6 in Osipov et al. (2013)). When $W = \cosh$, we refer to Section 7 in Gaillac and Gautier (2019*a*). We use the image of $g_m^{W,Rc}$ by the adjoint of $\mathcal{F}_c$ (see Appendix 3.6.1) and that $\varphi_m^{W,Rc}$ has norm 1 to get the rest of the SVD. We obtain the Fourier inverse by fast Fourier transform. We use a resolution of 2^{13}, which appears on simulations to realise a good trade-off between computational time and precision. For more details about the implementation, we refer to the vignette Gaillac and Gautier (2019*d*) of the package RandomCoefficients.



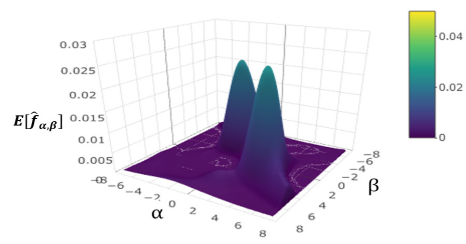
(a) True density (b) Mean of estimates

(c) 97.5% quantile of estimates (d) 2.5% quantile of estimates

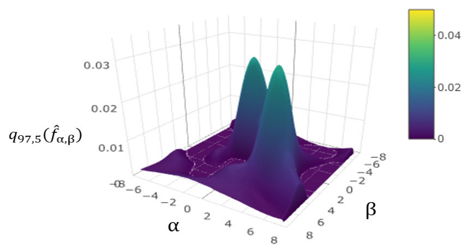
Figure 3-1: Case 1, $W = i_{[-7.5,7.5]}$



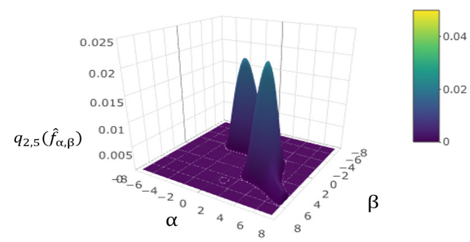
(a) True density



(b) Mean of estimates



(c) 97.5% quantile of estimates



(d) 2.5% quantile of estimates

Figure 3-2: Case 2, $W = \cosh(\cdot/7.5)$

3.6 Proofs of the main results

3.6.1 Notations and preliminaries

$\Re$ and $\Im$ denote the real and imaginary parts. For a differentiable function f of real variables, $f^{(\boldsymbol{m})}$ denotes $\prod_{j=1}^{d} \frac{\partial^{m_j}}{\partial x_j^{m_j}} f$. $C^{\infty}\left(\mathbb{R}^d\right)$ is the space of infinitely differentiable functions. Abusing notations, we sometimes use $\mathcal{F}_c[f]$ for the function in $L^2(\mathbb{R})$ and $\mathcal{E}xt[f]$ assigns the value 0 outside $[-1,1]^d$. Denote by $\Pi : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ such that $\Pi f(\boldsymbol{x}) = f(-\boldsymbol{x})$ and by

$$\begin{array}{cccc} \mathcal{C}_c : & L^2\left(\mathbb{R}^d\right) & \to & L^2\left(\mathbb{R}^d\right) \\ & f & \to & |c|^d f(c\,\cdot). \end{array} \tag{3.28}$$

Because $\mathcal{F}_c = \mathcal{F}\mathcal{C}_{1/c} = (1/|c|)\mathcal{C}_c\mathcal{F}$, $\Pi\mathcal{F}_c = \mathcal{F}_c\Pi$, $\mathcal{F}_c^* = (1/W)\Pi\mathcal{F}_c\mathcal{E}xt$, and W is even, we obtain $\mathcal{F}_c^* = \Pi\left((1/W)\mathcal{F}_c\mathcal{E}xt\right)$ and

$$\begin{aligned} \mathcal{F}_c\mathcal{F}_c^* &= \Pi\mathcal{F}_c\left(\frac{1}{W}\mathcal{F}_c\mathcal{E}xt\right) \\ &= \frac{2\pi}{|c|}\mathcal{F}^I\left(\mathcal{C}_{1/c}\left(\frac{1}{W}\mathcal{C}_c\mathcal{F}\mathcal{E}xt\right)\right) \\ &= 2\pi\mathcal{F}^I\left(\mathcal{C}_{1/c}\left(\frac{1}{W}\right)\mathcal{F}\mathcal{E}xt\right). \end{aligned}$$

The operator $\mathcal{Q}_c^W = (|c|/(2\pi))\mathcal{F}_c\mathcal{F}_c^*$ is a compact positive definite self-adjoint operator (see Osipov et al. (2013) and Widom (1964) for the two choices of W). Its eigenvalues in decreasing order repeated according to multiplicity are denoted by $\left(\rho_m^{W,c}\right)_{m\in\mathbb{N}_0}$ and a basis of eigenfunctions by $\left(g_m^{W,c}\right)_{m\in\mathbb{N}_0}$. The other elements of the SVD are $\sigma_m^{W,c} = \sqrt{2\pi\rho_m^{W,c}/|c|}$ and $\varphi_m^{W,c} = \mathcal{F}_c^* g_m^{W,c}/\sigma_m^{W,c}$. We denote, for all $m \in \mathbb{N}_0$, by ψ_m^c the function $g_m^{i_{[-1,1]},c}$ and $\mu_m^c = i^m\sigma_m^{i_{[-1,1]},c}$. Because $\psi_m^c = \mathcal{F}_c(\mathcal{E}xt[\psi_m^c])/\mu_m^c$ in $L^2([-1,1])$, ψ_m^c can be extended as an entire function which we denote with the same notation. Using the injectivity of $\mathcal{F}_c$ (see the proof of Proposition 1), we have $\varphi_m^{i_{[-1,1]},c} = i^{-m}\mathcal{E}xt[\psi_m^c]$.

We make use of

$$\forall a,b>0,\ \sup_{t\geq 1}\frac{\ln(t)^a}{t^b}=\left(\frac{a}{eb}\right)^a, \tag{3.29}$$

$$\forall c>0,\ \forall a,b\in\mathbb{R},\ ab\leq\frac{a^2}{2c}+\frac{b^2c}{2}. \tag{3.30}$$

Expectations are conditional on $\mathcal{G}_{n_0}$ when $f_{\boldsymbol{X}|\mathcal{X}}$ is unknown and we rely on $\mathcal{G}_{n_0}$ to estimate it. We remove the conditioning in the notations for simplicity.

3.6.2 Proofs of Proposition 1, 2, and 3

Proof of Proposition 1. The first assertion comes from the fact that W is bounded from below. The second uses Theorem IX.13 in Reed and Simon (1980) which implies that, for all $c\neq 0$, $\mathcal{F}_c$ defined in (3.3) is injective. We now show that $\mathcal{K}$ is continuous at 0. Let $f\in L^2\left(w\otimes W^{\otimes p}\right)$. The change of variables, the Plancherel identity, and the lower bounds on the weights yield

$$\|\mathcal{K}[f]\|^2_{L^2(\mathbb{R}\times[-1,1]^p)}\leq\int_{\mathbb{R}^{p+1}}|\mathcal{F}[f](t,\boldsymbol{v})|^2(t,\boldsymbol{v})dtd\boldsymbol{v}\leq\left(\frac{2\pi}{W(0)}\right)^p\|f\|^2_{L^2(w\otimes W^{\otimes p})}.$$

Let $w=1$. We exhibit a bounded sequence $(f_k)_{k\in\mathbb{N}_0}$ in $L^2(1\otimes W^{\otimes p})$ for which there does not exist a convergent subsequence of $(\mathcal{K}\left[f_k\right])_{k\in\mathbb{N}_0}$. Take v_0 such that $\mathrm{supp}(v_0)\subset[1,2]$, $\|v_0\|_{L^2(\mathbb{R})}=1$ and, for all $k\in\mathbb{N}_0$ and $(a,\boldsymbol{b}^\top)^\top\in\mathbb{R}^{p+1}$, $v_k(\cdot)=2^{-k/2}v_0(2^{-k}\cdot)$ and $f_k(a,\boldsymbol{b})=\mathcal{F}^I\left[v_k(\cdot)\varphi_{\underline{\mathbf{0}}}^{W,x_0\cdot}(\boldsymbol{b})\right](a)$. $(f_k)_{k\in\mathbb{N}_0}$ is bounded by the Plancherel identity and

$$\|f_k\|^2_{L^2(1\otimes W^{\otimes p})}=\frac{1}{2\pi}\int_{\mathbb{R}}v_k(t)^2\int_{\mathbb{R}^p}\left|\varphi_{\underline{\mathbf{0}}}^{W,x_0t}(\boldsymbol{b})\right|^2W^{\otimes p}(\boldsymbol{b})dtd\boldsymbol{b}\leq\frac{1}{2\pi}.$$

Using $\mathcal{K}\left[f_k\right](\cdot,\cdot_2)=\sigma_{\underline{\mathbf{0}}}^{W,x_0\cdot}v_k(\cdot)g_{\underline{\mathbf{0}}}^{W,x_0\cdot}(\cdot_2)\left|x_0\cdot\right|^{p/2}$, $c\in(0,\infty)\mapsto\rho_0^{W,c}$ is nondecreasing (by Lemma 3 in Gaillac and Gautier (2019*a*) which holds for all W satisfying (H1.2)), and using, for all $j \in \mathbb{N}_0$, $\|v_j\|_{L^2(\mathbb{R})} = 1$, we obtain, for all $(j,k) \in \mathbb{N}_0^2$, $j < k$,

$$\begin{aligned}\|\mathcal{K}[f_j] - \mathcal{K}[f_k]\|^2_{L^2(\mathbb{R}\times[-1,1]^p)} &\geq \rho_{\underline{\mathbf{0}}}^{W,2^j x_0} (2\pi)^p \int_{\mathbb{R}} \left(v_j(t)^2 + v_k(t)^2\right) dt \\ &\geq 2(2\pi)^p \rho_{\underline{\mathbf{0}}}^{W,x_0} > 0,\end{aligned}$$

so $\mathcal{K}$ is not compact. □

Proof of Proposition 2. This holds by Theorem 15.16 in Kress (1999) and the injectivity of $\mathcal{F}_c$. □

Proof of Proposition 3. Take $f \in L^2(\mathbb{R})$ and start by showing that $\mathcal{I}_{\underline{a},\epsilon}[f] \in L^2([-\epsilon,\epsilon])$. The terms $1 - \rho_m^{i_{[-1,1]},\underline{a}\epsilon}$ in the denominator of (3.9) are nonzero because $\left(\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\right)_{m\in\mathbb{N}_0}$ is nonincreasing and $\rho_0^{i_{[-1,1]},\underline{a}\epsilon} < 1$ (see (3.49) in Osipov et al. (2013)). Using that $\left(g_m^{i_{[-1,1]},\underline{a}\epsilon}(\cdot/\epsilon)/\sqrt{\epsilon}\right)_{m\in\mathbb{N}_0}$ is a basis of $L^2([-\epsilon,\epsilon])$, that $\left(\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\right)_{m\in\mathbb{N}_0}$ is nonincreasing, and the Cauchy-Schwarz inequality for the first inequality, using $\sum_{m\in\mathbb{N}_0} \rho_m^{i_{[-1,1]},\underline{a}\epsilon} = 2\underline{a}\epsilon/\pi$ (see (3.55) in Osipov et al. (2013)) and $\left\|g_m^{i_{[-1,1]},\underline{a}\epsilon}\right\|^2_{L^2(\mathbb{R})} = 1/\rho_m^{i_{[-1,1]},\underline{a}\epsilon}$ (see (3) in Bonami and Karoui (2014a)) for the second yield

$$\begin{aligned}&\sum_{m\in\mathbb{N}_0} \left(\frac{\rho_m^{i_{[-1,1]},\underline{a}\epsilon}}{\left(1-\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\right)\epsilon}\right)^2 \left|\left\langle f, g_m^{i_{[-1,1]},\underline{a}\epsilon}\left(\frac{\star}{\epsilon}\right)\right\rangle_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}\right|^2 \left\|g_m^{i_{[-1,1]},\underline{a}\epsilon}\left(\frac{\cdot}{\epsilon}\right)\right\|^2_{L^2([-\epsilon,\epsilon])} \\ &\leq \frac{\|f\|^2_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}}{\left(1-\rho_0^{i_{[-1,1]},\underline{a}\epsilon}\right)^2} \sum_{m\in\mathbb{N}_0} \left(\rho_m^{i_{[-1,1]},\underline{a}\epsilon}\right)^2 \left\|g_m^{i_{[-1,1]},\underline{a}\epsilon}\right\|^2_{L^2(\mathbb{R})} \leq \frac{2\underline{a}\epsilon\,\|f\|^2_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}}{\pi\left(1-\rho_0^{i_{[-1,1]},\underline{a}\epsilon}\right)^2}. \qquad (3.31)\end{aligned}$$

Let us now show the second statement. Take $\epsilon > 0$ and $g \in PW(\underline{a})$. Let $(\alpha_m)_{m\in\mathbb{N}}$ be the sequence of coefficients of $g(\epsilon\cdot) \in PW(\underline{a}\epsilon)$ on the complete orthogonal system $\left(g_m^{i_{[-1,1]},\underline{a}\epsilon}\right)_{m\in\mathbb{N}_0}$. Because $\left(g_m^{i_{[-1,1]},\underline{a}\epsilon}\right)_{m\in\mathbb{N}_0}$ is a basis of $L^2([-1,1])$, we have

$$\sum_{m\in\mathbb{N}_0} \alpha_m g_m^{i_{[-1,1]},\underline{a}\epsilon} = g(\epsilon\cdot)\mathbb{1}\{|\cdot| \geq 1\} + \sum_{m\in\mathbb{N}_0} \alpha_m g_m^{i_{[-1,1]},\underline{a}\epsilon}\mathbb{1}\{|\cdot| \leq 1\}.$$

We identify the coefficients by taking the Hermitian product in $L^2(\mathbb{R})$ with $g_m^{i_{[-1,1]},\underline{a}\epsilon}$

and obtain $\mathcal{I}_{\underline{a},\epsilon}[g]=g$ in $L^2(\mathbb{R})$ and, for all $f,h\in L^2(\mathbb{R})$,

$$\|f-\mathcal{I}_{\underline{a},\epsilon}[h]\|^2_{L^2([-\epsilon,\epsilon])}\le 2\left(\|f-\mathcal{P}_{\underline{a}}[f]\|^2_{L^2([-\epsilon,\epsilon])}+\|\mathcal{I}_{\underline{a},\epsilon}[\mathcal{P}_{\underline{a}}[f]-h]\|^2_{L^2([-\epsilon,\epsilon])}\right). \quad (3.32)$$

Replacing f by $\mathcal{P}_{\underline{a}}[f]-h$ in (3.31) yields

$$\|\mathcal{I}_{\underline{a},\epsilon}[\mathcal{P}_{\underline{a}}[f]-h]\|^2_{L^2([-\epsilon,\epsilon])}\le\frac{C_0(\underline{a}\epsilon)}{2}\|\mathcal{P}_{\underline{a}}[f]-h\|^2_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}. \quad (3.33)$$

Using (3.32) and (3.33) for the first display, $\mathcal{P}_{\underline{a}}[f]-h=(\mathcal{P}_{\underline{a}}[f]-f)+(f-h)$ and the Jensen inequality for the second display, we obtain

$$\begin{aligned}&\|f-\mathcal{I}_{\underline{a},\epsilon}[h]\|^2_{L^2([-\epsilon,\epsilon])}\\ &\le 2\|f-\mathcal{P}_{\underline{a}}[f]\|^2_{L^2([-\epsilon,\epsilon])}+C_0(\underline{a}\epsilon)\|\mathcal{P}_{\underline{a}}[f]-h\|^2_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}\\ &\le 2(1+C_0(\underline{a}\epsilon))\|f-\mathcal{P}_{\underline{a}}[f]\|^2_{L^2(\mathbb{R})}+2C_0(\underline{a}\epsilon)\|f-h\|^2_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}. \quad \square\end{aligned}$$

3.6.3 Lower bounds

We denote by $\mathbb{P}_j$ the law of density $f_{j,n}$ and by $\mathbb{P}_{j,n}$ the law of an iid sample of size n, and use

$$\inf_{\widehat{f}}\ \sup_{f\in\mathcal{H}}\mathbb{E}\left[\left\|\widehat{f}-f\right\|^2_{L^2(\mathbb{R}^{p+1})}\right]\ge\inf_{\widehat{f}}\ \max_{f_{j,n}\in\mathcal{H},\ j\in\{1,2\}}\mathbb{E}\left[\left\|\widehat{f}-f_{j,n}\right\|^2_{L^2(\mathbb{R}^{p+1})}\right]$$

and the next lemma (see Theorem 2.2, (2.5), and (2.9) in Tsybakov (2008)).

Lemma 1. *If there exists $\xi<\sqrt{2}$ such that*

(i) $\forall j\in\{1,2\},\ f_{j,n}\in\mathcal{H}$,

(ii) $\|f_{1,n}-f_{2,n}\|^2_{L^2(\mathbb{R}^{p+1})}\ge 4r(n)^2>0$,

(iii) $\chi_2(\mathbb{P}_{2,n},\mathbb{P}_{1,n})\le\xi^2$ *or* $K(\mathbb{P}_{2,n},\mathbb{P}_{1,n})\le\xi^2$,

then we have

$$\frac{1}{r(n)^2}\inf_{\widehat{f}}\ \max_{f_{j,n}\in\mathcal{H},\ j\in\{1,2\}}\mathbb{E}\left[\left\|\widehat{f}-f_{j,n}\right\|^2_{L^2(\mathbb{R}^{p+1})}\right]\ge\frac{1}{2}\left(\frac{e^{-\xi^2}}{2}\bigvee\left(1-\frac{\xi}{\sqrt{2}}\right)\right).$$

The proofs of the lower bounds consist in defining $f_{j,n} \in \mathcal{H}$, for $j \in \{1,2\}$ as a function of parameters, establishing the conditions on the parameters so that $f_{j,n}$ satisfy the three conditions (i)-(iii) of Lemma 1, and finally choosing the value of these parameters as a function of n to deduce the lower bounds.

Proof of (T1.1a). For $j = 1,2$, $f_{j,n}$ is a possible $f_{\alpha,\boldsymbol{\beta}}$, $(b^j_{\boldsymbol{m}})_{\boldsymbol{m}\in\mathbb{N}_0^p}$ the sequence of its coefficients (see Section 3.2.6), Steps 1-3 give conditions under which (i)-(iii) in Lemma 1 are satisfied when $f_{1,n} := f_0$ and

$$f_{2,n} := f_0 + F,\ \ f_0(a,\boldsymbol{b}) := \frac{1}{\pi\tau\left(1+(a/\tau)^2\right)}\frac{\mathbb{1}\{|\boldsymbol{b}|_\infty \le R\}}{(2R)^{p/2}}, \tag{3.34}$$

for all $(a,\boldsymbol{b}) \in \mathbb{R}^{p+1}$,

$$F(a,\boldsymbol{b}) := \gamma\mathcal{F}^I_{\rm 1st}\left[\left(\frac{c(|\star|)}{2\pi}\right)^{p/2}\lambda(\star)\psi^{Rc(\star)}_{\widetilde{\boldsymbol{N}}(q)}\left(\frac{\boldsymbol{b}}{R}\right)\right](a)\mathbb{1}\{|\boldsymbol{b}|_\infty \le R\}, \tag{3.35}$$

for all $U/2 \le |t| \le U$

$$\lambda(t) := \exp\left(1 - \frac{1}{1-(4|t|-3U)^2/U^2}\right),\ \text{else } \lambda(t) := 0, \tag{3.36}$$

$$\widetilde{\boldsymbol{N}}(1) := \left(N, \underline{\boldsymbol{N(Rx_0U)}^\top}\right)^\top,\ \widetilde{\boldsymbol{N}}(\infty) := \underline{\boldsymbol{N}} \in \mathbb{N}^p, \tag{3.37}$$

$N(Rx_0U) := \lceil H(Rx_0U)\rceil$, for H from Section 3.7.2, n large enough, N (odd), γ, $\tau \ge 1$, and U from Step 4 and such that $N \ge N(Rx_0U)$, hence $N \ge Rx_0U \vee 2$ by the discussion before Lemma 14. Note $\|\lambda\|_{L^\infty(\mathbb{R})} \le 1$.

Step 1.1. We prove that $f_{1,n}$ and $f_{2,n}$ are nonnegative when

$$\gamma U N^{k_q/2}\frac{((Rx_0U)/\pi)^{\frac{p}{2}}}{1+p/2}\left(\frac{5}{4}\right)^{\frac{k_q}{2}}\left(\frac{5}{4}N(Rx_0U)\right)^{\frac{p-1}{2q}} \le \frac{1}{\tau+1/\tau}, \tag{3.38}$$

$$\gamma U N^{k_q/2}\frac{2^{\frac{p}{2}}}{2}C_8(Rx_0U,p,U)N^2 \le \frac{1}{\tau+1/\tau}, \tag{3.39}$$

where $C_8(Rx_0U,p,U)$ is defined in Lemma 18. Let $(a,\boldsymbol{b}) \in \mathbb{R}\times[-R,R]^p$. We show that (3.38) and (3.39) yield $f_0(a,\boldsymbol{b}) \ge |F(a,\boldsymbol{b})|$ which ensures that $f_{2,n}$ is nonnegative.

(3.38) yields the result when $|a|<1$ because, by the third assertion in Lemma 15,

$$\begin{aligned}|F(a,\boldsymbol{b})| \leq & \frac{\gamma}{2\pi}\left(\frac{x_0}{2\pi}\right)^{p/2}\left(N+\frac{1}{2}\right)^{k_q/2}(N(Rx_0U)+1/2)^{(p-1)/(2q)}\int_{\mathbb{R}}|t|^{p/2}\lambda(t)dt \\ \leq & \frac{\gamma U}{\pi(1+p/2)}\left(\frac{Ux_0}{2\pi}\right)^{p/2}\left(\frac{5}{4}N\right)^{k_q/2}\left(\frac{5}{4}N(Rx_0U)\right)^{(p-1)/(2q)}. \end{aligned} \tag{3.40}$$

Because $t\mapsto \psi_{\widetilde{\boldsymbol{N}}(q)}^{Rx_0t}(\boldsymbol{b}/R)$ is analytic (see Fuchs (1964) page 320), the function

$$t\mapsto\left(\frac{x_0\,|t|}{2\pi}\right)^{p/2}\lambda(t)\psi_{\widetilde{\boldsymbol{N}}(q)}^{Rx_0t}\left(\frac{\boldsymbol{b}}{R}\right)\in C^\infty(\mathbb{R})$$

and its derivatives have compact support. By integration by parts, we obtain, for all $a\neq 0$,

$$|F(a,\boldsymbol{b})|\leq\frac{\gamma}{\pi a^2R^{p/2}}\int_{U/2}^{U}\left|\frac{\partial^2}{\partial t^2}\left(\left(\frac{Rx_0t}{2\pi}\right)^{p/2}\lambda(t)\psi_{\widetilde{\boldsymbol{N}}(q)}^{Rx_0t}\left(\frac{\boldsymbol{b}}{R}\right)\mathbb{1}\{|\boldsymbol{b}|_\infty\leq R\}\right)\right|dt.$$

The result when $|a|\geq 1$ is obtained by $1+(a/\tau)^2\leq(1+1/\tau^2)a^2$, so by (3.39), $\gamma UC_8(Rx_0U,p,U)N^{2+k_q/2}/(2a^2)\leq 1/\left(2^{p/2}\tau(1+(a/\tau)^2)\right)$, and by Lemma 18, for all $(a,\boldsymbol{b})\in\mathbb{R}^{p+1}$ such that $|a|\geq 1$,

$$|F(a,\boldsymbol{b})|\leq\frac{\gamma UC_8(Rx_0U,p,U)}{2\pi a^2R^{p/2}}N^{2+k_q/2}\mathbb{1}\{|\boldsymbol{b}|_\infty\leq R\}. \tag{3.41}$$

$f_{1,n}=f_0$ has integral 1 and so has $f_{2,n}$ by Fubini's theorem and that ψ_N^c is odd when N is odd.

Step 1.2. We give conditions for $f_{1,n},f_{2,n}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l)$. By (3.34)-(3.35), and because, by Step 1.1, for all $(a,\boldsymbol{b})\in\mathbb{R}^{p+1}$, $f_{2,n}(a,\boldsymbol{b})^2\leq 4f_{1,n}(a,\boldsymbol{b})^2$, $f_{1,n}$ and $f_{2,n}$ belong to $L^2\left(w\otimes i_{[-R,R]}^{\otimes p}\right)$. Let us show that $f_{2,n}$, hence $f_{1,n}$ ($f_{2,n}$ with $\gamma=0$), satisfy the first condition in $\mathcal{H}_{w,W}^{q,\phi,\omega}(l)$ if

$$2\left(\int_0^\infty\phi(t)^2e^{-2\tau t}dt+\gamma^2\left(\frac{Rx_0U}{2\pi}\right)^p\frac{\phi(U)^2U}{p+1}\right)\leq\pi^2 \tag{3.42}$$

$$\frac{C_{12}(\sigma,p)}{\tau p^{2\sigma/q}}+\gamma^2\frac{2Up^{2\sigma/q}N^{2\sigma}}{p+1}\left(\frac{Rx_0U}{2\pi}\right)^p\leq\pi^2. \tag{3.43}$$

Let $\boldsymbol{m} \in \mathbb{N}_0^p$ and $c_{\boldsymbol{m}}^P(t) := \left\langle 1/2^{p/2}, \psi_{\boldsymbol{m}}^{Rx_0t} \right\rangle_{L^2([-1,1]^p)}$. By Proposition 4 (iii), change of variables, and for all $t \in \mathbb{R}$, $\mathcal{F}_{\text{1st}}\left[f_0(\cdot,\star)\right](t) = e^{-|t|\tau}\mathbb{1}\{|\star|_\infty \le R\}/(2R)^{p/2}$, we have

$$b_{\boldsymbol{m}}^2(t) = \frac{1}{i^{|\boldsymbol{m}|_1}}\left(e^{-\tau|t|}c_{\boldsymbol{m}}^P(t) + \gamma\mathbb{1}\{\boldsymbol{m} = \widetilde{\boldsymbol{N}}(q)\}\left(\frac{Rx_0\,|t|}{2\pi}\right)^{p/2}\lambda(t)\right). \tag{3.44}$$

Because $\left(\psi_{\boldsymbol{m}}^{Rx_0t}\right)_{\boldsymbol{m}\in\mathbb{N}_0^p}$ is an orthonormal basis, we have

$$\forall t \neq 0,\ \sum_{\boldsymbol{m}\in\mathbb{N}_0^p}\left|b_{\boldsymbol{m}}^2(t)\right|^2 \le 2\left(e^{-2\tau|t|} + \gamma^2\left(\frac{Rx_0\,|t|}{2\pi}\right)^p\lambda(t)^2\right). \tag{3.45}$$

The first part of the first condition in $\mathcal{H}_{w,W}^{q,\phi,\omega}(l)$ holds by (3.42) and because, by (3.45),

$$\sum_{\boldsymbol{m}\in\mathbb{N}_0^p}\int_{\mathbb{R}}\phi(t)^2\left|b_{\boldsymbol{m}}^2(t)\right|^2dt \le 4\left(\int_0^\infty\frac{\phi(t)^2}{e^{2\tau t}}dt + \gamma^2\left(\frac{Rx_0}{2\pi}\right)^p\int_{U/2}^U\phi(t)^2t^p\lambda^2(t)dt\right).$$

The second part of the first condition holds by (3.43) and because, by (3.44) and Lemma 19, for all $\tau \ge \left(3e^{1/2}Rx_0/8\right)\vee(1/2)$ and $N \ge N(Rx_0U)$,

$$\begin{aligned}\sum_{\boldsymbol{m}\in\mathbb{N}_0^p}|\boldsymbol{m}|_q^{2\sigma}\int_{\mathbb{R}}\left|b_{\boldsymbol{m}}^2(t)\right|^2dt &\le 2\int_{\mathbb{R}}e^{-2\tau|t|}\sum_{\boldsymbol{m}\in\mathbb{N}_0^p}|\boldsymbol{m}|_q^{2\sigma}\left(c_{\boldsymbol{m}}^P(t)\right)^2dt \\ &\quad + 2\gamma^2\left(\frac{Rx_0}{2\pi}\right)^p\left|\widetilde{\boldsymbol{N}}(q)\right|_q^{2\sigma}\int_{\mathbb{R}}\lambda^2(t)\,|t|^p\,dt \\ &\le 2\left(\frac{C_{12}(\sigma,p)}{\tau p^{2\sigma/q}} + \gamma^2\frac{2Up^{2\sigma/q}N^{2\sigma}}{p+1}\left(\frac{Rx_0U}{2\pi}\right)^p\right).\end{aligned}$$

Step 2. (ii) holds with $4r(n)^2 = \gamma^2\left(Rx_0/(2\pi)\right)^p\int_{U/2}^U t^p\lambda(t)^2dt/\pi$.

Step 3. (ii) page 97 in Tsybakov (2008) yields $\chi_2(\mathbb{P}_{2,n},\mathbb{P}_{1,n}) = \left(1+\chi_2\left(\mathbb{P}_2,\mathbb{P}_1\right)\right)^n - 1$ so

$$\chi_2(\mathbb{P}_{2,n},\mathbb{P}_{1,n}) = n\int_0^{\chi_2(\mathbb{P}_2,\mathbb{P}_1)}(1+u)^{n-1}du \le n\chi_2\left(\mathbb{P}_2,\mathbb{P}_1\right)\exp\left((n-1)\chi_2\left(\mathbb{P}_2,\mathbb{P}_1\right)\right).$$

Thus, if $\chi_2(\mathbb{P}_2,\mathbb{P}_1)\le 1/n$, we obtain $\chi_2(\mathbb{P}_{2,n},\mathbb{P}_{1,n})\le en\chi_2(\mathbb{P}_2,\mathbb{P}_1)$. We have

$$\chi_2(\mathbb{P}_2,\mathbb{P}_1)=\int_{\mathbb{S}_{\boldsymbol{X}}}\int_{\mathbb{R}}\frac{f_{\boldsymbol{X}}(\boldsymbol{x})\left(f^1_{Y|\boldsymbol{X}}(y|\boldsymbol{x})-f^2_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\right)^2}{f^1_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}d\boldsymbol{x}dy$$

and, for all $(y,\boldsymbol{x})\in\mathbb{R}\times\mathbb{S}_{\boldsymbol{X}}$ such that $\boldsymbol{x}^{\underline{\mathbf{1}}}\neq 0$,

$$\begin{aligned}f^1_{Y|\boldsymbol{X}}(y|\boldsymbol{x})&=\frac{1}{\pi\tau(2R)^{p/2}\left|\boldsymbol{x}^{\underline{\mathbf{1}}}\right|}\int_{\mathbb{R}^p}\frac{\prod_{k=1}^p \mathbb{1}\{|\boldsymbol{u}_k|\le|\boldsymbol{x}_k|R\}}{\left(\left(y-\sum_{k=1}^p\boldsymbol{u}_k\right)/\tau\right)^2+1}d\boldsymbol{u}\\&\ge\frac{(2R)^{p/2}}{\pi\tau}\inf_{|u|\le|\boldsymbol{x}|_1R}\frac{1}{((y-u)/\tau)^2+1}.\end{aligned}$$

This yields, using $\mathbb{S}_{\boldsymbol{X}}=[-x_0,x_0]^p$ and Parseval's identity,

$$\begin{aligned}&\chi_2(\mathbb{P}_2,\mathbb{P}_1)\\&\le\frac{\pi\tau C_{\boldsymbol{X}}}{(2R)^{p/2}}\int_{[-x_0,x_0]^p}\int_{\mathbb{R}}\left(\frac{2y^2}{\tau^2}+\frac{2(|\boldsymbol{x}|_1R)^2}{\tau^2}+1\right)\left(f^1_{Y|\boldsymbol{X}}(y|\boldsymbol{x})-f^2_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\right)^2d\boldsymbol{x}dy\\&=\frac{C_{\boldsymbol{X}}x_0^p\gamma^2}{\tau(2R)^{p/2}}\int_{[-1,1]^p}\int_{\mathbb{R}}|\partial_t\mathcal{F}[F](t,x_0t\boldsymbol{x})|^2+\left((x_0pR)^2+\frac{\tau^2}{2}\right)|\mathcal{F}[F](t,x_0t\boldsymbol{x})|^2d\boldsymbol{x}dt.\end{aligned}$$

Lemmas 20 and 12 yield $\chi_2(\mathbb{P}_2,\mathbb{P}_1)\le C_{18}(U,x_0,R,\tau)\gamma^2N^2\left(eRx_0U/(4N)\right)^{2k_qN}$,

$$\begin{aligned}C_{18}(U,x_0,R,\tau):=&\frac{C_{\boldsymbol{X}}}{\tau}\left(\frac{R^{3/2}x_0Ue^3}{9\sqrt{2}}\right)^pC_{17}(Rx_0U,p,U)\left(\frac{eRx_0U}{4N(Rx_0U)}\right)^{\frac{2(p-1)N(Rx_0U)}{q}}\\&+\frac{C_{\boldsymbol{X}}}{\tau}\left(\frac{R^{3/2}x_0Ue^3}{9\sqrt{2}}\right)^pU\frac{2(x_0pR)^2+\tau^2}{2N(Rx_0U)^2}\left(\frac{eRx_0U}{4N(Rx_0U)}\right)^{\frac{2(p-1)N(Rx_0U)}{q}}.\end{aligned}$$

As a result, (iii) is satisfied if

$$n\gamma^2N^2\exp\left(-2k_qN\ln\left(\frac{4N}{eRx_0U}\right)\right)\le\frac{\xi^2}{eC_{18}(U,x_0,R,\tau)}.\tag{3.46}$$

Step 4. We take $U=4/(eRx_0)$, $N=2\lceil\underline{N}\rceil+1$ for $\underline{N}$ going to infinity with n, and τ such that $\int_0^\infty\phi(t)^2e^{-2\tau t}dt\bigvee C_{12}(\sigma,p)/(2\tau p^{2\sigma/q})\le\pi l^2/4$. Thus $N(Rx_0U)$ is universal and $\underline{N}\ge N(Rx_0U)$ and $N\le(9/2)\underline{N}$ for n large enough. (3.38), (3.42)-(3.43) (by the pigeonhole principle), and (3.46) hold for n large enough if

$$\gamma \underline{N}^{2+k_q/2} \leq \frac{1}{(\tau+1/\tau)\,(9/2)^{1+(p+k_q)/2} C_8(4/e,p,U)U}, \tag{3.47}$$

$$\gamma \underline{N}^{\sigma} \leq \frac{l}{2p^{\sigma/q}(9/2)2^{\sigma}} \sqrt{\frac{\pi(p+1)}{U}} \left(\frac{e\pi}{2}\right)^{p/2}, \tag{3.48}$$

$$\gamma \leq \frac{l}{\phi(U)} \sqrt{\frac{\pi(p+1)}{U}} \left(\frac{e\pi}{2}\right)^{p/2}, \tag{3.49}$$

$$n\gamma^2 \underline{N}^2 \exp\left(-4k_q \underline{N} \ln\left(\underline{N}\right)\right) \leq \frac{\xi^2}{(9/2)^2 e C_{18}(U,x_0,R,\tau)}, \tag{3.50}$$

and γ goes to 0 with n. Taking $\gamma = C_\gamma \underline{N}^{-(2+k_q/2)\vee\sigma} / \left((C_8(4/e,p,U)U) \wedge \sqrt{U}\right)$ for a small enough C_γ depending on l, ϕ, σ, p, and q, (3.47)-(3.48) hold because Rx_0, hence U is fixed. Then, with $\underline{N} = 3\ln(n)/\left(8k_q \ln_2(n)\right)$, (3.50) becomes, for n large enough,

$$\begin{aligned}
&\frac{C_\gamma^2}{\sqrt{n}} \exp\left(\frac{3\ln(n)\ln(8k_q\ln_2(n)/3)}{4\ln_2(n)} - \left(\left(2+\frac{k_q}{2}\right)\vee\sigma - 1\right)\ln\left(\frac{3\ln(n)}{8k_q\ln_2(n)}\right)\right)\\
&\leq \frac{\xi^2\left(C_8(4/e,p,U)U\right)^2 \wedge U}{8eC_{18}(U,x_0,R,\tau)}.
\end{aligned}$$

Moreover, we have $r(n)^2 = \underline{N}^{-2((2+k_q/2)\vee\sigma)} C_\gamma^2 \left(Rx_0/(2\pi)\right)^p \int_{U/2}^U t^p \lambda(t)^2 dt/(4\pi)$.□
All other steps 2 are the same as for (T1.1a).

Proof of (T1.2a). Denote by $E := L^2(\mathbb{R}) \times L^2(\mathbb{R})$. Equip E with $\langle \boldsymbol{g}, \boldsymbol{h}\rangle_E^2 = \langle \boldsymbol{g}_1, \boldsymbol{h}_1\rangle_{L^2(\mathbb{R})}^2 + \langle \boldsymbol{g}_2, \boldsymbol{h}_2\rangle_{L^2(\mathbb{R})}^2$. Denote by $\mathbb{P}_{j,n}^{\boldsymbol{m}}$ the law of $\left(\left(\Re\left(Z_{\boldsymbol{m}}^j(t)\right)\right)_{t\in\mathbb{R}}, \left(\Im\left(Z_{\boldsymbol{m}}^j(t)\right)\right)_{t\in\mathbb{R}}\right)$ in E and by $\mathbb{P}_{j,n}$ the law on $\ell_2(E)$ of the sequence indexed by $\boldsymbol{m} \in \mathbb{N}_0^p$. The latter can be defined as a function of $f_{j,n}$ or $(b_{\boldsymbol{m}}^j(t))_{\boldsymbol{m}\in\mathbb{N}_0^p,\ t\in\mathbb{R}}$, for $j=1,2$. Take $f_{1,n} := 0$ and $f_{2,n}$ like (3.34) with $\widetilde{\boldsymbol{N}}(1) := (N, \underline{\boldsymbol{0}}^\top)^\top \in \mathbb{N}_0^p$. (3.44) yields, for all $\boldsymbol{m} \in \mathbb{N}_0^p$, $b_{\boldsymbol{m}}^2(t) = i^{-|\boldsymbol{m}|_1}\gamma\mathbb{1}\{\boldsymbol{m} = \widetilde{\boldsymbol{N}}(q)\}\left(Rx_0\,|t|\,/2\pi\right)^{p/2}\lambda(t)$. By independence, we have, for $j=1,2$, $\mathbb{P}_{j,n} = \bigotimes_{\boldsymbol{m}\in\mathbb{N}_0^p} \mathbb{P}_{j,n}^{\boldsymbol{m}}$.

Step 1. Using (3.40) and (3.41), we have $f_{2,n} \in L^2\left(w \otimes \cosh\left(\cdot/R\right)^{\otimes p}\right)$ and $f_{2,n} \in$

$\mathcal{H}^{q,\phi,\omega}_{w,W}(l)$ if

$$\left(\frac{Rx_0U}{2\pi}\right)^p \frac{2U\gamma^2}{p+1}\left(\phi(U)\vee N^\sigma\right)^2 \le \pi^2. \tag{3.51}$$

Step 3. Let $\xi < \sqrt{2}$, $G^W_{\widetilde{\boldsymbol{N}}(q)} = \left(\Re\left(\sigma^{W,x_0\cdot}_{\widetilde{\boldsymbol{N}}(q)} b^2_{\widetilde{\boldsymbol{N}}(q)}(\cdot)\right), \Im\left(\sigma^{W,x_0\cdot}_{\widetilde{\boldsymbol{N}}(q)} b^2_{\widetilde{\boldsymbol{N}}(q)}(\cdot)\right)\right)^\top$, $\mathcal{Q}$ the covariance operator of $\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}$ on E, and, for all $\boldsymbol{h}\in E$,

$$\mathcal{L}[\boldsymbol{h}] := (\sigma/\sqrt{n})\left(\int_0^\cdot \boldsymbol{h}_1(s)ds, \int_0^\cdot \boldsymbol{h}_2(s)ds\right)^\top.$$

The reproducing kernel Hilbert space $H_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}}$ of $\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}$ on E is the image of $\mathcal{Q}^{1/2}$ with the scalar product of the image structure. By Corollary B.3 in Da Prato and Zabczyk (2014) and $\mathcal{Q} = \mathcal{L}\mathcal{L}^*$, it is the image of $\mathcal{L}$ with the norm $\|\boldsymbol{f}\|^2_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}} = (n/\sigma^2)\left(\|\boldsymbol{h}_1\|_2^2 + \|\boldsymbol{h}_2\|_2^2\right)$ for $\boldsymbol{f} = \mathcal{L}[\boldsymbol{h}]$ and derived scalar product. By (2.12) in Da Prato and Zabczyk (2014), the scalar product is also defined when one function belongs to $H_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}}$ for $\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}$ *a.e.* other function in E. By the Cameron-Martin formula (Proposition 2.26 in Da Prato and Zabczyk (2014)),

$$\frac{\mathrm{d}\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{2,n}}{\mathrm{d}\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}}(y) = \exp\left(\left\langle y, \frac{\sqrt{n}}{\sigma}\mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right\rangle_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}} - \frac{1}{2}\left|\frac{\sqrt{n}}{\sigma}\mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right|^2_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}}\right) \quad \mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}\ a.s.,$$

and, because $K(\mathbb{P}_{2,n},\mathbb{P}_{1,n}) = \int_E \ln\left(\mathrm{d}\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{2,n}/\mathrm{d}\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}(y)\right)d\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{2,n}(y)$, we have

$$K(\mathbb{P}_{2,n},\mathbb{P}_{1,n}) = \mathbb{E}\left[\left\langle Z^2_{\widetilde{\boldsymbol{N}}(q)}, \frac{\sqrt{n}}{\sigma}\mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right\rangle_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}}\right] - \frac{n}{2\sigma^2}\int_{\mathbb{R}}\left|\sigma^{W,x_0s}_{\widetilde{\boldsymbol{N}}(q)} b^2_{\widetilde{\boldsymbol{N}}(q)}(s)\right|^2 ds.$$

Because

$$\left\langle Z^2_{\widetilde{\boldsymbol{N}}(q)}, \frac{\sqrt{n}}{\sigma}\mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right\rangle_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}} = \left|\frac{\sqrt{n}}{\sigma}\mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right|^2_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}} + \left\langle \begin{pmatrix} B^{\Re}_{\widetilde{\boldsymbol{N}}(q)} \\ B^{\Im}_{\widetilde{\boldsymbol{N}}(q)} \end{pmatrix}, \mathcal{L}\left[G^W_{\widetilde{\boldsymbol{N}}(q)}\right]\right\rangle_{\mathbb{P}^{\widetilde{\boldsymbol{N}}(q)}_{1,n}},$$

and the second term in the right-hand side is a limit in quadratic mean of mean zero Gaussian random variables, hence has mean zero (see the arguments page 41 in

Da Prato and Zabczyk (2014)), we have

$$K(\mathbb{P}_{2,n}, \mathbb{P}_{1,n}) = \frac{n}{2\sigma^2} \int_{\mathbb{R}} \left| \sigma^{W,x_0 t}_{\widetilde{\boldsymbol{N}}(q)} b^2_{\widetilde{\boldsymbol{N}}(q)}(t) \right|^2 dt. \tag{3.52}$$

By Proposition 4 (ii), we have

$$\begin{aligned} K(\mathbb{P}_{2,n}, \mathbb{P}_{1,n}) &= \gamma^2 n R^p \int_{\mathbb{R}} \left(\sigma^{\cosh, Rx_0 t}_{\widetilde{\boldsymbol{N}}(q)} \right)^2 (Rx_0 \left|t\right| / (2\pi))^p \lambda(t)^2 dt/(2\sigma^2) \\ &= \gamma^2 n R^p \int_{\mathbb{R}} \rho^{\cosh, Rx_0 t}_{\widetilde{\boldsymbol{N}}(q)} \lambda(t)^2 dt/(2\sigma^2) \end{aligned}$$

and, by Theorem 7 in Gaillac and Gautier (2019*a*) (there is difference of normalisation for $\mathcal{Q}_t$ by a factor $1/(2\pi)$), for all $U/2 \leq |t| \leq U$ and $Rx_0 U < 1$, $\rho^{\cosh, Rx_0 t}_{\widetilde{\boldsymbol{N}}(q)} \leq (Rx_0 U e / (\pi(1 - (Rx_0 U)^2)))^p \exp(2k_q N \ln(Rx_0 U))$. Thus Lemma 1 (iii) holds if

$$n\gamma^2 \exp\left(-2k_q N \ln\left(\frac{1}{Rx_0 U} \right) \right) \leq \xi^2 \left(\frac{\pi(1 - (Rx_0 U)^2)}{R^2 x_0 U e} \right)^p \frac{2\sigma^2}{U}. \tag{3.53}$$

Step 4. Let $U = 1/(eRx_0)$, $N = \lceil \underline{N} \rceil$, $\gamma = C_\gamma \xi \sigma \sqrt{2eRx_0} \left(\pi\left(1 - e^{-2}\right)/R\right)^{p/2} / \underline{N}^{\overline{\sigma} \vee \sigma}$ for $\overline{\sigma} > 1/2$ and $C_\gamma = l\left(2eR/\left(1 - e^{-2}\right)\right)^{p/2} \sqrt{(p+1)e\pi R x_0} / \left(2^{\sigma+1} \xi \sigma \sqrt{2eRx_0}\right)$, and $\underline{N} = \ln(n/\ln(n))/(2k_q)$. (3.53) holds if

$$\begin{aligned} nC_\gamma^2 \exp\left(-2k_q \underline{N} - 2\left(\overline{\sigma} \vee \sigma\right) \ln\left(\underline{N}\right)\right) &= C_\gamma^2 (2k_q)^{2(\overline{\sigma} \vee \sigma)} \ln(n)^{1 - 2(\overline{\sigma} \vee \sigma)} \ln_2(n)^{2(\overline{\sigma} \vee \sigma)} \\ &\leq 1 \end{aligned}$$

so (3.51) and (3.53) hold for n large enough. □

Proof of (T1.1b). Step 1. By the proof of (T.1a), $f_{2,n} \in L^2\left(w \otimes i^{\otimes p}_{[-R,R]}\right)$ and $f_{2,n} \in \mathcal{H}^{q,\phi,\omega}_{w,W}(l)$ if

$$\left(\frac{Rx_0 U}{2\pi} \right)^p \frac{2U\gamma^2}{p+1} \left(\phi(U) \bigvee \exp\left(\kappa N \ln\left(N + 1\right)\right) \right)^2 \leq \pi l^2. \tag{3.54}$$

Step 3. Let $\xi < \sqrt{2}$ and $8/(eRx_0 U) \geq 1$. By Lemma 12, we have, for all $U/2 \leq |t| \leq U$, $\left(\sigma^{i_{[-1,1]}, Rx_0 t}_{\widetilde{\boldsymbol{N}}(q)} \right)^2 \leq (2\pi e^3/9)^p \exp\left(-2k_q N \ln\left(4(N + 3/2)/(eRx_0 U)\right)\right)$ and, by (3.52) and Proposition 4 (ii), Lemma 1 (iii) holds if

$$n\gamma^2 \exp\left(-2k_q N \ln\left(\frac{4(N+3/2)}{eRx_0U}\right)\right) \leq \frac{(p+1)\xi^2\sigma^2}{U}\left(\frac{9}{R^2Ux_0e^3}\right)^p. \tag{3.55}$$

Step 4. Let $U = 4/(eRx_0)$, $\gamma = \widetilde{C}_\gamma \exp(-\kappa N \ln(N+1))$, $\widetilde{C}_\gamma = l(\pi e/2)^{p/2}\sqrt{(p+1)e\pi Rx_0/8}$, $N = \lceil \underline{N} \rceil$, $2(\kappa + k_q)\underline{N}\ln(\underline{N}+1) = \ln\left(C_\gamma^2 n\right)$, $C_\gamma = l(2\pi Re^3/9)^{p/2}\sqrt{\pi/2}/(\xi\sigma)$. Under such a choice, (3.54) and (3.55) hold for n large enough. Moreover, we have $r(n) = C_r \exp(-\kappa N \ln(N+1))$, where $C_r = \widetilde{C}_\gamma\left(Rx_0/(2\pi)\right)^{p/2}\sqrt{\int_{U/2}^U |t|^p \lambda(t)^2 dt/(4\pi)}$, $\underline{N}\ln(\underline{N}) \leq N\ln(N+1) \leq (\underline{N}+1)\ln(\underline{N}+2)$

$$\begin{aligned} N\ln(N+1) &\leq \underline{N}\ln(\underline{N}+1) + \ln(\underline{N}+1) + 1 + o(1) \\ &= \ln\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) + (1+o(1))\ln_2\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right), \end{aligned}$$

indeed, using iteratively the definition of $\underline{N}$, $\ln(\underline{N}+1) = \ln(\underline{N}) + (1+o(1))/\underline{N}$ so $\ln(\underline{N}+1) = \ln(\underline{N})\left(1 + (1+o(1))/\ln\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right)\right)$ and

$$\begin{aligned} \ln(\underline{N}) &= \ln_2\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) - \ln_2(\underline{N}+1) \\ &= \ln_2\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) - \ln_2(\underline{N}) + (1+o(1))/\ln\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) \\ &= \ln_2\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) - (1+o(1))\ln_3\left(\left(C_\gamma^2 n\right)^{1/(2\kappa+2k_q)}\right) \end{aligned}$$

so $1/\ln\left(\left(C_\gamma^2 n\right)^{(1+o(1))\kappa/(2\kappa+2k_q)}\right) \leq r(n)\left(C_\gamma^2 n\right)^{\kappa/(2\kappa+2k_q)}/C_r \leq 1$.

Proof of (T.1.2b). Let $U = 2/(eRx_0)$, $\gamma = \widetilde{C}_\gamma \exp(-\kappa N)$, $\widetilde{C}_\gamma = l(\pi e)^{p/2}\sqrt{(p+1)e\pi Rx_0/2}$, $N = \lceil \underline{N} \rceil$, $\underline{N} = \ln(n)/(2\kappa + 2k_q)$, $C_\gamma = l\left(4\pi e/\left(\pi\left(1-e^{-2}\right)\right)\right)^{p/2}\sqrt{\pi/2}/(\xi\sigma)$. Under such a choice, $4\left(Rx_0U/(2\pi)\right)^p\left(2U\gamma^2/(p+1)\right)\left(\phi(U)\bigvee\exp(\kappa N)\right)^2 \leq \pi l^2$ and (3.53) hold for n large enough, hence steps 1 and 3. By Step 2, we have $r(n) = C_r\exp(-\kappa N) \geq C_r \exp(-\kappa\underline{N})/e$. □

3.6.4 Upper bounds

We use, for all $\epsilon > 0$, $N \in \mathbb{N}_0^{\mathbb{R}}$, $N_0 \in \mathbb{N}_0$, $T > 0$, $\widetilde{F}_1^{q,N,T,0}$ and $\widetilde{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon}$ which are defined like $\widehat{F}_1^{q,N,T,0}$ and $\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon}$ replacing $\widehat{c}_{\boldsymbol{m}}$ by $\widetilde{c}_{\boldsymbol{m}}$ and $\widehat{F}_1^{q,N,T,0}$ by $\widetilde{F}_1^{q,N,T,0}$, $\widetilde{c}_{\boldsymbol{m}} :=$

$\sum_{j=1}^n e^{i\star Y_j}\overline{g_{\boldsymbol{m}}^{W,x_0\star}}\left(\boldsymbol{X}_j/x_0\right)\mathbb{1}\{\boldsymbol{X}_j\in\mathcal{X}\}/(nx_0^p f_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j))$, $Z_{n_0}:=\sup_{f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}}\left\|\Delta_f f_{\boldsymbol{X}|\mathcal{X}}\right\|_{L^\infty(\mathcal{X})}^2$, $\Delta_f:=1/\widehat{f}_{\boldsymbol{X}|\mathcal{X}}^{\delta}-1/f_{\boldsymbol{X}|\mathcal{X}}$, $L:=(2\pi)^p\left\|\mathcal{F}_{\rm 1st}\left[f_{\alpha,\boldsymbol{\beta}}\right](\star,\cdot_2)\right\|_{L^2(\mathbb{R}^p)}^2$, $\widetilde{\omega}_{N(\star)}^{q,W,c}:=\sup_{|\boldsymbol{m}|_q\le N(\star)}1/\rho_{\boldsymbol{m}}^{W,c}$, $\Delta_{\boldsymbol{m}}:=\sum_{j=1}^n Z_j^{\boldsymbol{m},\star}/n$, $Z_j^{\boldsymbol{m},\star}:=e^{i\star Y_j}\Delta_f(\boldsymbol{X}_j)\overline{g_{\boldsymbol{m}}^{W,x_0\star}\left(\boldsymbol{X}_j/x_0\right)}\mathbb{1}\{\boldsymbol{X}_j\in\mathcal{X}\}/x_0^p$, $S_0^N(\star,\cdot_2):=\sum_{|\boldsymbol{m}|_q\le N(\star)}g_{\boldsymbol{m}}^{W,x_0\star}(\cdot_2)\Delta_{\boldsymbol{m}}(\star)$, $S_1^N(\star,\cdot_2):=\sum_{|\boldsymbol{m}|_q\le N(\star)}g_{\boldsymbol{m}}^{W,x_0\star}(\cdot_2)\mathbb{E}\left[\Delta_{\boldsymbol{m}}(\star)\right]$, $S_2^N(\star,\cdot_2):=\sum_{|\boldsymbol{m}|_q\le N(\star)}g_{\boldsymbol{m}}^{W,x_0\star}(\cdot_2)\left(\Delta_{\boldsymbol{m}}(\star)-\mathbb{E}\left[\Delta_{\boldsymbol{m}}(\star)\right]\right)$,

$$
\begin{aligned}
K_1 &:= \left\|\mathbb{1}\left\{\epsilon\le|\star|\right\}\left(\widehat{F}_1^{q,N,T,0}-\mathcal{F}_{\rm 1st}\left[f_{\alpha,\boldsymbol{\beta}}\right]\right)(\star,\cdot_2)\right\|_{L^2(1\otimes W^{\otimes p})}^2,\\
K_2 &:= \left\|\mathbb{1}\left\{|\star|<\epsilon\right\}\left(\mathcal{I}_{\underline{a},\epsilon}\left[\widehat{F}_1^{q,N,T,0}\right]-\mathcal{F}_{\rm 1st}\left[f_{\alpha,\boldsymbol{\beta}}\right]\right)(\star,\cdot_2)\right\|_{L^2(1\otimes W^{\otimes p})}^2\\
R_1(\star,\cdot_2) &:= \mathbb{1}\{\epsilon\le|\star|\}\left(\widetilde{F}_1^{q,N,T,0}-F_1^{q,N,T,0}\right)(\star,\cdot_2),\\
R_2(\star,\cdot_2) &:= \mathbb{1}\{\epsilon\le|\star|\}\left(\widehat{F}_1^{q,N,T,0}-\widetilde{F}_1^{q,N,T,0}\right)(\star,\cdot_2),\\
R_3(\star,\cdot_2) &:= \mathbb{1}\{\epsilon\le|\star|\}\left(F_1^{q,N,T,0}-F_1^{q,\infty,T,0}\right)(\star,\cdot_2),\\
R_4(\star,\cdot_2) &:= \mathbb{1}\{\epsilon\le|\star|\}\left(F_1^{q,\infty,T,0}-\mathcal{F}_{\rm 1st}\left[f_{\alpha,\boldsymbol{\beta}}\right]\right)(\star,\cdot_2),\\
\mathcal{R}_{n_0,\sup}^W &:= \sup_{f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l,M)\cap\mathcal{D},\ f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}}\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon},f_{\alpha,\boldsymbol{\beta}}\right),\\
\Delta_0^W(\star,N_0,n,z) &:= \frac{2}{\pi(2\pi)^p}\frac{c_{\boldsymbol{X}}|\star|^p}{n}\nu_q^W(x_0\star,N_0)\\
&\quad+\frac{2z}{\pi(2\pi)^p}\left(L(\star)+\frac{c_{\boldsymbol{X}}(N_0+1)^p|\star|^p}{n}\right)\widetilde{\omega}_{N_0}^{q,W,x_0\star}.
\end{aligned}
$$

Lemma 2. *For all $\boldsymbol{m}\in\mathbb{N}_0^p$, we have $\mathbb{E}\left[\widetilde{c}_{\boldsymbol{m}}(t)\right]=c_{\boldsymbol{m}}(t)$ and $\mathbb{E}\left[\left|\widetilde{c}_{\boldsymbol{m}}(t)-c_{\boldsymbol{m}}(t)\right|^2\right]\le c_{\boldsymbol{X}}/(nx_0^p)$.*

Proof. This comes from

$$\begin{aligned}\mathbb{E}\left[\widetilde{c}_{\boldsymbol{m}}(t)\right] &= \frac{1}{x_0^p}\mathbb{E}\left[\frac{e^{itY}}{f_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X})}\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\boldsymbol{X}}{x_0}\right)\mathbb{1}\{\boldsymbol{X}\in\mathcal{X}\}\right]\\ &= \frac{1}{x_0^p}\int_{\mathcal{X}}\mathbb{E}\left[e^{it\alpha+it\boldsymbol{\beta}^\top\boldsymbol{x}}\right]\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\boldsymbol{x}}{x_0}\right)d\boldsymbol{x},\\ \mathbb{E}\left[\left|\widetilde{c}_{\boldsymbol{m}}(t)-c_{\boldsymbol{m}}(t)\right|^2\right] &\le \frac{1}{nx_0^{2p}}\mathbb{E}\left[\left|\frac{e^{itY}}{f_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X})}\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\boldsymbol{X}}{x_0}\right)\right|^2\middle|\boldsymbol{X}\in\mathcal{X}\right]\\ &\le \frac{1}{nx_0^{2p}}\int_{\mathcal{X}}\frac{1}{f_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{x})}\left|\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\boldsymbol{x}}{x_0}\right)\right|^2d\boldsymbol{x}\\ &\le \frac{c_{\boldsymbol{X}}}{nx_0^p}\int_{[-1,1]^p}\left|\overline{g_{\boldsymbol{m}}^{W,x_0t}}(\boldsymbol{u})\right|^2d\boldsymbol{u}. \quad\square\end{aligned}$$

Lemma 3. *If $\widehat{f}_{\boldsymbol{X}|\mathcal{X}}$ satisfies (H1.2) then $Z_{n_0}=O_p\left(v(n_0,\mathcal{E})/\delta(n_0)\right)$.*

Proof. For all n_0 large enough so that $\sqrt{\delta(n_0)}c_{\boldsymbol{X}}\le 1$ and $x\in\mathcal{X}$, we have

$$\begin{aligned}\left|\left(\widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}}-f_{\boldsymbol{X}|\mathcal{X}}\right)(x)\right| &\le \left|\left(\widehat{f}_{\boldsymbol{X}|\mathcal{X}}-f_{\boldsymbol{X}|\mathcal{X}}\right)(x)\right|\mathbb{1}\left\{\widehat{f}_{\boldsymbol{X}|\mathcal{X}}(x)\ge\sqrt{\delta(n_0)}\right\}\\ &\quad+\left|\sqrt{\delta(n_0)}-f_{\boldsymbol{X}|\mathcal{X}}(x)\right|\mathbb{1}\left\{\widehat{f}_{\boldsymbol{X}|\mathcal{X}}(x)<\sqrt{\delta(n_0)}\right\}\\ &\le \left|\left(\widehat{f}_{\boldsymbol{X}|\mathcal{X}}-f_{\boldsymbol{X}|\mathcal{X}}\right)(x)\right|\ \text{(using }\sqrt{\delta(n_0)}c_{\boldsymbol{X}}\le 1)\end{aligned}$$

and $\delta(n_0)Z_{n_0}\le\sup_{f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}}\left\|\widehat{f}_{\boldsymbol{X}|\mathcal{X}}-f_{\boldsymbol{X}|\mathcal{X}}\right\|^2_{L^\infty(\mathcal{X})}$. We conclude by (H1.2). $\square$

In the remaining, $\mathcal{E}$ is a class of densities, $f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}$, $\eta,l,M>0$, and $f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}^{q,\phi,\omega}_{w,W}(l,M)\cap\mathcal{D}$. By Lemma 3, there exists $M_{\mathcal{E},\eta}$ such that, for all $n_0\in\mathbb{N}$, $\mathbb{P}\left(E\left(\mathcal{G}_{n_0},\mathcal{E},\eta\right)\right)\ge 1-\eta$, where $E\left(\mathcal{G}_{n_0},\mathcal{E},\eta\right):=\left\{Z_{n_0}\le M_{\mathcal{E},\eta}v(n_0,\mathcal{E})/\delta(n_0)\right\}$. We work on $E\left(\mathcal{G}_{n_0},\mathcal{E},\eta\right)$.

Proof of theorems 2 and 3. The proof consists in three parts.
In Part 1 we show, for $W=i_{[-R,R]}$ and $W=\cosh(\cdot/R)$,

$$\mathcal{R}^W_{n_0,\sup}\le C\left(\int_{\epsilon\le|t|\le T}\Delta_0^W(t,N,n,Z_{n_0})dt+4l^2\left(\sup_{t\in\mathbb{R}}\frac{1}{\omega^2_{N(t)+1}}+\frac{1}{\phi(T)^2}\right)\right)+CM^2\widetilde{w}(\underline{a}). \tag{3.56}$$

In Part 2 we take $W=i_{[-R,R]}$ and, particularising (3.56) to the different smoothness cases, obtain (T2.1), (T2.2a), (T2.2b), and (T2.3) in Theorem 2. In Part 3 we proceed similarly for the weight $W = \cosh(\cdot/R)$ and prove (T3.1) and (T3.2) in Theorem 3. We use $\theta := 7e\pi/(Rx_0)$, $\theta_0 := \pi/(4Rx_0)$, $\theta_1 := 7e^2\pi/(2Rx_0)$,

$$Q_q := 2^{k_q}\left((p/2)^p/(p!q) + \mathbb{1}\{q=\infty\}\right),$$

for all $k, l \geq 0$, $N \geq 1$, $f_{\alpha,\boldsymbol{\beta}} \in \mathcal{H}_{w,W}^{q,\phi,\omega}(l, M)$,

$$(N+l)^k \leq ((l+1)N)^k, \ \int_{\epsilon\leq|t|\leq T} L(t)dt \leq (2\pi)^{p+1}l^2. \tag{3.57}$$

Part 1. The Plancherel and Chasles identities yield $\left\|\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,N,T,\epsilon} - f_{\alpha,\boldsymbol{\beta}}\right\|_{L^2(1\otimes W^{\otimes p})}^2 \leq (K_1+K_2)/(2\pi)$. By the Jensen inequality, we have $K_1 \leq 4\sum_{j=1}^4 \|R_j\|_{L^2(1\otimes W^{\otimes p})}^2$ and, using (3.11) for the first display and Lemma 9 for the second,

$$\begin{aligned}
&K_1 + K_2 \\
&\leq K_1 + \int_{\mathbb{R}^p} 2(1+C_0(\underline{a}\epsilon))\left\|\mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](\star,\boldsymbol{b}) - \mathcal{P}_{\underline{a}}\left[\mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](\cdot,\boldsymbol{b})\right](\star)\right\|_{L^2(\mathbb{R})}^2 W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b} \\
&\quad + \int_{\mathbb{R}^p} 2C_0(\underline{a}\epsilon)\left\|\mathbb{1}\{|\star|\geq\epsilon\}\left(\widehat{F}_1^{q,N,T,0} - \mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right]\right)(\star,\boldsymbol{b})\right\|_{L^2(\mathbb{R})}^2 W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b} \\
&\leq K_1 + 4\pi(1+C_0(\underline{a}\epsilon))\widetilde{w}(\underline{a})\int_{\mathbb{R}^p}\|f_{\alpha,\boldsymbol{\beta}}(\cdot,\boldsymbol{b})\|_{L^2(w)}^2 W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b} + 2C_0(\underline{a}\epsilon)K_1 \\
&\leq (1+2C_0(\underline{a}\epsilon))K_1 + 4\pi(1+C_0(\underline{a}\epsilon))M^2\widetilde{w}(\underline{a}) \leq C\left(K_1 + 2\pi M^2\widetilde{w}(\underline{a})\right).
\end{aligned} \tag{3.58}$$

Using successively Proposition 2 and lemmas 2 and 10, we have

$$\begin{aligned}
\mathbb{E}\left[\|R_1\|_{L^2(1\otimes W^{\otimes p})}^2\right] &= \int_{\epsilon\leq|t|\leq T}\sum_{|\boldsymbol{m}|_q\leq N(t)}\frac{\mathbb{E}\left[|\widetilde{c}_{\boldsymbol{m}}(t) - c_{\boldsymbol{m}}(t)|^2\right]}{\left(\sigma_{\boldsymbol{m}}^{W,x_0t}\right)^2}dt \\
&\leq \frac{c_{\boldsymbol{X}}}{(2\pi)^p n}\int_{\epsilon\leq|t|\leq T}|t|^p\,\nu_q^W(x_0t, N(t))dt,
\end{aligned} \tag{3.59}$$

also

$$\begin{aligned}
\|R_2\|_{L^2(1\otimes W^{\otimes p})}^2 &\leq \int_{\epsilon\leq|t|\leq T}\left(\frac{x_0|t|}{2\pi}\right)^p \widetilde{\omega}_{N(t)}^{q,W,x_0t}\left\|S_0^N(t,\cdot_2)\right\|_{L^2([-1,1]^p)}^2 dt, \\
\mathbb{E}\left[\left\|S_0^N(t,\cdot_2)\right\|_{L^2([-1,1]^p)}^2\right] &= \left\|S_1^N(t,\cdot_2)\right\|_{L^2([-1,1]^p)}^2 + \mathbb{E}\left[\left\|S_2^N(t,\cdot_2)\right\|_{L^2([-1,1]^p)}^2\right],
\end{aligned}$$

$$
\begin{aligned}
&\left\|S_1^N(t,\cdot_2)\right\|^2_{L^2([-1,1]^p)}\\
&=\left\|\sum_{|\boldsymbol{m}|_q\le N(t)} g_{\boldsymbol{m}}^{W,x_0t}\left\langle \mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\cdot_2}\right](t)\left(\Delta_f f_{\boldsymbol{X}|\mathcal{X}}\right)(x_0\cdot_2),g_{\boldsymbol{m}}^{W,x_0t}\right\rangle_{L^2([-1,1]^p)}\right\|^2_{L^2([-1,1]^p)}\\
&\le\left\|\mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\cdot_2}\right](t)\left(\Delta_f f_{\boldsymbol{X}|\mathcal{X}}\right)(x_0\cdot_2)\right\|^2_{L^2([-1,1]^p)}\\
&\le Z_{n_0}\left\|\mathcal{F}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,x_0t\cdot_2)\right\|^2_{L^2([-1,1]^p)}\le Z_{n_0}\left(\frac{2\pi}{x_0|t|}\right)^p\left\|\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot_2)\right\|^2_{L^2(\mathbb{R}^p)}, \quad (3.60)
\end{aligned}
$$

and, by independence and $\sum_{|\boldsymbol{m}|_q\le N}1=\binom{N+p}{p}\mathbb{1}\{q=1\}+(N+1)^p\mathbb{1}\{q=\infty\}\le(N+1)^p$,

$$
\begin{aligned}
\mathbb{E}\left[\left\|S_2^N(t,\cdot_2)\right\|^2_{L^2([-1,1]^p)}\right]&=\sum_{|\boldsymbol{m}|_q\le N(t)}\frac{1}{n}\mathbb{E}\left[\left|Z_j^{\boldsymbol{m},t}-\mathbb{E}\left[Z_j^{\boldsymbol{m},t}\right]\right|^2\right]\\
&\le\sum_{|\boldsymbol{m}|_q\le N(t)}\frac{Z_{n_0}}{nx_0^{2p}}\int_{\mathcal{X}}\frac{1}{f_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{x})}\left|\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\boldsymbol{x}}{x_0}\right)\right|^2d\boldsymbol{x}\\
&\le\frac{(N(t)+1)^pc_{\boldsymbol{X}}Z_{n_0}}{nx_0^p}. \quad (3.61)
\end{aligned}
$$

Collecting (3.60) and (3.61), we obtain

$$
\mathbb{E}\left[\|R_2\|^2_{L^2(1\otimes W^{\otimes p})}\right]\le\frac{Z_{n_0}}{(2\pi)^p}\int_{\epsilon\le|t|\le T}\left(L(t)+\frac{c_{\boldsymbol{X}}(N(t)+1)^p\,|t|^p}{n}\right)\widetilde{\omega}_{N(t)}^{q,W,x_0t}dt. \quad (3.62)
$$

By Lemma 11 and Proposition 2, we have

$$
\begin{aligned}
\|R_3\|^2_{L^2(1\otimes W^{\otimes p})}&\le\int_{\mathbb{R}}\sum_{k>N(t)}\sum_{|\boldsymbol{m}|_q=k}\left|b_{\boldsymbol{m}}(t)\right|^2dt\le\int_{\mathbb{R}}\sum_{k>N(t)}\frac{\omega_k^2\theta_{q,k}^2(t)}{\omega_{N(t)+1}^2}dt\\
&\le\sup_{t\in\mathbb{R}}\frac{2\pi l^2}{\omega_{N(t)+1}^2} \quad (3.63)
\end{aligned}
$$

and, by Proposition 2,

$$
\begin{aligned}
\|R_4\|^2_{L^2(1\otimes W^{\otimes p})}&\le\sum_{k\in\mathbb{N}_0}\int_{|t|\ge T}\sum_{|\boldsymbol{m}|_q=k}\left|b_{\boldsymbol{m}}(t)\right|^2dt\le\sum_{k\in\mathbb{N}_0}\int_{\mathbb{R}}\frac{\phi^2(|t|)}{\phi^2(T)}\theta_{q,k}^2(t)dt\\
&\le\frac{2\pi l^2}{\phi^2(T)}. \quad (3.64)
\end{aligned}
$$

Thus we have (3.56).

Part 2. We consider now all smoothness cases when $q \in \{1, \infty\}$. Let $t \neq 0$ and $z > 0$. (3.105) and $k_q k'_q = p$ yields

$$\widetilde{\omega}_N^{q,W,x_0 t} \leq 2^p \left(1 \bigvee \frac{\theta(N+1)}{|t|}\right)^{2k_q N + p}. \tag{3.65}$$

This yields, for all $N \geq 1$, $\Delta_0^{i_{[-R,R]}}(t, N, n, z) \leq \Delta^{i_{[-R,R]}}(t, N, n, z)$, where

$$\begin{aligned}&\Delta^{i_{[-R,R]}}(\star, N, n, z)\\ &:= \left(1 \bigvee \frac{\theta(N+1)}{|\star|}\right)^{2k_q N + p} \frac{2}{\pi^{p+1}} \left(\frac{Q_q c_{\boldsymbol{X}} N^p |\star|^p}{n} + z\left(L(t) + \frac{c_{\boldsymbol{X}}(N+1)^p |\star|^p}{n}\right)\right).\end{aligned}$$

Let n_e be large enough to ensure $\underline{N} \geq 1 \geq 1/k_q$. Using $N \leq \underline{N}$, $\epsilon \leq \theta \leq \theta(\underline{N}+1)$,

$$\begin{aligned}\int_\epsilon^T \left(1 \bigvee \frac{\theta(\underline{N}+1)}{t}\right)^{2k_q \underline{N} + p} t^p dt &= (\theta(\underline{N}+1))^{2k_q \underline{N}+p} \int_\epsilon^{\theta(\underline{N}+1)} t^{-2k_q \underline{N}} dt\\ &\quad + \mathbb{1}\{\theta(\underline{N}+1) \leq T\} \int_{\theta(\underline{N}+1)}^T t^p dt\\ &\leq \frac{\epsilon^{p+1}}{2k_q \underline{N} - 1}\left(\frac{\theta(\underline{N}+1)}{\epsilon}\right)^{2k_q \underline{N}+p} + \frac{T^{p+1}}{p+1},\end{aligned} \tag{3.66}$$

$$\int_{\epsilon \leq |t| \leq T} \left(1 \bigvee \frac{\theta(\underline{N}+1)}{|t|}\right)^{2k_q \underline{N}+p} L(t) dt \leq (2\pi)^{p+1} l^2 \left(\frac{\theta(\underline{N}+1)}{\epsilon}\right)^{2k_q \underline{N}+p} \quad \text{(by (3.57))}, \tag{3.67}$$

$n_e/n \le 1$, and $n_e v(n_0,\mathcal{E})/\delta(n_0) \le 1$, we have

$$\begin{aligned}
&\int_{\epsilon\le|t|\le T} \Delta^{i_{[-R,R]}}(t,N,n,Z_{n_0})dt\\
&\le \frac{4c_{\boldsymbol{X}}\underline{N}^p}{\pi^{p+1}n_e}\left(\frac{\epsilon^{p+1}}{k_q\underline{N}}\left(\frac{\theta(\underline{N}+1)}{\epsilon}\right)^{2k_q\underline{N}+p}+\frac{T^{p+1}}{p+1}\right)\left(Q_q+\frac{M_{\mathcal{E},\eta}2^p}{n}\right)\\
&\quad+\frac{M_{\mathcal{E},\eta}2^{p+2}l^2}{n_e}\left(\frac{\theta(\underline{N}+1)}{\epsilon}\right)^{2k_q\underline{N}+p}\\
&\le \frac{\tau_0\underline{N}^{p-1}}{e^{2k_q}2^p n_e}\left(\frac{\theta(\underline{N}+1)}{\epsilon}\right)^{2k_q\underline{N}+p}+\frac{\tau_1\underline{N}^{p-1}T^{p+1}}{n_e},\\
&\tau_0 := \frac{e^{2k_q}2^{p+2}c_{\boldsymbol{X}}\theta^{p+1}}{\pi^{p+1}k_q}\left(Q_q+e^2M_{\mathcal{E},\eta}2^p\right)+M_{\mathcal{E},\eta}2^{p+2}l^2,\ \ \tau_1 := \frac{4c_{\boldsymbol{X}}}{\pi^{p+1}(p+1)}\left(Q_q+M_{\mathcal{E},\eta}2^p\right).
\end{aligned}$$

(3.56), $(\underline{N}+1)^{2k_q\underline{N}+p} \le e^{2k_q}(1+1/k_q)^p\underline{N}^{2k_q\underline{N}+p}$, $\theta/\epsilon = K_{\underline{a}}(1)$, $N+1 \ge \underline{N}$, and the definition of $\underline{a}$, yield

$$\mathcal{R}^W_{n_0,\sup} \le C\left(\tau_0\underline{N}^{p-1}\frac{\left(\underline{N}K_{\underline{a}}(1)\right)^{2k_q\underline{N}+p}}{n_e}+\tau_1\underline{N}^{p-1}\frac{T^{p+1}}{n_e}+\frac{8l^2+M^2\mathbb{1}\{w\ne i_{[-\underline{a},\underline{a}]}\}}{\omega^2_{\underline{N}}}\right). \tag{3.68}$$

The choices of $\underline{N}$ are such that the first and third terms have the same and largest order.

Proof of (T2.1). Let $n_e \ge e^e$ be large enough so that $(\ln(n_e)/\tau_2)^{\sigma((p+1)/s+2)+p-1} \le n_e$, where $\tau_2 := 4k_q(2\sigma/\mu+1)\mathcal{W}(1/(4k_q(2\sigma/\mu+1)))$. We have

$$\begin{aligned}
&2k_q(\underline{N}+k_q'/2)\ln\left(\underline{N}w^I\left(\omega^2_{\underline{N}}\right)\right)+\ln\left(\omega^2_{\underline{N}}\right)+(p-1)\ln(\underline{N})\\
&=2k_q\left(\frac{2\sigma}{\mu}+1\right)\underline{N}\ln\left(\underline{N}\right)+2\sigma\ln\left(\underline{N}\right)+\left(2\left(\frac{\sigma}{\mu}+1\right)p-1\right)\ln(\underline{N})\\
&\ge 2k_q\left(\frac{2\sigma}{\mu}+1\right)\underline{N}\ln\left(\underline{N}\right)
\end{aligned}$$

and, for all $x \ge 1/e$, $\mathcal{W}(x\ln(x)) = \ln(x)$. Using as well the definition of $\mathcal{W}$, this yields

$$\underline{N} \le \frac{\ln(n_e)}{4k_q(2\sigma/\mu+1)\mathcal{W}(\ln(n_e)/(4k_q(2\sigma/\mu+1)))} \le \frac{\ln(n_e)}{\tau_2}. \tag{3.69}$$

Using (3.69), we have $T^{p+1}\underline{N}^{p-1}/n_e = \underline{N}^{\sigma(p+1)/s+p-1}/n_e \le (\ln(n_e)/\tau_2)^{\sigma(p+1)/s+p-1}/n_e \le$

$(\ln(n_e)/\tau_2)^{-2\sigma} \leq \underline{N}^{-2\sigma} = \omega_{\underline{N}}^{-2}$. Using the definition of $\underline{N}$ and (3.68), we obtain

$$\mathcal{R}^W_{n_0,\sup} \leq \frac{C}{\underline{N}^{2\sigma}}\left(\tau_0 + \tau_1 + 8l^2 + M^2\right). \tag{3.70}$$

We also have

$$\begin{aligned}\ln(n_e) &= 2k_q(\underline{N} + k_q'/2)\ln\left(\underline{N}w^I\left(\omega_{\underline{N}}^2\right)\right) + \ln\left(\omega_{\underline{N}}^2\right) + (p-1)\ln(\underline{N}) \\ &\leq \left(2k_q\left(\frac{2\sigma}{\mu}+1\right) + \left(2\sigma - 1 + 2p\left(\frac{\sigma}{\mu}+1\right)\right)\right)\underline{N}\ln(\underline{N}),\end{aligned}$$

hence $\underline{N}\ln(\underline{N}) \geq \ln(n_e)/\tau_3$, $\tau_3 := 2k_q\left(2\sigma/\mu + 1\right) + \left(2\sigma - 1 + 2p\left(\sigma/\mu + 1\right)\right)$. Similarly to (3.69) and using for the second inequality, for all $x > 0$, $\mathcal{W}(x) \leq \ln(x+1)$ (see Theorem 2.3 in Hoorfar and Hassani (2007)), we have

$$\underline{N} \geq \frac{\ln(n_e)}{\tau_3\mathcal{W}\left(\ln(n_e)/\tau_3\right)} \geq \frac{\ln(n_e)}{\tau_3\ln\left(\ln(n_e) + \tau_3\right)} \geq \frac{\ln(n_e)}{\tau_3\ln_2(n_e)\left(1 + \ln(1 + \tau_3/e)\right)}. \tag{3.71}$$

This and (3.70) yield the result.

Proof of (T2.2a). Let $\tau_4 := \kappa(p+1)/\left(2s\left(k_q(\nu+1)+\kappa\right)\right)$. We have

$$2k_q\left(\underline{N} + \frac{k_q'}{2}\right)\ln\left(\underline{N}w^I\left(\omega_{\underline{N}}^2\right)\right) + \ln\left(\omega_{\underline{N}}^2\right) \geq 2\left(k_q(\nu+1)+\kappa\right)\underline{N}\ln\left(\underline{N}\right), \tag{3.72}$$

hence $\underline{N}\ln\left(\underline{N}\right) \leq \left(\ln(n_e) - (p-1)\ln(\underline{N})\right)/\left(2\left(k_q(\nu+1)+\kappa\right)\right)$ and

$$\frac{\underline{N}^{p-1}T^{p+1}}{n_e} = \frac{\underline{N}^{p-1}e^{\kappa(p+1)\underline{N}\ln(\underline{N}+1)/s}}{n_e} \leq e^{\kappa(p+1)/s}n_e^{\tau_4 - 1}\underline{N}^{(p-1)(1-\tau_4)}.$$

Because $s \geq \kappa(p+1)/(2k_q(1+\nu))$, we have $\tau_4 - 1 \leq -\kappa/\left(\kappa + k_q(1+\nu)\right)$ and

$$\mathcal{R}^W_{n_0,\sup} \leq C\left(\frac{\tau_1 e^{\kappa(p+1)/s}}{n_e^{\kappa/(\kappa + k_q(1+\nu))}}\underline{N}^{(p-1)(1-\tau_4)} + \frac{\tau_0 + 8l^2 + M^2\mathbb{1}\{w \neq i_{[-\underline{a},\underline{a}]}\}}{e^{2\kappa\underline{N}\ln(\underline{N}+1)}}\right). \tag{3.73}$$

Let n_e large enough so that $\underline{N} \geq 1$. Using that $\tau_4 - 1 \leq -\kappa/\left(\kappa + k_q(1+\nu)\right)$, $\tau_5 :=$

$2(k_q(1+\nu)+\kappa)\ln(2)$, that

$$\begin{aligned}\ln\left(n_e\right)-\left(p-1+k_qk_q'(1+\nu)\right)\ln\left(\underline{N}\right)&=2k_q\underline{N}\ln\left(\underline{N}w^I\left(\omega_{\underline{N}}^2\right)\right)+\ln\left(\omega_{\underline{N}}^2\right)\\&\leq\frac{\tau_5}{\ln(2)}\underline{N}\ln\left(\underline{N}+1\right),\end{aligned}$$

and that $k_qk_q'=p$, we obtain

$$\begin{aligned}e^{2\kappa\underline{N}\ln(\underline{N}+1)}&\geq n_e^{\kappa/(\kappa+k_q(1+\nu))}\underline{N}^{-\kappa(p-1+k_qk_q'(1+\nu))/(k_q(1+\nu)+\kappa)}\\&\geq n_e^{\kappa/(\kappa+k_q(1+\nu))}\underline{N}^{-((2+\nu)p-1)(1-\tau_4)}.\end{aligned}$$

We conclude because, by (3.72), $\underline{N}\leq\ln(n_e)/\tau_5$.

Proof of (T2.2b). It is derived from (3.73) with $w=i_{[-\underline{a},\underline{a}]}$ and $\nu=0$.

Proof of (T2.3). By $\ln(n_e)\geq 2k_q(\underline{N}+k_q'/2)\ln\left(\underline{N}w^I\left(\omega_{\underline{N}}^2\right)\right)+\ln\left(\omega_{\underline{N}}^2\right)\geq\ln\left(\omega_{\underline{N}}^2\right)$, we have

$$\left(\underline{N}\ln(\underline{N}+1)\right)^r\leq\frac{\ln(n_e)}{2\kappa},\tag{3.74}$$

hence, using the value of T and $\underline{N}\leq\underline{N}\ln\left(\underline{N}+1\right)/\ln(2)$,

$$\frac{T^{p+1}\underline{N}^{p-1}}{n_e}=\frac{\kappa^{p+1}(\underline{N}\ln(\underline{N}+1))^{r(p+1)}\underline{N}^{p-1}}{\gamma^{p+1}n_e}\leq\frac{\ln(n_e)^{p+1+(p-1)/r}}{\kappa^{(p-1)/r}2^{p+1+(p-1)/r}\gamma^{p+1}\ln(2)^{p-1}n_e}.$$

Moreover, because $\ln(n_e)^{p+1+(p-1)/r}$ is smaller than $\varphi(n_e)$ by definition,

$$\mathcal{R}_{n_0,\sup}^W\leq C\left(\frac{\tau_1\varphi(n_e)}{\kappa^{(p-1)/r}2^{p+1+(p-1)/r}\gamma^{p+1}\ln(2)^{p-1}n_e}+\frac{\tau_0+8l^2+M^2}{e^{2\kappa(\underline{N}\ln(1+\underline{N}))^r}}\right).\tag{3.75}$$

Let n_e large enough to ensure $\underline{N}\geq 1+1/p$. We also have,

$$2k_q(\underline{N}+k_q'/2)\ln\left(\underline{N}w^I\left(\omega_{\underline{N}}^2\right)\right)+\ln\left(\omega_{\underline{N}}^2\right)+(p-1)\ln(\underline{N})\leq 2\kappa\left(\underline{N}\ln\left(\underline{N}+1\right)\right)^r\left(1+h(\underline{N})\right),\tag{3.76}$$

where $h:=\left(k_q(1+\nu)\cdot+(2+\nu)p-1\right)\ln\left(\cdot\right)/\left(\kappa\left(\cdot\ln\left(\cdot+1\right)\right)^r\right)$. This yields, for n_e

large enough,

$$\exp\left(2\kappa\left(\underline{N}\ln(\underline{N}+1)\right)^r\right) \geq \exp\left(\frac{\ln(n_e)}{1+h(\underline{N})}\right) = n_e \exp\left(\sum_{k=1}^{\infty}(-1)^k h(\underline{N})^k \ln(n_e)\right). \tag{3.77}$$

By (3.76), we have $\underline{N}\ln(\underline{N}+1) \geq \ln(n_e)^{1/r}/d_0^{1/r}$. We obtain, by (3.74) for the second inequality,

$$h\left(\underline{N}\right) \leq \frac{k_q(1+\nu)+((2+\nu)p-1)p/(p+1)}{\kappa(\underline{N}\ln(\underline{N}+1))^{r-1}} \leq \frac{(k_q(1+\nu)+(2+\nu)p-1)d_0^{1-1/r}}{\kappa\ln(n_e)^{1-1/r}},$$

$$h\left(\underline{N}\right) \geq \frac{k_q(1+\nu)}{\kappa(1+\ln(2)/\ln(1+1/p))^r(\underline{N}\ln(\underline{N}))^{r-1}} \geq \frac{k_q(1+\nu)(2\kappa)^{1-1/r}}{\kappa(1+\ln(2)/\ln(1+1/p))^r\ln(n_e)^{1-1/r}},$$

and we conclude using that, for n_e large enough so that the remainder below is smaller in absolute value than a converging geometric series,

$$\exp\left(\sum_{k=1}^{\infty}(-1)^k h(\underline{N})^k \ln(n_e)\right) \geq \exp\left(\sum_{k=1}^{k_0}(-1)^k d_k \ln(n_e)^{(1/r-1)k+1} + O(1)\right). \quad \square$$

Part 3. Let $q \in \{1,\infty\}$. Let $t \neq 0$ and $z > 0$. By (3.100), (3.102), and Proposition 4 (ii), we have, for $q \in 1,\infty$ and $|t| \leq \theta_0$

$$\begin{aligned}\widetilde{\omega}_N^{q,W,x_0 t} \leq &\left(\frac{\pi^2}{56}\right)^p \exp\left(2k_q \ln\left(\frac{\theta_1}{|t|}\right)\left(N+\frac{k_q'}{2}\right)\right)\mathbb{1}\{|t| \leq \theta_0\} \\ &+ 2^p \exp\left(\frac{2k_q\theta_0(N+k_q')}{|t|}\right)\mathbb{1}\{|t| > \theta_0\}.\end{aligned} \tag{3.78}$$

For all $N \geq 1$, we have $\Delta_0^{\cosh(\cdot/R)}(t,N,n,z) \leq \Delta^{\cosh(\cdot/R)}(t,N,n,z)$, where

$$\begin{aligned}&\Delta^{\cosh(\cdot/R)}(\star,N,n,z) \\ &:= \frac{2}{\pi}\left(\frac{\pi}{112}\right)^p\left(\frac{2^{p/q}Q_q c_{\mathbf{X}} N^{(p-1)/q}|\star|^p}{n} + z\left(L(\star)+\frac{c_{\mathbf{X}}(N+1)^p|\star|^p}{n}\right)\right)\left(\frac{\theta_1}{|\star|}\right)^{2k_qN+p}\mathbb{1}\{|\star| \leq \theta_0\} \\ &\quad + \frac{2}{\pi^{p+1}}\left(\frac{2^{p/q}Q_q c_{\mathbf{X}} N^{(p-1)/q}|\star|^{p+k_q}}{(4\theta_0/e)^{k_q}n} + z\left(L(\star)+\frac{c_{\mathbf{X}}(N+1)^p|\star|^p}{n}\right)\right)\exp\left(\frac{2k_q\theta_0(N+k_q')}{|\star|}\right)\mathbb{1}\{|\star| > \theta_0\}.\end{aligned}$$

Let $n_e \geq e^e$ be large enough so $\underline{N} \geq (p+2)/(2k_q)$. We have, using $k_q k'_q = p$,

$$\int_{\theta_0}^{T} t^{p+k_q} e^{2k_q\theta_0(\underline{N}+k'_q)/t} dt = \int_{1/T}^{1/\theta_0} \frac{e^{2k_q\theta_0(\underline{N}+k'_q)u}}{u^{p+2+k_q}} du \leq \frac{\theta_0^{p+1+k_q} e^{2k_q(\underline{N}+k'_q)}}{2k_q(\underline{N}+k'_q)} \leq \frac{\theta_0^{p+1+k_q} e^{2k_q\underline{N}+2p}}{3p+2}. \tag{3.79}$$

Then, using $N \leq \underline{N}$, $n_e/n \leq 1$, $n_e v(n_0, \mathcal{E})/\delta(n_0) \leq 1$, and $\epsilon \leq \theta_0$, for the first display,

$$\int_{\epsilon}^{\theta_0} t^p \left(\frac{\theta_1}{t}\right)^{2k_q\underline{N}+p} dt \leq \frac{\epsilon^{p+1}}{2k_q\underline{N}-1} \left(\frac{\theta_1}{\epsilon}\right)^{2k_q\underline{N}+p}, \tag{3.80}$$

$$\begin{aligned} &\int_{\epsilon\leq|t|\leq\theta_0} \left(\frac{\theta_1}{|t|}\right)^{2k_q\underline{N}+p} L(t)dt + \int_{\theta_0\leq|t|\leq T} e^{2k_q\theta_0(\underline{N}+k'_q)/|t|} L(t)dt \\ &\leq (2\pi)^{p+1} l^2 \left(\left(\frac{\theta_1}{\epsilon}\right)^{2k_q\underline{N}+p} \mathbb{1}\{\epsilon < \theta_0\} \bigvee e^{2k_q\underline{N}+2p}\right), \end{aligned}$$

and (3.79) for the second display, we obtain

$$\begin{aligned}
&\int_{\epsilon\leq|t|\leq T}\Delta^{\cosh(\cdot/R)}(t,N,n,Z_{n_0})\\
&\leq \frac{2^{2+p/q}Q_q c_{\boldsymbol{X}}\underline{N}^{(p-1)/q}}{\pi n_e}\left(\frac{\pi}{112}\right)^p\int_{\epsilon}^{\theta_0} t^p\left(\frac{\theta_1}{t}\right)^{2k_q\underline{N}+p}dt\\
&+\frac{M_{\mathcal{E},\eta}}{n_e}\left(\left(\frac{\pi}{112}\right)^p\bigvee\frac{1}{\pi^p}\right)2^{p+2}\pi^p l^2\left(\left(\frac{\theta_1}{\epsilon}\right)^{2k_q\underline{N}+p}\mathbb{1}\{\epsilon<\theta_0\}\bigvee e^{2k_q\underline{N}+2p}\right)\\
&+\frac{4c_{\boldsymbol{X}}}{\pi n_e}\left(\frac{\pi}{112}\right)^p\frac{2^pM_{\mathcal{E},\eta}\underline{N}^p}{n}\int_{\epsilon}^{\theta_0} t^p\left(\frac{\theta_1}{t}\right)^{2k_q\underline{N}+p}dt\\
&+\frac{2^{p/q}Q_q c_{\boldsymbol{X}}\underline{N}^{(p-1)/q}e^{k_q}}{\pi^{p+1}4^{k_q-1}\theta_0^{k_q}n_e}\int_{\theta_0}^{T} t^{p+k_q}e^{2k_q\theta_0(\underline{N}+k_q')/t}dt\\
&+\frac{4c_{\boldsymbol{X}}}{\pi^{p+1}n_e}\frac{2^pM_{\mathcal{E},\eta}\underline{N}^p}{n}\int_{\theta_0}^{T} t^{p}e^{2k_q\theta_0(\underline{N}+k_q')/t}dt\\
&\leq G_1\left(\frac{\underline{N}^{k_q}}{n}\right)\frac{\underline{N}^{(p-1)/q}}{n_e}\left(\frac{\theta_1}{\epsilon}\right)^{2k_q\underline{N}+p}\mathbb{1}\{\epsilon<\theta_0\}+G_2\left(\frac{\underline{N}^{k_q}}{n}\right)\frac{\underline{N}^{(p-1)/q}}{n_e}e^{2k_q\underline{N}+2p},
\end{aligned}\tag{3.81}$$

$$G:=4\left(\frac{2k_q}{p+2}\right)^{(p-1)/q}M_{\mathcal{E},\eta}\left(\frac{\pi}{56}\right)^p\pi^p l^2,\ G_1:=\frac{4c_{\boldsymbol{X}}\theta_0^{p+1}}{\pi}\left(2^{p/q}Q_q+2^pM_{\mathcal{E},\eta}\cdot\right)\left(\frac{\pi}{112}\right)^p+G,$$
$$G_2:=\frac{4c_{\boldsymbol{X}}\theta_0^{p+1+k_q}}{(3p+2)\pi^{p+1}}\left(2^{p/q}Q_q\left(\frac{e}{4\theta_0}\right)^{k_q}+2^pM_{\mathcal{E},\eta}\cdot\right)+G.$$

Proof of (T3.1). We have, using $K_{\underline{a}}(e)=w^I\left(\omega_{\underline{N}}^2\right)$,

$$\begin{aligned}
\ln(n_e)-\frac{p-1}{q}\ln(\underline{N})&=2k_q\left(\underline{N}+\frac{k_q'}{2}\right)\ln\left(w^I\left(\omega_{\underline{N}}^2\right)\right)+\ln\left(\omega_{\underline{N}}^2\right)\\
&=\frac{4k_q\sigma}{\mu}\left(\underline{N}+\frac{k_q'}{2}\right)\ln\left(\underline{N}\right)+2\sigma\ln\left(\underline{N}\right)\geq\frac{4k_q\sigma}{\mu}\underline{N}\ln\left(\underline{N}\right),
\end{aligned}$$

hence, for n_e large enough so that $\ln(\underline{N})\geq\mu/(k_q\sigma)$,

$$\underline{N}\leq\frac{\ln(n_e)}{4k_q\sigma\ln(\underline{N})/\mu}\leq\frac{\ln(n_e)}{4k_q}.\tag{3.82}$$

Thus, using (3.29), we have $\underline{N}/n\leq 1/(4k_qe)$. This yields, using (3.81), (3.56), $\theta_1/\epsilon=$

$w^I\left(\omega_{\underline{N}}^2\right)$, $N+1 \geq \underline{N}$, and the definition of $\underline{a}$,

$$\mathcal{R}^W_{n_0,\sup} \leq C\left(G_1\left(\frac{k_q^{k_q-1}}{4e^{k_q}}\right)\underline{N}^{p-1}\frac{\left(w^I\left(\omega_{\underline{N}}^2\right)\right)^{2k_q\underline{N}+p}}{n_e} + G_2\left(\frac{k_q^{k_q-1}}{4e^{k_q}}\right)\underline{N}^{p-1}\frac{e^{2k_q\underline{N}+2p}}{n_e} + \frac{8l^2+M^2}{\omega_{\underline{N}}^2}\right).$$

By (3.82), we obtain $\omega_{\underline{N}}^2\underline{N}^{(p-1)/q}e^{2k_q\underline{N}+2p}/n_e \leq \ln(n_e)^{2\sigma+(p-1)/q}e^{2p}/(4^{2\sigma+p-1}\sqrt{n_e})$. Thus, using (3.29), we have $\underline{N}^{(p-1)/q}e^{2k_q\underline{N}+2p}/n_e \leq e^{2p}((2\sigma+(p-1)/q)/(2e))^{2\sigma+(p-1)/q}/\omega_{\underline{N}}^2$ and using the definition of $\underline{N}$,

$$\mathcal{R}^W_{n_0,\sup} \leq \frac{C}{\underline{N}^{2\sigma}}\left(G_1\left(\frac{k_q^{k_q-1}}{4e^{k_q}}\right) + G_2\left(\frac{k_q^{k_q-1}}{4e^{k_q}}\right)e^{2p}\left(\frac{2\sigma+(p-1)/q}{2e}\right)^{2\sigma+(p-1)/q} + 8l^2+M^2\right). \tag{3.83}$$

We also have

$$\begin{aligned}\ln(n_e) =& 2k_q\underline{N}\ln\left(w^I\left(\omega_{\underline{N}}^2\right)\right) + \ln\left(\omega_{\underline{N}}^2\right) + \left(\frac{p-1}{q}+\frac{2\sigma p}{\mu}\right)\ln(\underline{N})\\ \leq& \left(2\left(\frac{2k_q}{\mu}+1\right)\sigma + \left(\frac{p-1}{q}+\frac{2\sigma p}{\mu}\right)\frac{2}{p+2}\right)\underline{N}\ln(\underline{N}),\end{aligned}$$

hence $\underline{N}\ln(\underline{N}) \geq \ln(n_e)/\tau_6$, $\tau_6 := 2\left(2k_q/\mu+1\right)\sigma + 2\left((p-1)/q+(2\sigma p)/\mu\right)(p+2)$. Similarly to (3.71), we have $\underline{N} \geq \ln(n_e)/(\tau_6\left(1+\ln(1+\tau_6/e)\right)\ln_2(n_e))$, which yields the result with (3.83).

Proof of (T3.2). Because $K_{\underline{a}}(e) = e$ then $2(k_q+\kappa)\underline{N} + (p-1)\ln(\underline{N}) + p \geq 2(k_q+\kappa)\underline{N}$, we obtain $\underline{N} \leq \ln(n_e)/(2(k_q+\kappa))$. Thus using $n \geq n_e$ and (3.29), we have $G_2(\underline{N}^{k_q}/n) \leq G_2(k_q^{k_q}/(2(k_q+\kappa)e)^{k_q})$. Using (3.56), $w = i_{\mathcal{A}}$, (3.81), $\epsilon = \theta_0$, yield

$$\mathcal{R}^W_{n_0,\sup} \leq C\left(G_2\left(\frac{k_q^{k_q}}{2(k_q+\kappa)e^{k_q}}\right)\underline{N}^{(p-1)/q}\frac{e^{2k_q\underline{N}+2p}}{n_e} + \frac{8l^2}{\omega_{\underline{N}}^2}\right).$$

We conclude using the definition of $\underline{N}$, which yields

$$\underline{N}^{(p-1)/q}e^{2k_q\underline{N}}/n_e = n_e^{-\kappa/(k_q+\kappa)}\underline{N}^{\kappa(p-1)/(q(1+\kappa))}$$

and $\omega_{\underline{N}}^{-2} = e^{-2\kappa\underline{N}} = n_e^{-\kappa/(k_q+\kappa)}\underline{N}^{\kappa(p-1)/(q(k_q+\kappa))}$. □

3.6.5 Auxiliary lemmas and proof of Theorem 4

The proof of Theorem 4 uses several auxiliary lemmas. Lemmas 5 and 6 are particularly important.

Let $T_{\max} := 2^{K_{\max}}$, $\mathcal{N}_n$ be the set of functions $N \in \mathbb{N}_0^{\mathbb{R}}$ such that, for all $t \in \mathbb{R} \setminus (-\epsilon, \epsilon)$, $N(t) \in \{0, \dots, N^W_{\max,q}\}$ and $\Pi(n, Z_{n_0}, T_{\max}, N^W_{\max,q})$. Let, for all $t \neq 0$ and $N \in \mathbb{N}_0^{\mathbb{R}}$, $\Delta_{\boldsymbol{m}} := \widehat{c}_{\boldsymbol{m}} - \widetilde{c}_{\boldsymbol{m}}$, $\widetilde{\Delta}_{\boldsymbol{m}} := \widetilde{c}_{\boldsymbol{m}} - c_{\boldsymbol{m}}$,

$$\Xi(t,N) := \sum_{|\boldsymbol{m}|_q > N} \left|\frac{c_{\boldsymbol{m}}(t)}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right|^2, \; S_1(t,N) := \sum_{|\boldsymbol{m}|_q \le N} \left|\frac{\mathbb{E}\left[\Delta_{\boldsymbol{m}}(t)\right]}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right|^2,$$

$$S_2(t,N) := \sum_{|\boldsymbol{m}|_q \le N} \left|\frac{\Delta_{\boldsymbol{m}}(t) - \mathbb{E}\left[\Delta_{\boldsymbol{m}}(t)\right]}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right|^2, \; S_3(t,N) := \sum_{|\boldsymbol{m}|_q \le N} \left|\frac{\widetilde{\Delta}_{\boldsymbol{m}}(t)}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right|^2,$$

$$K_n(t) := H_W(t)\left(N^W_{\max,q} + \frac{1}{2}\right)^p, \; L := \frac{1}{42}\sqrt{\frac{2x_0^p}{c_{\boldsymbol{X}}}},$$

$$\Psi_{0,n}(t) := \exp\left(-\frac{p_n}{6}\right) + \frac{294 c_{\boldsymbol{X}} K_n^2(t)}{x_0^p n} \exp\left(-\frac{L\sqrt{p_n n}}{K_n(t)}\right),$$

$$\widetilde{B}\left(\widehat{N}\right) := \mathbb{E}\left[\sup_{T' \in \mathcal{T}_n} \int_{\epsilon \le |t| \le T'} \left(S_3\left(t, \widehat{N}(t)\right) - \frac{\Sigma\left(t, \widehat{N}(t)\right)}{2(2+\sqrt{5})}\right)_+ dt\right],$$

$$\Pi(n, Z_{n_0}, T_{\max}, N^W_{\max,q}) := Z_{n_0} \int_{\epsilon \le |t| \le T_{\max}} \Psi_n(t) dt + \Pi_1\left(n, T_{\max}, N^W_{\max,q}\right), \tag{3.84}$$

$$\Pi_1(n, T_{\max}, N^W_{\max,q}) := \frac{96\left(1 + 2\sqrt{5}\right) c_{\boldsymbol{X}} K_{\max}}{(2\pi)^p n} \int_{\epsilon}^{T_{\max}} \left(N^W_{\max,q} + 1\right) t^p \nu_q^W\left(x_0 t, N^W_{\max,q}\right) \Psi_{0,n}(t) dt,$$

$$\Psi_n := \left(2 + \frac{1}{\sqrt{5}}\right)\left(\left(\frac{2\pi}{x_0|\star|}\right)^p \widetilde{\omega}^{q,W,x_0t}_{N^W_{\max,q}} \left\|\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](\star, \cdot_2)\right\|^2_{L^2(\mathbb{R}^p)} + \left(\frac{|\star|}{2\pi}\right)^p \frac{c_{\boldsymbol{X}} \nu_q^W\left(x_0\star, N^W_{\max,q}\right)}{n}\right),$$

$$\begin{aligned}\widetilde{\Delta}_0^W(\star, N, n, z) := &\frac{c_{\boldsymbol{X}} |\star|^p}{n}\left(1 + 2(1+2p_n)(1+c_1)\right)\nu_q^W(x_0\star, N) \\ &+ z\left(L(\star) + \frac{c_{\boldsymbol{X}}(N+1)^p |\star|^p}{n}\right)\widetilde{\omega}_N^{q,W,x_0\star},\end{aligned} \tag{3.85}$$

where $H_W(t)$ is defined in Proposition 5. For all $t \in [-T, T] \setminus [-\epsilon, \epsilon]$ and $N \in \mathbb{N}_0$, using (3.30) with $c = \sqrt{5}$, we have

$$\mathcal{L}_q^W(t,N) \leq \Xi(t,N) + \left(1 + \frac{2}{\sqrt{5}}\right)(S_1(t,N) + S_2(t,N)) + (1+2\sqrt{5})S_3(t,N). \quad (3.86)$$

Lemma 4. *For all $q \in \{1,\infty\}$, $0 < \epsilon < 1 < T < T_{\max} = 2^{K_{\max}}$, $t \in [-T,T] \setminus (-\epsilon,\epsilon)$, and $N \in \{0,\ldots,N_{\max,q}^W\}$, we have*

$$\mathbb{E}\left[S_1\left(t,\widehat{N}(t)\right)\right] \leq Z_{n_0}\left(\frac{2\pi}{x_0|t|}\right)^p \widetilde{\omega}_{N_{\max,q}^W}^{q,W,x_0t} \left\|\mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot_2)\right\|_{L^2(\mathbb{R}^p)}^2, \quad (3.87)$$

$$\mathbb{E}\left[S_2\left(t,\widehat{N}(t)\right)\right] \leq Z_{n_0}\frac{c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p \nu_q^W(x_0t, N_{\max,q}^W), \quad (3.88)$$

$$\mathbb{E}\left[\left(S_3(t,N) - \frac{\Sigma(t,N)}{2(2+\sqrt{5})}\right)_+\right] \leq 48\frac{c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p \nu_q^W(x_0t,N)\Psi_{0,n}(t). \quad (3.89)$$

Proof of Lemma 4. Let the parameters in the for all statement be given. (3.87) follows from

$$\begin{aligned} S_1\left(t,\widehat{N}(t)\right) &\leq \widetilde{\omega}_{\widehat{N}(t)}^{q,W,x_0t}\left\|\mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\cdot}\right](t)\left(\Delta_f f_{\boldsymbol{X}|\mathcal{X}}\right)(x_0\cdot)\right\|_{L^2([-1,1]^p)}^2 \\ &\leq Z_{n_0}\left(\frac{2\pi}{x_0|t|}\right)^p \widetilde{\omega}_{N_{\max,q}^W}^{q,W,x_0t}\left\|\mathcal{F}_{1\text{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot_2)\right\|_{L^2(\mathbb{R}^p)}^2 \quad \text{(by (3.60))}. \end{aligned}$$

By Lemma 10, $\sum_{|\boldsymbol{m}|_q \leq N}\left(\sigma_{\boldsymbol{m}}^{W,x_0t}\right)^{-2} \leq |x_0t|^p\nu_q^W(x_0t,N)/(2\pi)^p$, so we obtain (3.88) by the following sequence of inequalities, which uses (3.61) for the second display,

$$\mathbb{E}\left[S_2\left(t,\widehat{N}(t)\right)\right] \leq \sum_{|\boldsymbol{m}|_q \leq N_{\max,q}^W}\frac{\mathbb{E}\left[\left|\Delta_{\boldsymbol{m}}(t) - \mathbb{E}\left[\Delta_{\boldsymbol{m}}(t)\right]\right|^2\right]}{\left(\sigma_{\boldsymbol{m}}^{W,x_0t}\right)^2} \leq \frac{c_{\boldsymbol{X}}Z_{n_0}|t|^p\nu_q^W\left(x_0t,N_{\max,q}^W\right)}{(2\pi)^p n}.$$

To prove (3.89), we use

$$S_3(t,N) = \int_{\mathbb{R}^p}\left|\widetilde{F}_1^{q,N,T,0}(t,\boldsymbol{b}) - F_1^{q,N,T,0}(t,\boldsymbol{b})\right|^2 W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b} = \sup_{u\in\mathcal{U}}\left|\nu_n^t(u)\right|^2,$$

$$\nu_n^t(u) := \left\langle\widetilde{F}_1^{q,N,T,0}(t,\cdot_2) - F_1^{q,N,T,0}(t,\cdot_2), u(\cdot)\right\rangle_{L^2(W^{\otimes p})} = \frac{1}{n}\sum_{j=1}^n\left(f_u^t(Y_j,X_j) - \mathbb{E}\left[f_u^t(Y_j,X_j)\right]\right),$$

$$f_u^t(\star,\cdot) := \mathbb{1}\{\cdot\in\mathcal{X}\}\frac{e^{it\star}}{x_0^p f_{\boldsymbol{X}|\mathcal{X}}(\cdot)}\int_{\mathbb{R}^p}\sum_{|\boldsymbol{m}|_q\leq N}\overline{g_{\boldsymbol{m}}^{W,x_0t}}\left(\frac{\cdot}{x_0}\right)\frac{1}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\varphi_{\boldsymbol{m}}^{W,x_0t}(\boldsymbol{b})\overline{u}(\boldsymbol{b})W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b},$$

and $\mathcal{U}$ is a countable dense set of measurable functions of $\left\{u: \ \|u\|_{L^2(W^{\otimes p})}=1\right\}$ and check the conditions of the Talagrand inequality given in Lemma 21 with $\eta=p_n$ and $\Lambda(p_n)=1$. For all $u\in\mathcal{U}$, the Cauchy-Schwarz inequality yield

$$\begin{aligned}\left\|f_u^t\right\|_{L^\infty(\mathbb{R}\times\mathcal{X})} &\le c_{\boldsymbol{X}}\left(\frac{|t|}{2\pi x_0}\right)^{p/2}\left\|\left(\sum_{|\boldsymbol{m}|_q\le N}\frac{\left|\overline{g_{\boldsymbol{m}}^{W,x_0t}}(\cdot/x_0)\right|^2}{\rho_{\boldsymbol{m}}^{W,x_0t}}\int_{\mathbb{R}^p}\left|\varphi_{\boldsymbol{m}}^{W,x_0t}(\boldsymbol{b})\right|^2W^{\otimes p}(\boldsymbol{b})d\boldsymbol{b}\right)^{1/2}\right\|_{L^\infty(\mathcal{X})}\\ &\le c_{\boldsymbol{X}}K_n(t)\left(\frac{|t|}{2\pi x_0}\right)^{p/2}\sqrt{\nu_q^W(x_0t,N)}.\end{aligned}$$

By the Cauchy-Schwarz inequality and the computation leading to (3.59), we have

$$\begin{aligned}\mathbb{E}\left[\sup_{u\in\mathcal{U}}\left|\nu_n^t(u)\right|\right]^2\le\mathbb{E}\left[\sup_{u\in\mathcal{U}}\left|\nu_n^t(u)\right|^2\right]&\le\mathbb{E}\left[\left\|\widetilde{F}_1^{q,N,T,0}(t,\cdot_2)-F_1^{q,N,T,0}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}\right]\\ &\le\frac{c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p\nu_q^W(x_0t,N)=\frac{\Sigma(t,N)}{8(2+\sqrt{5})(1+2p_n)}.\end{aligned}$$

Finally, by the Cauchy-Schwarz inequality and Proposition 5 for the second display and Lemma 10 for the third display, we have

$$\begin{aligned}\mathrm{Var}\left(\Re(f_u^t(Y_j,X_j))\right)\vee\mathrm{Var}\left(\Im(f_u^t(Y_j,X_j))\right)&\le\int_{\mathbb{R}\times\mathcal{X}}\left|f_u^t(y,\boldsymbol{x})\right|^2f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})dyd\boldsymbol{x}\\ &\le c_{\boldsymbol{X}}\left(\frac{|t|}{2\pi}\right)^p\nu_q^W(x_0t,N).\quad\square\end{aligned}$$

Lemma 5. *For all $\epsilon>0$, $q\in\{1,\infty\}$, and $T\in\mathcal{T}_n$, we have*

$$\begin{aligned}\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,\widehat{N},\widehat{T},\epsilon},f_{\alpha,\boldsymbol{\beta}}\right)\le&\frac{C\left(2+\sqrt{5}\right)^2}{2\pi}\int_{\epsilon\le|t|}\mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),T\right)\right]+\frac{\mathbb{E}\left[\mathbb{1}\{|t|\le T\}\Sigma\left(t,\widehat{N}(t)\right)\right]}{2+\sqrt{5}}dt\\ &+\frac{C2(2+\sqrt{5})^2}{\pi}\Pi(n,Z_{n_0},T_{\max},N^W_{\max,q})+CM^2\widetilde{w}(\underline{a}).\end{aligned}$$

Proof of Lemma 5. Let $\epsilon>0$, $q\in\{1,\infty\}$, and $T\in\mathcal{T}_n$.
Start from (3.23). Using, for all $T_1,T_2\ge\epsilon$, $R_{T_1}^{T_2}:=\mathbb{1}\left\{\epsilon\le|\star|\right\}\left(\widehat{F}_1^{q,\widehat{N},T_1,0}-\widehat{F}_1^{q,\widehat{N},T_2\vee T_1,0}\right)(\star,\cdot_2)$ and $R^{T_1}:=\mathbb{1}\left\{\epsilon\le|\star|\right\}\left(\widehat{F}_1^{q,\widehat{N},T_1,0}-\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\beta}\right]\right)(\star,\cdot_2)$, we have $R^{\widehat{T}}=R^T_{\widehat{T}}-R^{\widehat{T}}_T+R^T$

and $\left\|R^{\widehat{T}}\right\|^2_{L^2(W^{\otimes p})} = \mathbb{1}\{\epsilon \leq |\star|\}\, \mathcal{L}^W_q\left(\star, \widehat{N}(\star), \widehat{T}\right)$. Because

$$B_2\left(T_1, \widehat{N}\right) = \max_{T'\in\mathcal{T}_n}\left(\int_{T_1\leq|t|\leq T_1\vee T'} \left\|R^{T'}_{T_1}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} - \Sigma\left(t, \widehat{N}(t)\right) dt\right)_+, \qquad (3.90)$$

we have $\mathbb{E}\left[\|R^{T_2}_{T_1}\|^2_{L^2(1\otimes W^{\otimes p})}\right] \leq \mathbb{E}\left[B_2\left(T_1, \widehat{N}\right)\right] + \mathbb{E}\left[\Sigma_2\left(T_2, \widehat{N}\right)\right]$ for possibly random T_1 and T_2 on $\mathcal{T}_n$. By (3.30) with $c = \sqrt{5}$ and (3.25), we have

$$\begin{aligned}\mathbb{E}\left[\left\|R^{\widehat{T}}\right\|^2_{L^2(1\otimes W^{\otimes p})}\right] \leq & 2(2+\sqrt{5})\left(\mathbb{E}\left[B_2\left(T, \widehat{N}\right)\right] + \mathbb{E}\left[\Sigma_2\left(T, \widehat{N}\right)\right]\right) \\ & + \left(1 + \frac{2}{\sqrt{5}}\right)\mathbb{E}\left[\|R^T\|^2_{L^2(1\otimes W^{\otimes p})}\right].\end{aligned}$$

Using, for all $T' \in \mathcal{T}_n$, $R^{T'}_{T,1} := \widehat{F}^{q,\widehat{N},T\vee T',0}_1 - F^{q,\widehat{N},T\vee T',0}_1$, $R^{T'}_{T,2} := F^{q,\widehat{N},T,0}_1 - \widehat{F}^{q,\widehat{N},T,0}_1$, and $R^{T'}_{T,3} := F^{q,\widehat{N},T\vee T',0}_1 - F^{q,\widehat{N},T,0}_1$, by (3.30), the objective function in (3.90) is smaller than

$$\int_{T\leq|t|\leq T\vee T'}\left(k_1\sum_{j=1}^2\left\|R^{T'}_{T,j}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} + k_2\left\|R^{T'}_{T,3}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} - \Sigma\left(t, \widehat{N}(t)\right)\right)_+ dt.$$

where $k_1 = 2+\sqrt{5}$ and $k_2 = 1 + 2/\sqrt{5}$. Using that $F^{q,\infty,\infty,0}_1 = \mathcal{F}_{\mathrm{1st}}\left[f_{\alpha,\beta}\right]$, we have, for all $t \in \mathbb{R}\setminus(-\epsilon,\epsilon)$,

$$\begin{aligned}\left\|R^{T'}_{T,3}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} &= \mathbb{1}\{T \leq |t| \leq T\vee T'\}\sum_{0\leq|\boldsymbol{m}|_q\leq\widehat{N}}\left|\frac{c_{\boldsymbol{m}}(t)}{\sigma^{W,x_0t}_{\boldsymbol{m}}}\right|^2 \\ &\leq \left\|\left(F^{q,\widehat{N},T,0}_1 - \mathcal{F}_{\mathrm{1st}}\left[f_{\alpha,\beta}\right]\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})},\end{aligned}$$

hence

$$\begin{aligned}&B_2\left(T, \widehat{N}\right) \\ &\leq \max_{T'\in\mathcal{T}_n}\int_{T\leq|t|\leq T'}\left(2(2+\sqrt{5})\left\|\left(\widehat{F}^{q,\widehat{N},T',0}_1 - F^{q,\widehat{N},T',0}_1\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} - \Sigma\left(t, \widehat{N}(t)\right)\right)_+ dt \\ &\quad + \left(1 + \frac{2}{\sqrt{5}}\right)\int_{\epsilon\leq|t|}\left\|R^T(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} dt.\end{aligned}$$

Finally, we have

$$\mathbb{E}\left[\left\|R^{\widehat{T}}\right\|^2_{L^2(1\otimes W^{\otimes p})}\right] \le 2(2+\sqrt{5})\mathbb{E}\left[\Sigma_2\left(T,\widehat{N}\right)\right] + (5+2\sqrt{5})\left(1+\frac{2}{\sqrt{5}}\right)\mathbb{E}\left[\left\|R^{T}\right\|^2_{L^2(1\otimes W^{\otimes p})}\right]$$
$$+4(2+\sqrt{5})^2\mathbb{E}\left[\max_{T'\in\mathcal{T}_n}\int_{T\le|t|\le T'}\left(\left\|\left(\widehat{F}_1^{q,\widehat{N},T',0}-F_1^{q,\widehat{N},T',0}\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}-\frac{\Sigma\left(t,\widehat{N}(t)\right)}{2(2+\sqrt{5})}\right)_+\right].$$

Using (3.86) and Lemma 4, we have

$$\mathbb{E}\left[\max_{T'\in\mathcal{T}_n}\int_{T\le|t|\le T'}\left(\left\|\left(\widehat{F}_1^{q,\widehat{N},T',0}-F_1^{q,\widehat{N},T',0}\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}-\frac{\Sigma\left(t,\widehat{N}(t)\right)}{2(2+\sqrt{5})}\right)_+dt\right]$$
$$\le \mathbb{E}\left[\max_{T'\in\mathcal{T}_n}\int_{\epsilon\le|t|\le T'}\left(\sum_{|\boldsymbol{m}|_q\le\widehat{N}(t)}\left(\frac{|\widehat{c}_{\boldsymbol{m}}(t)-c_{\boldsymbol{m}}(t)|}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right)^2-\frac{\Sigma\left(t,\widehat{N}(t)\right)}{2(2+\sqrt{5})}\right)_+dt\right]$$
$$\le \left(1+2\sqrt{5}\right)\widetilde{B}\left(\widehat{N}\right)+Z_{n_0}\int_{\epsilon\le|t|\le T_{\max}}\Psi_n(t)dt.$$

Considering the first term of the last inequality and using (3.89) for the second display yields

$$\widetilde{B}\left(\widehat{N}\right)\le\mathbb{E}\left[\sum_{T'\in\mathcal{T}_n}\int_{\epsilon\le|t|\le T'}\left(S_3\left(t,\widehat{N}(t)\right)-\frac{\Sigma\left(t,\widehat{N}(t)\right)}{2(2+\sqrt{5})}\right)_+dt\right]$$
$$\le\sum_{T'\in\mathcal{T}_n}\int_\epsilon^{T'}\sum_{0\le N\le N^W_{\max,q}}96\frac{c_{\boldsymbol{X}}}{n}\left(\frac{t}{2\pi}\right)^p\nu_q^W(x_0t,N)\Psi_{0,n}(t)dt$$
$$\le\frac{96c_{\boldsymbol{X}}K_{\max}}{(2\pi)^pn}\int_\epsilon^{T_{\max}}\left(N^W_{\max,q}+1\right)t^p\nu_q^W\left(x_0t,N^W_{\max,q}\right)\Psi_{0,n}(t)dt.\qquad\square$$

Lemma 6. *For all $\epsilon>0$, $q\in\{1,\infty\}$, and $(T,N)\in\mathcal{T}_n\times\mathcal{N}_n$,*

$$\int_{\epsilon\le|t|\le T}\mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),T\right)\right]+\frac{\mathbb{1}\{|t|\le T\}}{2+\sqrt{5}}\mathbb{E}\left[\Sigma\left(T,\widehat{N}(t)\right)\right]dt$$
$$\le\left(2+\sqrt{5}\right)^2\left(\int_{\epsilon\le|t|\le T}\mathbb{E}\left[\mathcal{L}_q^W(t,N(t),T)\right]dt+\frac{1+c_1}{2+\sqrt{5}}\mathbb{E}\left[\Sigma_2(T,N)\right]\right)$$
$$+\left(2+\sqrt{5}\right)^24\Pi(n,Z_{n_0},T_{\max},N^W_{\max,q}).$$

Proof of Lemma 6. Let $t \in [-T,T] \setminus (-\epsilon,\epsilon)$, $N \in \{0,\ldots,N^W_{\max,q}\}$, $T \in \mathcal{T}_n$. Using, for all $N_1, N_2 \in \mathbb{N}_0$, $\widetilde{R}^{N_2}_{N_1}(t,\cdot_2) := \left(\widehat{F}^{q,N_1,T,0}_1 - \widehat{F}^{q,N_2\vee N_1,T,0}_1\right)(t,\cdot_2)$, we have $\mathcal{L}^W_q\left(t,\widehat{N}(t),T\right) = \widetilde{R}^N_{\widehat{N}(t)} - \widetilde{R}^{\widehat{N}(t)}_N + \mathcal{L}^W_q(t,N,T)$. Using (3.30) yields

$$\begin{aligned}\mathbb{E}\left[\mathcal{L}^W_q\left(t,\widehat{N}(t),T\right)\right] \le &(2+\sqrt{5})\left(\mathbb{E}\left[\|\widetilde{R}^N_{\widehat{N}(t)}(t,\cdot_2)\|^2_{L^2(W^{\otimes p})}\right] + \mathbb{E}\left[\|\widetilde{R}^{\widehat{N}(t)}_N(t,\cdot_2)\|^2_{L^2(W^{\otimes p})}\right]\right)\\ &+\left(1+\frac{2}{\sqrt{5}}\right)\mathbb{E}\left[\mathcal{L}^W_q(t,N,T)\right].\end{aligned}$$

Because $B_1(t,N) = \max_{N'\in\mathbb{N}_0:\ N'\le N^W_{\max,q}} \left(\sum_{N\le|\boldsymbol{m}|_q\le N'\vee N}\left(\left|\widehat{c}_{\boldsymbol{m}}(t)\right|/\sigma^{W,x_0t}_{\boldsymbol{m}}\right)^2 - \Sigma(t,N')\right)_+$, we have $\mathbb{E}\left[\left\|\widetilde{R}^{N_2}_{N_1}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}\right] \le \mathbb{E}\left[B_1(t,N_1)\right] + \mathbb{E}\left[\Sigma(t,N_2)\right]$ for possibly random N_1 and N_2. Using $c_1 \ge 1 + 1/(2+\sqrt{5})^2$ and (3.24) yield

$$\begin{aligned}&\mathbb{E}\left[\mathcal{L}^W_q\left(t,\widehat{N}(t),T\right)\right] + \frac{1}{2+\sqrt{5}}\mathbb{E}\left[\Sigma\left(t,\widehat{N}(t)\right)\right]\\ &\le (2+\sqrt{5})\left(2\mathbb{E}\left[B_1(t,N)\right] + (1+c_1)\mathbb{E}\left[\Sigma(t,N)\right]\right) + \left(1+\frac{2}{\sqrt{5}}\right)\mathbb{E}\left[\mathcal{L}^W_q(t,N,T)\right].\end{aligned}$$

By (3.30) and, for all $N' \in \mathcal{N}_n$, $\widetilde{R}^{N'}_{N,1}(t,\cdot_2) := \left(\widehat{F}^{q,N\vee N',T,0}_1 - F^{q,N\vee N',T,0}_1\right)(t,\cdot_2)$, $\widetilde{R}^{N'}_{N,2}(t,\cdot_2) := \left(F^{q,N,T,0}_1 - \widehat{F}^{q,N,T,0}_1\right)(t,\cdot_2)$, and $\widetilde{R}^{N'}_{N,3}(t,\cdot_2) := \left(F^{q,N\vee N',T,0}_1 - F^{q,N,T,0}_1\right)(t,\cdot_2)$, we have that $B_1(t,N)$ is lower or equal than

$$\max_{\substack{0\le N'\le N^W_{\max,q}\\ N'\in\mathbb{N}_0}}\left((2+\sqrt{5})\sum_{j=1}^2\left\|\widetilde{R}^{N'}_{N,j}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} + \left(1+\frac{2}{\sqrt{5}}\right)\left\|\widetilde{R}^{N'}_{N,3}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} - \Sigma(t,N')\right)_+.$$

Using $F^{q,\infty,T,0}_1(t,\cdot) = \mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\beta}\right](t,\cdot)$, we have

$$\left\|\widetilde{R}^{N'}_{N,3}(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} = \sum_{N<|\boldsymbol{m}|_q\le N\vee N'}\left|\frac{c_{\boldsymbol{m}}(t)}{\sigma^{W,x_0t}_{\boldsymbol{m}}}\right|^2 \le \left\|\left(F^{q,N,T,0}_1 - \mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\beta}\right]\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})},$$

hence

$$B_1(t,N) \leq \max_{\substack{0\leq N'\leq N^W_{\max,q}\\ N'\in\mathbb{N}_0}} \left(2(2+\sqrt{5})\left\|\left(F_1^{q,N',T,0}-\widehat{F}_1^{q,N',T,0}\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})} - \Sigma(t,N')\right)_+ + \left(1+\frac{2}{\sqrt{5}}\right)\left\|\left(F_1^{q,N,T,0}-\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right]\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}.$$

Finally, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{L}_q^W\left(t,\widehat{N}(t),T\right)\right]+\frac{1}{2+\sqrt{5}}\mathbb{E}\left[\Sigma\left(t,\widehat{N}(t)\right)\right]\\
&\leq 4(2+\sqrt{5})^2\mathbb{E}\left[\max_{0\leq N'\leq N^W_{\max,q}}\left(\left\|\left(\widehat{F}_1^{q,N',T,0}-F_1^{q,N',T,0}\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}-\frac{\Sigma(t,N')}{2(2+\sqrt{5})}\right)_+\right]\\
&\quad+(2+\sqrt{5})(1+c_1)\mathbb{E}\left[\Sigma\left(t,N\right)\right]+\left(2+\sqrt{5}\right)^2\mathbb{E}\left[\mathcal{L}_q^W\left(t,N,T\right)\right].
\end{aligned}$$

Using (3.86) for the second display and Lemma 4 for the third, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\max_{0\leq N'\leq N^W_{\max,q}}\left\|\left(\widehat{F}_1^{q,N',T,0}-F_1^{q,N',T,0}\right)(t,\cdot_2)\right\|^2_{L^2(W^{\otimes p})}-\frac{\Sigma(t,N')}{2(2+\sqrt{5})}\right]\\
&=\mathbb{E}\left[\max_{0\leq N'\leq N^W_{\max,q}}\left(\sum_{|\boldsymbol{m}|_q\leq N'}\left(\frac{|\widehat{c}_{\boldsymbol{m}}(t)-c_{\boldsymbol{m}}(t)|}{\sigma_{\boldsymbol{m}}^{W,x_0t}}\right)^2-\frac{\Sigma(t,N')}{2(2+\sqrt{5})}\right)_+\right]\\
&\leq\left(1+2\sqrt{5}\right)\mathbb{E}\left[\max_{0\leq N'\leq N^W_{\max,q}}\left(S_3(N',t)-\frac{\Sigma(t,N')}{2(2+\sqrt{5})}\right)_+\right]\\
&\quad+\left(2+\frac{1}{\sqrt{5}}\right)\mathbb{E}\left[\max_{0\leq N'\leq N^W_{\max,q}}\left(S_1(N',t)+S_2(N',t)\right)\right]\\
&\leq\left(N^W_{\max,q}+1\right)\left(1+2\sqrt{5}\right)48\frac{c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p\nu_q^W\left(x_0t,N^W_{\max,q}\right)\Psi_{0,n}(t)+Z_{n_0}\Psi_n(t).
\end{aligned}$$

Hence the result. □

Hereafter, let $(n,n_0)\in\mathbb{N}^2$ such that $v(n_0,\mathcal{E})/\delta(n_0)\leq n^{-2}\ln(n)^{-p}$, $n\geq e^{7e^2/(2\pi)}$ large enough so that $N^W_{\max,q}\geq(p+1)/k_q$. Using $\theta_1/\ln(n)\leq 2\theta_0$, let $(\theta\mathbb{1}\{W=i_{[-R,R]}\}+\theta_1\mathbb{1}\{W=\cosh(\cdot/R)\})/\ln(n)\leq\epsilon\leq\theta\mathbb{1}\{W=i_{[-R,R]}\}+2\theta_0\mathbb{1}\{W=\cosh(\cdot/R)\}$. Using the definition of $N^W_{\max,q}$ yields $N^W_{\max,q}\leq\underline{N^W_{\max,q}}$ and $\epsilon\leq\theta$ yields $\underline{N^{i_{[-R,R]}}_{\max,q}}\ln\left(\underline{N^{i_{[-R,R]}}_{\max,q}}\right)\leq\ln(n)/(2k_q)$. Then, using that, for all $x\geq 1/e$, $\mathcal{W}\left(x\ln(x)\right)=\ln(x)$, and the definition of $\mathcal{W}$ for the bound on $N_{\max,q}^{i_{[-R,R]}}$ and else the definition of $N_{\max,q}^{\cosh(\cdot/R)}$, we have, for all $t \neq 0$,

$$N_{\max,q}^{W} \leq \frac{\ln(n)}{\tau_7},\ \tau_7 := 2k_q \mathcal{W}\left(\frac{7e^2}{4\pi k_q}\right) \mathbb{1}\{W = i_{[-R,R]}\} + 2k_q \mathbb{1}\{W = \cosh(\cdot/R)\}. \tag{3.91}$$

Using, for all $|t| \geq \epsilon$ and $N \geq 1$, $2k_q(N + k_q'/2)\ln(\theta(N+1)/|t|) \leq 2k_q(N + k_q'/2)\ln(\theta N/\epsilon) + 2k_q + p$, we have

$$\left(\theta\left(N_{\max,q}^{i_{[-R,R]}} + 1\right)/|t|\right)^{2k_q N_{\max,q}^{i_{[-R,R]}} + p} \leq e^{2k_q + p}\left(\theta \underline{N_{\max,q}^{i_{[-R,R]}}}/\epsilon\right)^{2k_q \underline{N_{\max,q}^{i_{[-R,R]}}} + p}.$$

(3.65) and the definition of $N_{\max,q}^{i_{[-R,R]}}$ yield

$$\forall |t| \geq \epsilon,\ \widetilde{\omega}_{N_{\max,q}^{W}}^{q, i_{[-R,R]}, x_0 t} \leq 2^{2p} e^{2k_q} n. \tag{3.92}$$

Lemma 7. *For all $q \in \{1, \infty\}$, $\epsilon \leq T_{\max} \leq n^{\zeta_0}$, and $W \in \{i_{[-R,R]}, \cosh(\cdot/R)\}$, we have*

$$\begin{aligned}
&\int_{\epsilon}^{T_{\max}} t^p \nu_q^W\left(x_0 t, N_{\max,q}^W\right) dt \leq A_3^{W,q} \ln(n)^{a_0} n,\\
&a_0 := p\mathbb{1}\{W = i_{[-R,R]}\} + \frac{p-1}{q}\mathbb{1}\{W = \cosh(\cdot/R)\},\\
&A_3^{i_{[-R,R]},q} := \frac{2^p Q_q}{\tau_7^p (p+1)}\left(\frac{\theta 2^p e^{2k_q}}{k_q} + e^{(1-\zeta_0(p+1))7e^2/(2\pi)}\right),\\
&A_3^{\cosh(\cdot/R),q} := \frac{2^{1/q} Q_q}{k_q^{(p-1)/q}}\left(\left(\frac{\pi}{56}\right)^p \mathbb{1}\{\epsilon < \theta_0\}\theta_0 \theta_1^p + 2^p\left(\frac{e}{4}\right)^{k_q} \frac{\theta_0^{p+1} e^{2p}}{3p+2}\right).
\end{aligned}$$

Proof. Let $W = i_{[-R,R]}$. Using the definition of $\nu_q^{i_{[-R,R]}}$, $N_{\max,q}^W \geq 1$, and (3.57) for the first inequality, (3.66) and (3.91) for the second, and the arguments which yield

(3.92) for the last inequality, the result follows from

$$\begin{aligned}
&\int_{\epsilon}^{T_{\max}} t^p \nu_q^W\left(x_0 t, N_{\max,q}^W\right) dt \\
&\leq 2^p Q_q \left(\underline{N_{\max,q}^W}\right)^p \int_{\epsilon}^{T_{\max}} t^p \left(1 \bigvee \left(\frac{\theta\left(\underline{N_{\max,q}^W}+1\right)}{|t|}\right)\right)^{2k_q \underline{N_{\max,q}^W}+p} dt \\
&\leq \frac{2^p Q_q \ln(n)^p}{\tau_7^p}\left(\frac{\epsilon^{p+1}}{k_q \underline{N_{\max,q}^W}}\left(\frac{\theta(\underline{N_{\max,q}^W}+1)}{\epsilon}\right)^{2k_q \underline{N_{\max,q}^W}+p} + \frac{T_{\max}^{p+1}}{p+1}\right) \\
&\leq \frac{2^p Q_q \ln(n)^p}{\tau_7^p (p+1)}\left(\frac{\theta^{p+1}2^p e^{2k_q}}{k_q} + \frac{1}{n^{(p+1)\zeta_0 - 1}}\right) n.
\end{aligned}$$

Let now $W = \cosh(\cdot/R)$. Using the definition of ν_q^W, we have for all $N \geq 1$ and $t \neq 0$,

$$\begin{aligned}
\nu_q^{\cosh(\cdot/R)}\left(x_0 t, N\right) =& 2^{p/q} Q_q N^{(p-1)/q}\left(\frac{\pi}{56}\right)^p \left(\frac{\theta_1}{|t|}\right)^{2k_q N+p} \mathbb{1}\{|t| \leq \theta_0\} \\
&+ 2^{p(1+1/q)} Q_q N^{(p-1)/q}\left(\frac{e|t|}{4\theta_0}\right)^{k_q} \exp\left(2\theta_0 k_q (N + k_q')\right) \mathbb{1}\{|t| > \theta_0\}.
\end{aligned}$$

Because of (3.91), we have, for $t \neq 0$, $e^{2k_q N_{\max,q}^{\cosh(\cdot/R)}} \leq n$. By definition of $\underline{N_{\max,q}^{\cosh(\cdot/R)}}$, when $\epsilon < \theta_0$, we also have, for $|t| \leq \theta_0$, $(\theta_1/\epsilon)^{2k_q N_{\max,q}^{\cosh(\cdot/R)}} \leq n$. Then, using (3.91) for the first display and using (3.79) and (3.80) for the second, the result follows from

$$\begin{aligned}
&\int_{\epsilon}^{T_{\max}} t^p \nu_q^W\left(x_0 t, N_{\max,q}^W\right) dt \\
\leq& 2^{p/q} Q_q \left(\frac{\ln(n)}{2k_q}\right)^{(p-1)/q}\left(\frac{\pi}{56}\right)^p \mathbb{1}\{\epsilon < \theta_0\} \int_{\epsilon}^{\theta_0} t^p \left(\frac{\theta_1}{t}\right)^{2k_q \underline{N_{\max,q}^W}+p} dt \\
&+ Q_q \left(\frac{\ln(n)}{2k_q}\right)^{(p-1)/q} \frac{2^{p(1+1/q)} e^{k_q}}{(4\theta_0)^{k_q}} \int_{\theta_0}^{T_{\max}} t^{p+k_q} e^{2\theta_0 k_q\left(\underline{N_{\max,q}^W}+k_q'\right)} dt \\
\leq& 2^{p/q} Q_q \left(\frac{\ln(n)}{2k_q}\right)^{(p-1)/q}\left(\frac{\pi}{56}\right)^p \mathbb{1}\{\epsilon < \theta_0\}\theta_0 \theta_1^p \left(\frac{\theta_1}{\epsilon}\right)^{2k_q \underline{N_{\max,q}^W}} \\
&+ Q_q \left(\frac{\ln(n)}{2k_q}\right)^{(p-1)/q} \frac{2^{p(1+1/q)} e^{k_q}}{(4\theta_0)^{k_q}} \frac{\theta_0^{p+1+k_q} e^{2p}}{3p+2} e^{2k_q \underline{N_{\max,q}^W}} \\
\leq& A_3^{\cosh(\cdot/R),q} \ln(n)^{(p-1)/q} n. \quad \square
\end{aligned}$$

Lemma 8. *For all $q \in \{1, \infty\}$, $W \in \{i_{[-R,R]}, \cosh(\cdot/R)\}$, and $(\epsilon \vee 1) \leq T_{\max} \leq n^{\zeta_0}$,* *we have* $\Pi(n, Z_{n_0}, T_{\max}, N_{\max,q}) \leq (A_0 + A_1)/n$, *where*

$$
\begin{aligned}
A_0 &:= \frac{M_{\mathcal{E},\eta}}{n}\left(2+\frac{1}{\sqrt{5}}\right)\left(\left(\frac{4\pi^2}{\theta x_0}\right)^p b_0(2\pi)l^2 + \frac{2c_{\boldsymbol{X}}A_3^{W,q}}{(2\pi)^p e^{7e^2/(2\pi)}}\right),\\
b_0 &:= 2^{2p}e^{2k_q}\mathbb{1}\{W = i_{[-R,R]}\} + \left(\frac{\pi}{56}\right)^p \mathbb{1}\{W = \cosh(\cdot/R)\},\\
A_1 &:= \frac{96\left(1+2\sqrt{5}\right)c_{\boldsymbol{X}}\zeta_0 A_3^{W,q}}{(2\pi)^p k_q \ln(2)\left(1/\tau_7 + \pi/(7e^2)\right)^p}\left(\frac{a_0+2}{e\zeta_0}\right)^{a_0+2}\\
&\quad + \frac{96\left(1+2\sqrt{5}\right)c_{\boldsymbol{X}}\zeta_0 A_3^{W,q}}{(2\pi)^p k_q \ln(2)\left(1/\tau_7 + \pi/(7e^2)\right)^p}\frac{e^{1/b_1}294 c_{\boldsymbol{X}} a_1^2}{x_0^p}\left(\frac{a_0+2(p+1)}{e}\right)^{a_0+2(p+1)},\\
a_1 &:= \left(\frac{1}{\tau_7}+\frac{\pi}{7e^2}\right)^p \left(H_0\mathbb{1}\{W = i_{[-R,R]}\} + H_1\mathbb{1}\{W = \cosh(\cdot/R)\}\right)^p\left(1+x_0^2\right)^p,\\
b_1 &:= \frac{L\sqrt{p_2}}{a_1}\left(\frac{e(1-4(1+\mathbb{1}\{W=\cosh(\cdot/R)\})p\zeta_0)}{2p+3}\right)^{p+3/2}.
\end{aligned}
$$

Proof. Let us show

$$
Z_{n_0}\int_{\epsilon\leq|t|\leq T_{\max}}\Psi_n(t)dt \leq \frac{A_0}{n} \quad \text{and} \quad \Pi_1(n, T_{\max}, N_{\max,q}) \leq \frac{A_1}{n}. \tag{3.93}
$$

Let $W = i_{[-R,R]}$. Using for the second display $v(n_0,\mathcal{E})/\delta(n_0) \leq n^{-2}\ln(n)^{-p}$, (3.92), $\epsilon \geq \theta/\ln(n)$, (3.57), and Lemma 7, we obtain

$$
\begin{aligned}
&Z_{n_0}\int_{\epsilon\leq|t|\leq T_{\max}}\Psi_n(t)\\
&= Z_{n_0}\left(2+\frac{1}{\sqrt{5}}\right)\left(\frac{2\pi}{x_0}\right)^p\int_{\epsilon\leq|t|\leq T_{\max}}\widetilde{\omega}^{q,W,x_0t}_{N^W_{\max,q}}|t|^{-p}\left\|\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot_2)\right\|^2_{L^2(\mathbb{R}^p)}dt\\
&+ Z_{n_0}\left(2+\frac{1}{\sqrt{5}}\right)\frac{c_{\boldsymbol{X}}}{(2\pi)^p n}\int_{\epsilon\leq|t|\leq T_{\max}}|t|^p\nu_q^W(x_0t, N^W_{\max,q})dt\\
\leq& \frac{M_{\mathcal{E},\eta}}{n^2\ln(n)^p}\left(2+\frac{1}{\sqrt{5}}\right)\left(\left(\frac{8\pi}{\theta x_0}\right)^p e^{2k_q}n\ln(n)^p(2\pi)^{p+1}l^2 + \frac{2c_{\boldsymbol{X}}A_3^{W,q}\ln(n)^p}{(2\pi)^p}\right).
\end{aligned}
$$

Using $n \geq e^{7e^2/(2\pi)}$ and (3.29) yield the first inequality in (3.93). Similarly, by definition of $N^{\cosh(\cdot/R)}_{\max,q}$ and (3.78), we have $\widetilde{\omega}^{q,\cosh(\cdot/R),x_0t}_{N^W_{\max,q}} \leq (e\pi/2)^{2p}n$. This and (3.91) yield the first inequality in (3.93) for the other instances of W and q.

By (3.91), we have

$$K_n(t) \le \left(\frac{\ln(n)}{\tau_7} + \frac{1}{2}\right)^p T_{\max}^{2p} H_0^p \left(\frac{1}{T_{\max}^2} + x_0^2\right)^p \le a_1 \ln(n)^p T_{\max}^{2p}. \tag{3.94}$$

We obtain, using $T_{\max} \le n^{\zeta_0}$ for the third inequality and (3.29) for the fourth,

$$\frac{L\sqrt{p_n n}}{K_n(t)} \ge \frac{L\sqrt{p_2 \ln(n) n}}{a_1 \ln(n)^p T_{\max}^{2p}} \ge \frac{L\sqrt{p_2} n^{(1-4\zeta_0 p)/2}}{a_1 \ln(n)^{p-1/2}} \ge b_1 \ln(n)^2.$$

Using (3.91) for the first inequality, Lemma 7 for the second, and using the definition of p_n, $6(1+\zeta_0)\ln(n) > 3$, (3.94), and $T_{\max}^{4p} \le n^{4p\zeta_0}$, for the third, we have

$$\begin{aligned}
&\Pi_1(n, T_{\max}, N_{\max,q}) \\
&\le \frac{96\left(1+2\sqrt{5}\right) c_{\boldsymbol{X}} \zeta_0 \ln(n)^2}{(2\pi)^p k_q \ln(2) n} \left(\frac{1}{\tau_7} + \frac{\pi}{7e^2}\right)^p \int_\epsilon^{T_{\max}} t^p \nu_q^W\left(x_0 t, N_{\max,q}^W\right) dt \sup_{\epsilon \le t \le T_{\max}} \Psi_{0,n}(t) \\
&\le \frac{96\left(1+2\sqrt{5}\right) c_{\boldsymbol{X}} \zeta_0 A_3^{W,q}}{(2\pi)^p k_q \ln(2) n} \left(\frac{1}{\tau_7} + \frac{\pi}{7e^2}\right)^p \sup_{\epsilon \le t \le T_{\max}} \ln(n)^{p+2} n \Psi_{0,n}(t) \\
&\le \frac{96\left(1+2\sqrt{5}\right) c_{\boldsymbol{X}} \zeta_0 A_3^{W,q}}{(2\pi)^p k_q \ln(2) n} \left(\frac{1}{\tau_7} + \frac{\pi}{7e^2}\right)^p \left(\frac{\ln(n)^{p+2}}{n^{\zeta_0}} + \sup_{n>0}\left(e^{-b_1 \ln(n)^2} n^2\right) \frac{294 c_{\boldsymbol{X}} a_1^2}{x_0^p} \frac{\ln(n)^{3p+2}}{n^{2(1-2p\zeta_0)}}\right).
\end{aligned}$$

Thus, (3.29), $1 - 2p\zeta_0 > 1/2$, $\sup_{x>0}\left(e^{-b_1 \ln(x)^2} x^2\right) = e^{1/b_1}$ yield the second inequality in (3.93). We obtain similarly the bounds for the other instances of W and q. □

Proof of Theorem 4. Let n, n_0 such that $v(n_0, \mathcal{E})/\delta(n_0) \le n^{-2}\ln(n)^{-p}$, $T \in \mathcal{T}_n$, and $N \in \mathcal{N}_n$. The proof of Theorem 4 has two parts. First, we prove

$$\begin{aligned}
&\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,\widehat{N},\widehat{T},\epsilon}, f_{\alpha,\boldsymbol{\beta}}\right) \\
&\le \frac{C2\left(2+\sqrt{5}\right)^4}{\pi}\left(\int_{\epsilon \le |t| \le T} \frac{\widetilde{\Delta}_0^W(t, N(t), n, Z_{n_0})}{(2\pi)^p} dt + \sup_{t \in [-T,T]} \frac{2\pi l^2}{\omega_{N(t)+1}^2} + \frac{2\pi l^2}{\phi(T)^2}\right) \\
&\quad + CM^2 \widetilde{w}(\underline{a}) + \frac{2(2+\sqrt{5})^2 C\left(1+\left(2+\sqrt{5}\right)^2\right)(A_0 + A_1)}{\pi n},
\end{aligned} \tag{3.95}$$

where $\widetilde{\Delta}_0^W$ is defined in (3.84) and (3.85). Second, we particularise (3.95) to the different smoothness cases and prove (T4.1),(T4.2),(T5.1), and (T5.2).

Part 1. By Lemma 5 and Lemma 6, we have

$$\begin{aligned}
&\mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,\widehat{N},\widehat{T},\epsilon}, f_{\alpha,\boldsymbol{\beta}}\right)\\
&\le \frac{C\left(2+\sqrt{5}\right)^4}{2\pi}\left(\int_{\epsilon\le|t|}\mathbb{E}\left[\mathcal{L}_q^W\left(t,N(t),T\right)\right]dt+\frac{1+c_1}{2+\sqrt{5}}\mathbb{E}\left[\Sigma_2(T,N)\right]\right)\\
&\quad+\frac{2(2+\sqrt{5})^2C\left(1+\left(2+\sqrt{5}\right)^2\right)}{\pi}\Pi(n,Z_{n_0},T_{\max},N_{\max,q})+CM^2\widetilde{w}(\underline{a}).
\end{aligned}$$

The definition of Σ, (3.59), (3.62), (3.63), (3.64), and Lemma 8 yield (3.95).

Part 2. We start from (3.95) and use $A_4 := 2(2+\sqrt{5})^2C\left(1+\left(2+\sqrt{5}\right)^2\right)(A_0+A_1)/\pi$, $T^* := 2^{k^*}$, $k^* := \lfloor \ln(\phi^I(\omega_{\underline{N}^*}))/\ln(2)\rfloor$, $N^*(t) := \lfloor \underline{N}^*\rfloor$, where $\underline{N}^*$ is defined below, and

$$\mathcal{R}_{n_0,\sup}^{W,\mathrm{adp}} := \sup_{f_{\alpha,\boldsymbol{\beta}}\in\mathcal{H}_{w,W}^{q,\phi,\omega}(l,M)\cap\mathcal{D},\ f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}} \mathcal{R}_{n_0}^W\left(\widehat{f}_{\alpha,\boldsymbol{\beta}}^{q,\widehat{N},\widehat{T},\epsilon}, f_{\alpha,\boldsymbol{\beta}}\right).$$

We have, for all $|t|\ge\epsilon$, $W\in\{i_{[-R,R]},\cosh(\cdot/R)\}$, and $N\ge1$, $2\widetilde{\Delta}_0^W(t,N,n,Z_{n_0})/(\pi(2\pi)^p)\le \widetilde{\Delta}^W(t,N,n,Z_{n_0})$ where $\widetilde{\Delta}^W$ is defined like Δ^W replacing Q_q by $Q_{q,n} := Q_q\left(1+2(1+2p_n)(1+c_1)\right)$. Thus, by (3.95), we obtain, for all $W\in\{i_{[-R,R]},\cosh(\cdot/R)\}$,

$$\begin{aligned}
\mathcal{R}_{n_0,\sup}^{W,\mathrm{adp}} \le& C\left(2+\sqrt{5}\right)^4\int_{\epsilon\le|t|\le T}\widetilde{\Delta}^W\left(t,N(t),n,\frac{M_{\mathcal{E},\eta}v(n_0,\mathcal{E})}{\delta(n_0)}\right)dt\\
&+C\left(2+\sqrt{5}\right)^4\left(\sup_{t\in[-T,T]}\frac{4l^2}{\omega_{N(t)+1}^2}+\frac{4l^2}{\phi(T)^2}\right)+CM^2\widetilde{w}(\underline{a})+\frac{A_4}{n}.
\end{aligned}\tag{3.96}$$

Proof of (T4.1). Let $\underline{N}^*$ solution of

$$2k_q\underline{N}^*\ln\left(\underline{N}^*\ln(n)\right)+\ln\left(\omega_{\underline{N}^*}^2\right)+(p-1)\ln\left(\underline{N}^*\right)+\ln_2(n)=\ln(n),\tag{3.97}$$

$n\ge e^{7e^2/(2\pi)}$ large enough so $N^*\ge1$, and $(\ln(n)/\tau_2')^{\sigma/s}\le n^{\zeta_0}/2$, where $\tau_2' := 2k_q\mathcal{W}(e/(2k_q))$. By (3.97) and the definition of $\underline{N_{\max,q}^W}$, we have $N^*\le N_{\max,q}^W$ for all $t\in\mathbb{R}\setminus(-\epsilon,\epsilon)$, hence $N^*\in\mathcal{N}_n$. Also $T^*\in\mathcal{T}_n$ because, by the arguments in the proof of (T2.1),

$$\underline{N}^* \leq \frac{\ln(n)}{2k_q \mathcal{W}(\ln(n)/(2k_q))} \leq \frac{\ln(n)}{\tau_2'},$$

so we have $T^* \leq (\ln(n)/\tau_2')^{\sigma/s} \leq n^{\zeta_0}/2 \leq T_{\max}$. (3.66), (3.67), and $p_n = 6\,(1+\zeta_0)\ln(n)$, yield

$$\mathcal{R}^{W,\mathrm{adp}}_{n_0,\sup} \leq C\left(2+\sqrt{5}\right)^4 \ln(n)(N^*)^{p-1}\left(\frac{\tau_0'}{n}\left(N^*\ln(n)\right)^{2k_qN^*+p} + \frac{\tau_1'}{n}(T^*)^{p+1}\right) + \frac{8C\left(2+\sqrt{5}\right)^4 l^2}{\omega^2_{N^*}}$$
$$+ \frac{\theta C M^2}{\ln(n)} + \frac{A_4}{n},$$
$$\tau_0' := \frac{e^{2k_q}4c_{\boldsymbol{X}}\theta^{p+1}}{\pi^{p+1}k_q}\left(Q_q\left(\frac{1}{e}+\left(\frac{1}{e}+6\,(1+\zeta_0)\right)2(1+c_1)\right)+eM_{\mathcal{E},\eta}2^p\right) + \frac{M_{\mathcal{E},\eta}2^{p+2}l^2}{e},$$
$$\tau_1' := \frac{4c_{\boldsymbol{X}}}{\pi^{p+1}(p+1)}\left(Q_q\left(\frac{1}{e}+\left(\frac{1}{e}+6\,(1+\zeta_0)\right)2(1+c_1)\right)+\frac{M_{\mathcal{E},\eta}2^p}{e}\right).$$

The computation below gives lower bounds on $N^*\ln(N^*)$ and N^*:

$$\begin{aligned}\ln(n) &= 2k_q(N^*+k_q'/2)\ln\left(N^*\ln(n)\right) + (2\sigma+p-1)\ln\left(N^*\right) + \ln_2(n)\\ &\leq 2\left((2(k_q+\sigma)+2p-1)N^*\ln(N^*)\bigvee(2k_qN^*+p+1)\ln_2(n)\right)\\ &\leq 2\left((2(k_q+\sigma)+2p-1)N^*\ln(N^*)\bigvee(2k_q+p+1)N^*\ln_2(n)\right).\end{aligned}$$

This yields, using $\mathcal{W}(x) \leq \ln(x+1)$ for all $x>0$, $N^* \geq \ln(n)/((\tau_8\bigvee(2k_q+p+1))\ln_2(n)(1+\ln(1+\tau_8/e))$, where $\tau_8 := 2(2(k_q+\sigma)+2p-1)$. We conclude proceeding like for (3.70).

Proof of (T4.2). Start from (3.96), where, because $w = i_{\mathcal{A}}$, the term $M^2\widetilde{w}(\underline{a})$ is zero. Let $\underline{N}^*$ solution of $2k_q(\underline{N}^*+k_q'/2)\ln\left(\underline{N}^*\right)+\ln\left(\omega^2_{\underline{N}^*}\right)+(p-1)\ln\left(\underline{N}^*\right)+\ln_2(n) = \ln(n)$. By definition of $\underline{N}^W_{\max,q}$ this yields $N^* \leq N^W_{\max,q}$ hence $N^* \in \mathcal{N}_n$. Using arguments from the proof of (T2.2a) we have $T^* \leq n^{\kappa/(2(\kappa+k_q)s)}$ and, using $s > 2p+1/2$, for n large enough $n^{\kappa/(2(\kappa+k_q)s)} \leq n^{1/(4p+1)}/2 \leq T_{\max}$, hence $T^* \in \mathcal{T}_n$. Thus, we obtain

$$\mathcal{R}^{W,\mathrm{adp}}_{n_0,\sup} \leq C\left(2+\sqrt{5}\right)^4 \ln(n)(N^*)^{p-1}\left(\frac{\tau_0'}{n}\left(N^*\right)^{2k_qN^*+p} + \frac{\tau_1'}{n}(T^*)^{p+1}\right) + \frac{8C\left(2+\sqrt{5}\right)^4 l^2}{\omega^2_{N^*}} + \frac{A_4}{n}.$$

This yields the result following the proof of (T2.2a).

Proof of (T5.1). Starting from (3.96), the proof is similar to that of (T4.1) with elements from that of (T3.1), using $\underline{N}^*$ solution of $2k_q(\underline{N}+k_q'/2)\ln_2(n)+(p-1)\ln(\underline{N})/q+\ln\left(\omega_{\underline{N}^*}^2\right)=\ln(n)$.

Proof of (T5.2). The proof is similar to that of (T4.2). Start from (3.96). Let $\underline{N}^*$ solution of $2(k_q+\kappa)\underline{N}^*+(p-1)\ln(\underline{N}^*)/q+\ln_2(n)=\ln(n)$. Then, using the definition of $\underline{N_{\max,q}^W}$, which satisfies $2k_q\underline{N_{\max,q}^W}=\ln(n)$, we have $N^*\in\mathcal{N}_n$. Using arguments from the proof of (T3.2), we have $T^*\le n_e^{\kappa/(2s(\kappa+k_q))}$ and, using $s>4p+1/2$, for n large enough $n_e^{\kappa/(2s(\kappa+k_q))}\le n^{1/(8p+1)}/2\le T_{\max}$, hence $T^*\in\mathcal{T}_n$. This yields the result using the proof of (T3.2). □

3.7 Harmonic analysis

3.7.1 Preliminaries

P_m is the Legendre polynomial of degree m with $\|P_m\|_{L^2([-1,1])}=1$.

Lemma 9. *For all $f\in L_w^2(\mathbb{R})$, w even, nondecreasing on $[0,\infty)$, and $w(0),R>0$, we have $\left\|\mathcal{P}_R\left[\mathcal{F}\left[f\right]\right]-\mathcal{F}\left[f\right]\right\|_{L^2(\mathbb{R})}^2\le(2\pi/w(R))\|f\|_{L^2(w)}^2$.*

Proof. The result uses the Plancherel identity and

$$\left\|\mathcal{P}_R\left[\mathcal{F}\left[f\right]\right]-\mathcal{F}\left[f\right]\right\|_{L^2(\mathbb{R})}^2=2\pi\int_{\mathbb{R}}\mathbb{1}\{|a|>R\}\left|f\left(a\right)\right|^2da\le\frac{2\pi}{w(R)}\int_{\mathbb{R}}\left|f\left(a\right)\right|^2w(a)da.\quad\square$$

Proposition 4. *For all weighting function W, $c\in\mathbb{R}$, $R>0$, and $m\in\mathbb{N}_0$, we have*

(i) $g_m^{W(\cdot/R),c}=g_m^{W,Rc}$ *in* $L^2([-1,1])$,

(ii) $\sigma_m^{W(\cdot/R),c}=\sigma_m^{W,Rc}\sqrt{R}$,

(iii) $\varphi_m^{W(\cdot/R),c}=\varphi_m^{W,Rc}\left(\star/R\right)/\sqrt{R}$ *a.e.*

Proof. (i) follows from $\mathcal{Q}_c^{W(\cdot/R)}=\mathcal{Q}_{Rc}^W$ and (ii) from $\sigma_m^{W(\cdot/R),c}=\sqrt{2\pi\rho_m^{W(\cdot/R),c}/\left|c\right|}=\sqrt{2\pi\rho_m^{W,Rc}/\left|c\right|}$ (by the argument yielding (i)). Now, using (i) in the first display and

(ii) in the last display, we have, for a.e. $t \in \mathbb{R}$,

$$\begin{aligned}\sigma_m^{W,Rc}\varphi_m^{W,Rc}\left(\frac{t}{R}\right) &= \mathcal{F}_{Rc}^*\left[g_m^{W(\cdot/R),c}\right]\left(\frac{t}{R}\right) \quad (\text{where } \mathcal{F}_{Rc}^*:\ L^2([-1,1]) \to L^2(W))\\ &= \mathcal{F}_c^*\left[g_m^{W(\cdot/R),c}\right](t) \quad \left(\text{where } \mathcal{F}_c^*:\ L^2([-1,1]) \to L^2(W(\cdot/R))\right)\\ &= \sigma_m^{W(\cdot/R),c}\varphi_m^{W(\cdot/R),c}(t) = \sigma_m^{W,Rc}\sqrt{R}\varphi_m^{W(\cdot/R),c}(t),\end{aligned}$$

hence (iii) when we divide by $\sigma_m^{W,Rc}$ which is nonzero. □

Proposition 5. *For all $\boldsymbol{m} \in \mathbb{N}_0^p$, $R > 0$, $W = i_{[-R,R]}$ or $W = \cosh(\cdot/R)$, $t \neq 0$, we have $\left\|g_{\boldsymbol{m}}^{W,x_0t}\right\|_{L^\infty([-1,1]^p)} \le H_W(t)\prod_{j=1}^p \sqrt{m_j + 1/2}$, where $H_{i_{[-R,R]}}(t) = H_0^p\left(1 + (x_0|t|)^2\right)^p$, $H_0 = 2(1 + 1/\sqrt{3})$, $H_{\cosh(\cdot/R)}(t) = H_1^p(1 \vee (x_0|t|)^4)^p$, $H_1 > 0$.*

Proof. When $W = i_{[-R,R]}$, this is (66) in Bonami and Karoui (2016) else this is (50) in Gaillac and Gautier (2019*a*). □

Lemma 10. *For all $q \in \{1, \infty\}$, $t \neq 0$, $R > 0$, $N \in \mathbb{N}_0$, in the two cases of weights W in Section 3.4.3, we have $\sum_{|\boldsymbol{m}|_q \le N} 1/\rho_{\boldsymbol{m}}^{W,t} \le \nu_q^W(t, N)$.*

Proof. Let $R > 0$. We use repeatedly, for all $x > 0$ and $N \in \mathbb{N}_0$,

$$\sum_{k \le N} \exp(kx) \le \frac{\exp((N + 1/2)x)}{2\sinh(x/2)} \le \frac{\exp((N + 1/2)x)}{x} \quad (\text{because } \sinh(|x|) \ge |x|), \tag{3.98}$$

$$\le \frac{\exp(Nx)}{1 - \exp(-x)}, \tag{3.99}$$

the cardinal of $\{\boldsymbol{m} \in \mathbb{N}_0^p:\ |\boldsymbol{m}|_1 = k\}$ is $\binom{k+p-1}{p-1}$, and $(k+p-1)!/k! \le (k+p-1)^{p-1}$, and for all $m \in \mathbb{N}_0$, $\rho_m^{\cosh,Rt} = \rho_m^{\cosh(\cdot/R),t}$ and $\rho_m^{i_{[-1,1]},Rt} = \rho_m^{i_{[-R,R]},t}$.
Start by case (N.2). Let $|t| > \pi/4$ and $q = 1$. By (11) in Gaillac and Gautier (2019*a*) (there $\mathcal{Q}_t$ differs by a factor $1/(2\pi)$), we have, for all $m \in \mathbb{N}_0$,

$$\rho_m^{\cosh,t} \ge \frac{1}{2}\exp\left(-\frac{\pi(m+1)}{2|t|}\right). \tag{3.100}$$

The result is obtained from the above with (3.98) and

$$\begin{aligned}\sum_{|\boldsymbol{m}|_1\le N}\frac{1}{\rho_{\boldsymbol{m}}^{\cosh,t}}\le&2^p\sum_{k\le N}\sum_{|\boldsymbol{m}|_1=k}\exp\left(\frac{\pi(|\boldsymbol{m}|_1+p)}{2|t|}\right)\\ \le&\frac{2^{p+1}(N+p-1)^{p-1}e|t|}{\pi(p-1)!}\exp\left(\frac{\pi(N+p)}{2|t|}\right).\end{aligned}\tag{3.101}$$

Let $|t|\le\pi/4$ and $q=1$. By Theorem 4 in Gaillac and Gautier (2019a) and using that $x\mapsto\sin(x)/x$ is decreasing on $(0,\pi/2]$, we have, for all $m\in\mathbb{N}_0$,

$$\rho_m^{\cosh,t}\ge\frac{56}{\pi^2}\exp\left(-2\ln\left(\frac{7e^2\pi}{2|t|}\right)\left(m+\frac{1}{2}\right)\right).\tag{3.102}$$

The result is obtained from the above with (3.99) and

$$\begin{aligned}\sum_{|\boldsymbol{m}|_1\le N}\frac{1}{\rho_{\boldsymbol{m}}^{\cosh,t}}\le&\left(\frac{\pi^2}{56}\right)^p\sum_{k\le N}\sum_{|\boldsymbol{m}|_1=k}\exp\left(2\ln\left(\frac{7e^2\pi}{2|t|}\right)\left(|\boldsymbol{m}|_1+\frac{p}{2}\right)\right)\\ \le&\left(\frac{\pi^2}{56}\right)^p\frac{(N+p-1)^{p-1}}{(p-1)!}\exp\left(2\ln\left(\frac{7e^2\pi}{2|t|}\right)\left(N+\frac{p}{2}\right)\right)\frac{1}{1-(1/(14e^2\pi))^2}.\end{aligned}\tag{3.103}$$

The results for $q=\infty$ are obtained using (3.101) and (3.103) with $p=1$ and

$$\sum_{|\boldsymbol{m}|_\infty\le N}\frac{1}{\rho_{\boldsymbol{m}}^{\cosh,t}}\le\prod_{j=1}^p\left(\sum_{\boldsymbol{m}_j=0}^N\frac{1}{\rho_{\boldsymbol{m}_j}^{\cosh,t}}\right).\tag{3.104}$$

Consider case (N.1). Let $t\neq0$. Because $14e\ge1$ and by Lemma 13, we have, for all $m\in\mathbb{N}_0$,

$$\rho_m^{i_{[-1,1]},t}\ge\frac{1}{2}\left(\frac{|t|}{7e\pi(m+1)}\bigwedge1\right)^{2m+1}.\tag{3.105}$$

When $q=1$, the result follows from the following sequence of inequalities

$$\begin{aligned}\sum_{|\boldsymbol{m}|_1\le N}\frac{1}{\rho_{\boldsymbol{m}}^{i_{[-1,1]},t}}\le & 2^p\sum_{k\le N}\sum_{|\boldsymbol{m}|_1=k}\prod_{j=1}^p\exp\left(2\left(\boldsymbol{m}_j+\frac{1}{2}\right)\ln\left(\frac{7e\pi(\boldsymbol{m}_j+1)}{|t|}\bigvee 1\right)\right)\\ \le & \frac{2^p(N+p-1)^{p-1}(N+1)}{(p-1)!}\exp\left(2\left(N+\frac{p}{2}\right)\ln\left(\frac{7e\pi(N+1)}{|t|}\bigvee 1\right)\right).\end{aligned}$$

When $q=\infty$, we obtain the result using the above with $p=1$ and (3.104). $\square$

The proof of the next lemma is straightforward.

Lemma 11. *Let $f_{\alpha,\boldsymbol{\beta}}\in L^2\left(w\otimes W^{\otimes p}\right)$. For all $\boldsymbol{m}\in\mathbb{N}_0^p$, $t\neq 0$, we have $c_{\boldsymbol{m}}=\sigma_{\boldsymbol{m}}^{W,x_0t}b_{\boldsymbol{m}}$ a.e.*

3.7.2 Properties of the PSWF and eigenvalues

Lemma 12. *For all $c\neq 0$ and $m\in\mathbb{N}_0$, we have $|\mu_m^c|\le\sqrt{2\pi}e^{3/2}\left(e\,|c|\,/\left(4(m+3/2)\right)\right)^m/3$.*

Proof. Let $c\neq 0$ and $m\in\mathbb{N}_0$. By (69) in Rokhlin and Xiao (2007), 6.1.18 in Abramowitz and Stegun (1965), (7) in Gautschi (1959), (1.3) in Mortici and Chen (2011), we obtain

$$\begin{aligned}|\mu_m^c| &\le \frac{\sqrt{\pi}\,|c|^m\,(m!)^2}{(2m)!\Gamma(m+3/2)}\\ &\le \frac{\pi\,|c|^m}{4^m\Gamma\left(m+3/2\right)}\frac{\Gamma(m+1)}{\Gamma\left(m+1/2\right)}\\ &\le \frac{\pi\,|c|^m}{4^m\Gamma(m+3/2)}(m+1)^{1/2}\le\frac{\sqrt{\pi e^3}(e\,|c|)^m(m+1)^{1/2}}{4^m\sqrt{2}(m+3/2)^{m+1}}\end{aligned}$$

and conclude using $\sup_{x\ge 0}(x+1)^{1/2}/(x+3/2)\le 2/3$. $\square$

Lemma 13. *For all $c\neq 0$ and $m\in\mathbb{N}_0$, we have*

$$\rho_m^{i_{[-1,1]},c}\ge\frac{1}{2}\mathbb{1}\left\{m\le\frac{2\,|c|}{\pi}-1\right\}+7e\left(\frac{c}{7e\pi(m+1)}\right)^{2m+1}\mathbb{1}\left\{m>\frac{2\,|c|}{\pi}-1\right\}.$$

Proof. When $m\ge 2\,|c|\,/\pi-1$, the result follows from the fact that, by Proposition 5.1 in Bonami et al. (2018) and the Turán-Nazarov inequality (see Nazarov (2000) page 240 or (12) in Gaillac and Gautier (2019a)), $\rho_m^{i_{[-1,1]},c} \geq 7e\left(c/\left(7e\pi(m+1)\right)\right)^{2m+1}$. For all $m \leq 2|c|/\pi - 1$, the result follows from Remark 5.2 in Bonami et al. (2018) and that, for all $m \in \mathbb{N}_0$, $c \in (0,\infty) \mapsto \rho_m^c$ is nondecreasing (by the arguments in the proof of Lemma 3 in Gaillac and Gautier (2019a)). $\square$

We now use $\Pi(c) := 3c^2 \exp\left(2c^2/\sqrt{3}\right)/16$, $H(c) := \sqrt{2\Pi(c)} \vee 2$,

$$l(c) := \left(1 + 4c^2/3^{3/2}\right)\left(1 + 2c^2 3^{3/2}\right),$$

if $N \geq H(c)$ then $N \geq c$ (for all $c \geq 2$, $N \geq c\sqrt{3\exp(8/\sqrt{3})/16} > c$ else $N \geq H(c) \geq 2 > c$), $f(x) := |x|/(1-x^2)$, $g(x) := |x|/(1-x)^2$, $h(x) := |x|/(1-|x|)$, $c_f := 4/3$, $c_g := 4$, $c_h := 2$,

$$\forall x \in [-1/2, 1/2],\ f(x) \leq c_f|x|,\ g(x) \leq c_g|x|,\ h(x) \leq c_h|x|; \tag{3.106}$$

$$2\sum_{k\equiv N[2],\ 0<k<N} 2k+1 = N(N-1). \tag{3.107}$$

(3.107) is obtained because for all N even the sum is $2\sum_{p=1}^{N/2-1} 4p+1$ and else $2\sum_{p=0}^{(N-1)/2-1} 4p+3$.

Lemma 14. *For all $c \neq 0$ and $m \geq 2$, we have $\left|\mu_m^c/\mu_{m-2}^c\right| \leq \Pi(c)/m^2$.*

Proof. Let $c > 0$ and $m \in \mathbb{N}_0$ (for $c < 0$, we use $\mu_m^c = \overline{\mu_m^{-c}}$). By Theorem 8.1 in Osipov et al. (2013), we have

$$|\mu_m^c| = \frac{\sqrt{\pi}c^m(m!)^2}{(2m)!\Gamma(m+3/2)}e^{F_m(c)},\ F_m(c) = \int_0^c \left(\frac{2\left(\psi_m^t(1)\right)^2 - 1}{2t} - \frac{m}{t}\right)dt.$$

Moreover, by (65) in Bonami and Karoui (2016), for all $t > 0$,

$$\left(\sqrt{m+\frac{1}{2}} - \frac{t^2}{\sqrt{3}\sqrt{m+1/2}}\right)^2 \leq \left(\psi_m^t(1)\right)^2 \leq \left(\sqrt{m+\frac{1}{2}} + \frac{t^2}{\sqrt{3}\sqrt{m+1/2}}\right)^2$$

which yields, if $m \geq 2$,

$$\begin{aligned}
&\left(\psi_m^t(1)\right)^2 - \left(\psi_{m-2}^t(1)\right)^2 \\
&\leq \left(\sqrt{m+\frac{1}{2}} + \frac{t^2}{\sqrt{3}\sqrt{m+1/2}}\right)^2 - \left(\sqrt{(m-2)+\frac{1}{2}} - \frac{t^2}{\sqrt{3}\sqrt{(m-2)+1/2}}\right)^2 \\
&= 2 + \frac{4t^2}{\sqrt{3}} + \frac{t^4}{3}\left(\frac{1}{m+1/2} - \frac{1}{m-3/2}\right) \leq 2 + \frac{4t^2}{\sqrt{3}}. \qquad (3.108)
\end{aligned}$$

Using $\sup_{x\geq 2} x^3(x-1)/\left((x^2-1/4)(x-1/2)(x-3/2)\right) \leq 3$ and (3.108), for all $m \geq 2$,

$$\begin{aligned}
\left|\frac{\mu_m^c}{\mu_{m-2}^c}\right| &= \frac{c^2}{16}\frac{m(m-1)}{(m^2-1/4)(m-1/2)(m-3/2)} \exp\left(F_m(c) - F_{m-2}(c)\right) \\
&\leq \frac{3c^2}{16m^2}\exp\left(\int_0^c \left(\frac{\left(\psi_m^t(1)\right)^2 - \left(\psi_{m-2}^t(1)\right)^2}{t} - \frac{2}{t}\right)dt\right) \leq \frac{3c^2}{16m^2}\exp\left(\frac{2c^2}{\sqrt{3}}\right). \quad \square
\end{aligned}$$

Lemma 15. *For all $c \neq 0$ and $k \in \mathbb{N}$, we have $\left(\psi_k^c(1)\right)^2 \leq (k+1/2)\left(1+2c^2/3^{3/2}\right)^2$ and $\|\psi_k^c\|^2_{L^\infty([-1,1])} \leq (k+1/2)\left(1+4c^2/3^{3/2}\right)^2$. For all $c \neq 0$ and $k \geq c$, we have $\|\psi_k^c\|^2_{L^\infty([-1,1])} \leq k+1/2$. We also have $\|\psi_0^c\|^2_{L^\infty([-1,1])} \leq 2|c|/\pi$.*

Proof. The first assertion follows from (65) in Bonami and Karoui (2016). For the second, we use (66) in Bonami and Karoui (2016) in the first display, 22.14.7 and 22.2.10 in Abramowitz and Stegun (1965), hence $\|P_k\|_{L^\infty([-1,1])} \leq \sqrt{k+1/2}$, in the second inequality,

$$\begin{aligned}
\|\psi_k^c\|_{L^\infty([-1,1])} &\leq \|P_k\|_{L^\infty([-1,1])} + \frac{c^2}{\sqrt{3(k+1/2)}}\left(1 + \frac{\sqrt{3/2}}{\sqrt{k+1/2}}\right) \\
&\leq \sqrt{k+1/2}\left(1 + \frac{c^2}{\sqrt{3}(k+1/2)}\left(1 + \frac{\sqrt{3/2}}{\sqrt{k+1/2}}\right)\right) \\
&\leq \sqrt{k+1/2}\left(1 + \frac{4c^2}{3^{3/2}}\right).
\end{aligned}$$

The third uses (3.4) and (3.125) in Osipov et al. (2013). We obtain the last by the proof of Proposition 1 in Karoui and Moumni (2008) which yields $\|\psi_0^c\|^2_{L^\infty([-1,1])} \leq 2/(\mu_0^c)^2$ and Lemma 13. For all $c < 0$, we use $\psi_m^{-c} = \psi_m^c$. $\square$

Lemma 16. *For all $c \neq 0$ and $N \geq H(c)$, we have*

$$\left\| \frac{\partial \psi_N^c}{\partial c} \right\|_{L^\infty([-1,1])} \leq \frac{2c_f\left(C_1(c) + C_2(c)\right) C_3(c) \Pi(c)}{|c|} \sqrt{N},$$

$$C_1(c) := \frac{2H(c)+9}{(H(c)+2)^2},\ C_2(c) := \frac{2\,|c|}{\pi H(c)(H(c)-1)} + \frac{l(c)}{4},\ C_3(c) := \sqrt{1 + \frac{1}{2H(c)}}.$$

Proof. Take $c \neq 0$, $N \geq H(c)$, and $w \in [-1,1]$. Theorem 7.11 in Osipov et al. (2013) yields

$$\frac{\partial \psi_N^c}{\partial c}(w) = \frac{2\psi_N^c(1)}{|c|} \sum_{k \equiv N[2],\ k \neq N} \frac{\mu_N^c \mu_k^c}{\left(\mu_N^c\right)^2 - \left(\mu_k^c\right)^2} \psi_k^c(1) \psi_k^c(w). \tag{3.109}$$

Using $\mu_k^c / \mu_N^c \in \mathbb{R}$ if $k \equiv N[2]$ and Lemma 15, we obtain

$$\left| \frac{\partial \psi_N^c}{\partial c}(w) \right| \leq \frac{\sqrt{4N+2}}{|c|} \mathcal{C}(f, N, c),$$

$$\mathcal{C}(f, N, c) := f\left(\frac{\mu_N^c}{\mu_0^c}\right) \frac{2|c| \mathbb{1}\{N \equiv 0[2]\}}{\pi} + \sum_{\substack{0<k<N \\ k \equiv N[2]}} f\left(\frac{\mu_N^c}{\mu_k^c}\right) l(c) \left(k + \frac{1}{2}\right)$$

$$+ \sum_{\substack{k>N \\ k \equiv N[2]}} f\left(\frac{\mu_k^c}{\mu_N^c}\right) \left(k + \frac{1}{2}\right).$$

Lemma 14 yields, if $k \equiv N[2]$,

$$\left| \frac{\mu_N^c}{\mu_k^c} \right| \leq \left| \frac{\mu_N^c}{\mu_{N-2}^c} \right| \leq \frac{\Pi(c)}{N^2} \leq \frac{1}{2} \text{ if } k < N \text{ and } \left| \frac{\mu_k^c}{\mu_N^c} \right| \leq \left(\frac{\sqrt{\Pi(c)}}{N+2} \right)^{k-N} \leq \frac{1}{2} \text{ if } k > N. \tag{3.110}$$

Using (3.107), (3.106), (3.110), and $\sum_{k \in \mathbb{N}} k 2^{-k} = 2$ in the third display, the result follows from

$$\begin{aligned}\mathcal{C}(f,N,c) &\leq c_f\left(\left(\frac{2\,|c|}{\pi}+\frac{l(c)N(N-1)}{4}\right)\frac{\Pi(c)}{N^2}+\sum_{k\equiv N[2],\ k>N}\frac{k+1/2}{2^{(k-N)/2}}\left(\frac{\sqrt{2\Pi(c)}}{N+2}\right)^{k-N}\right)\\ &\leq c_f\Pi(c)\left(\frac{2\,|c|}{\pi H(c)(H(c)-1)}+\frac{l(c)}{4}+\frac{2}{(N+2)^2}\sum_{l\equiv 0[2],\ l\geq 2}\left(l+N+\frac{1}{2}\right)\frac{1}{2^{l/2}}\right)\\ &\leq c_f\Pi(c)\left(C_2(c)+\frac{2}{(N+2)^2}\left(N+\frac{9}{2}\right)\right)\leq c_f\Pi(c)\left(C_1(c)+C_2(c)\right). \qquad\square\end{aligned} \tag{3.111}$$

Lemma 17. *For all $c\neq 0$ and $N\geq H(c)$, we have*

$$\left\|\frac{\partial^2\psi_N^c}{\partial c^2}\right\|_{L^\infty([-1,1])}\leq\frac{\Pi(c)C_3(c)}{c^2}\left(C_4(c)N^{5/2}+C_5(c)N^{3/2}+C_6(c)\sqrt{N}+C_7(c)\right),$$

$$C_4(c):=c_g\left(C_2(c)-C_1(c)\right),\ C_7(c):=\frac{c_g}{(H(c)+2)^{1/2}}\left(85+\frac{246}{H(c)+2}\right),$$

$$C_5(c):=8\left(c_f\left(C_1(c)+C_2(c)\right)C_3(c)\right)^2\Pi(c)+\left(c_g+4c_f\right)C_2(c)+\left(8c_f-c_g\right)C_1(c)+2c_g,$$

$$C_6(c):=8c_hc_f(C_1(c)+C_2(c))^2\Pi(c)+\left(C_1(c)+C_2(c)\right)\left(c^2c_g+4c_f\right)+19c_g.$$

Proof. For all $c<0$, $\mu_m^c=\overline{\mu_m^{-c}}$ and $\psi_m^{-c}=\psi_m^c$, hence we only consider $c>0$. Using $c\in(0,\infty)\mapsto\psi_N^c(x)$ is analytic (see Fuchs (1964) page 320) and (7.99) in Osipov et al. (2013), we have by differentiating

$$\mu_N^c\psi_N^c(x)=\int_{-1}^1 e^{icxt}\psi_N^c(t)dt: \tag{3.112}$$

$$\mu_N^c\frac{\partial\psi_N^c}{\partial x}(x)=\int_{-1}^1 icte^{icxt}\psi_N^c(t)dt, \tag{3.113}$$

$$\mu_N^c\frac{\partial^2\psi_N^c}{\partial x^2}(x)=-\int_{-1}^1(ct)^2e^{icxt}\psi_N^c(t)dt, \tag{3.114}$$

$$\left(\frac{\partial^2\mu_N^c}{\partial c^2}\psi_N^c+2\frac{\partial\mu_N^c}{\partial c}\frac{\partial\psi_N^c}{\partial c}+\mu_N^c\frac{\partial^2\psi_N^c}{\partial c^2}\right)(x)=\int_{-1}^1 e^{icxt}\left(\frac{\partial^2\psi_N^c}{\partial c^2}(t)+2ixt\frac{\partial\psi_N^c}{\partial c}(t)-(xt)^2\psi_N^c(t)\right)dt. \tag{3.115}$$

Multiplying (3.115) by $\psi_k^c(x)$, integrating, and using (3.112)-(3.114), we obtain, for all $k \neq N$,

$$2\frac{\partial \mu_N^c}{\partial c}\int_{-1}^1 \frac{\partial \psi_N^c}{\partial c}(x)\psi_k^c(x)dx + \mu_N^c\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_k^c(x)dx$$
$$= \mu_k^c\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_k^c(x)dx + 2\frac{\mu_k^c}{c}\int_{-1}^1 x\frac{\partial \psi_N^c}{\partial c}(x)\frac{\partial \psi_k^c}{\partial x}(x)dx + \frac{\mu_k^c}{c^2}\int_{-1}^1 x^2\psi_N^c(x)\frac{\partial^2 \psi_k^c}{\partial x^2}(x)dx.$$

Recombining and using that, for all $k \neq N$, $\mu_k^c \neq \mu_N^c$ (see (3.45) in Osipov et al. (2013)), we obtain

$$(\mu_N^c - \mu_k^c)\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_k^c(x)dx$$
$$= 2\frac{\mu_k^c}{c}\int_{-1}^1 x\frac{\partial \psi_N^c}{\partial c}(x)\frac{\partial \psi_k^c}{\partial x}(x)dx + \frac{\mu_k^c}{c^2}\int_{-1}^1 x^2\psi_N^c(x)\frac{\partial^2 \psi_k^c}{\partial x^2}(x)dx - 2\frac{\partial \mu_N^c}{\partial c}\int_{-1}^1 \frac{\partial \psi_N^c}{\partial c}(x)\psi_k^c(x)dx.$$

Using (3.109), (7.69)-(7.70), and Theorem 7.11 in Osipov et al. (2013), yield, for all $k \not\equiv N[2]$, $\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_k^c(x)dx = 0$, while, for all $k \equiv N[2]$ and $k \neq N$, using (7.69)-(7.70), Theorem 7.11, (7.99) and the eigenvalues $(\chi_N^c)_{N\in\mathbb{N}_0}$ of the differential operator in (1.1) in Osipov et al. (2013),

$$\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_k^c(x)dx = \frac{2}{c}\frac{\mu_k^c}{\mu_N^c - \mu_k^c}\int_{-1}^1 x\frac{\partial \psi_N^c}{\partial c}(x)\frac{\partial \psi_k^c}{\partial x}(x)dx + \Xi_{N,k},$$

$$\Xi_{N,k} := \frac{\psi_N^c(1)\psi_k^c(1)}{c^2}\left(\frac{\mu_N^c\mu_k^c(\chi_k^c - \chi_N^c)}{(\mu_N^c - \mu_k^c)^2} - 2\frac{\mu_N^c\mu_k^c}{(\mu_N^c)^2 - (\mu_k^c)^2}\left(2 + \frac{\mu_N^c\,(2\psi_N^c(1)^2 - 1)}{\mu_N^c - \mu_k^c}\right)\right).$$

Differentiating (7.114) in Osipov et al. (2013) in c yields $\int_{-1}^1 \frac{\partial^2 \psi_N^c}{\partial c^2}(x)\psi_N^c(x)dx = -\int_{-1}^1 \left(\frac{\partial \psi_N^c}{\partial c}(x)\right)^2 dx$. Also, by (3.110), for all $k \equiv N[2]$,

$$\frac{|\mu_N^c|}{|\mu_N^c - \mu_k^c|} \leq 1 \text{ if } k < N \text{ and else } \frac{|\mu_N^c|}{|\mu_N^c - \mu_k^c|} \leq 2. \tag{3.116}$$

We obtain, using Lemma 15 and $N \geq c$ for the first term,

$$\left\| \frac{\partial^2 \psi_N^c}{\partial c^2} \right\|_{L^\infty([-1,1])} \leq \sqrt{N+\frac{1}{2}} \int_{-1}^{1} \left(\frac{\partial \psi_N^c}{\partial c}(x) \right)^2 dx + \sum_{k \equiv N[2],\ k \neq N} |\Xi_{N,k}| \, \|\psi_k^c\|_{L^\infty([-1,1])}$$
$$+ \sum_{k \equiv N[2],\ k \neq N} \frac{2\,|\mu_k^c|}{c\,|\mu_N^c - \mu_k^c|} \left| \int_{-1}^{1} x \frac{\partial \psi_N^c}{\partial c}(x) \frac{\partial \psi_k^c}{\partial x}(x) dx \right| \|\psi_k^c\|_{L^\infty([-1,1])}. \tag{3.117}$$

For the first term on the right-hand side of (3.117), using Lemma 16, we obtain

$$\sqrt{N+\frac{1}{2}} \int_{-1}^{1} \left(\frac{\partial \psi_N^c}{\partial c}(x) \right)^2 dx \leq 8 \left(c_f \left(C_1(c) + C_2(c) \right) C_3(c) \right)^2 C_3(c) \left(\frac{\Pi(c)}{c} \right)^2 N^{3/2}.$$

For the second term in (3.117), using that for all $k \equiv N[2]$, $\mu_N^c / \mu_k^c \in \mathbb{R}$ and (3.110) we obtain

$$|\Xi_{N,k}| \leq \frac{|\psi_N^c(1)|\,|\psi_k^c(1)|}{c^2} \left(g\left(\rho_k\right) \left(\chi_k^c - \chi_N^c\right) + 2 \left(2 + \frac{|2\psi_N^c(1)^2 - 1|\,|\mu_N^c|}{|\mu_N^c - \mu_k^c|} \right) f\left(\rho_k\right) \right),$$

where $\rho_k = \mu_N^c / \mu_k^c$ when $k < N$ and $\rho_k = \mu_k^c / \mu_N^c$ when $k > N$. Using $N \geq c$, (3.116), $|\chi_N^c - \chi_k^c| \leq |N-k|\,(k+N+1) + c^2$ (see (13) in Bonami and Karoui (2014a)), (3.106), and $|2\psi_N^c(1)^2 - 1| \leq 2N$ (by Lemma 15) for the first inequality, $(N-k)(k+N+1) \leq N(N+1)$ for all $0 < k < N$, (3.110), and (3.107) for the second, $(k-N)(k+N+1) = k(k+1) - N^2 - N$ for the third, the computations in (3.111), $\sum_{k=1}^{\infty} k^2 2^{-k} = 6$ and

$\sum_{k=1}^{\infty} k^3 2^{-k} = 26$, and Euclidean division for the fourth, yield

$$\begin{aligned}
&\sum_{k\equiv N[2],\ k\neq N} |\Xi_{N,k}| \, \|\psi_k^c\|_{L^\infty([-1,1])} \leq \frac{c_g\sqrt{4N+2}\mathbb{1}\{N\equiv 0[2]\}}{|c|\,\pi} \left|\frac{\mu_N^c}{\mu_0^c}\right| \left(N(N+1)+c^2+\frac{4c_f}{c_g}(N+1)\right) \\
&\quad + \frac{c_g\sqrt{4N+2}}{2c^2} \sum_{k\equiv N[2],\ 0<k<N} \left(k+\frac{1}{2}\right) l(c) \left|\frac{\mu_N^c}{\mu_k^c}\right| \left((N-k)(k+N+1)+c^2+\frac{4c_f}{c_g}(N+1)\right) \\
&\quad + \frac{c_g\sqrt{4N+2}}{2c^2} \sum_{k\equiv N[2],\ k>N} \left(k+\frac{1}{2}\right) \left|\frac{\mu_k^c}{\mu_N^c}\right| \left(|N-k|\,(k+N+1)+c^2+\frac{4c_f}{c_g}(2N+1)\right) \\
&\leq \frac{c_g\sqrt{4N+2}}{2c^2} \left(N(N+1)+c^2+\frac{4c_f}{c_g}(N+1)\right) \left(\frac{2\,|c|}{\pi}+\frac{l(c)N(N-1)}{4}\right) \frac{\Pi(c)}{N^2} \\
&\quad + \frac{c_g\sqrt{4N+2}}{2c^2} \sum_{k\equiv N[2],\ k>N} \frac{k+1/2}{2^{(k-N)/2}} \left(\frac{\sqrt{2\Pi(c)}}{N+2}\right)^{k-N} \left((k-N)(k+N+1)+c^2+\frac{4c_f}{c_g}(2N+1)\right) \\
&\leq \frac{c_g\sqrt{4N+2}\Pi(c)}{2c^2} \left(N(N+1)+c^2+\frac{4c_f}{c_g}(N+1)\right) \left(\frac{2\,|c|}{\pi H(c)(H(c)-1)}+\frac{l(c)}{4}\right) \\
&\quad + \frac{c_g\sqrt{4N+2}}{2c^2} \frac{2\Pi(c)}{(N+2)^2} \sum_{l\equiv 0[2],\ l\geq 2} \frac{l+N+1/2}{2^{l/2}} \left(c^2+\frac{4c_f}{c_g}(2N+1)-N-N^2\right) \\
&\quad + \frac{c_g\sqrt{4N+2}}{2c^2} \frac{2\Pi(c)}{(N+2)^2} \sum_{l\equiv 0[2],\ l\geq 2} \left(l+N+\frac{1}{2}\right)(l+N)(l+N+1)\frac{1}{2^{l/2}} \\
&\leq \frac{c_g\sqrt{4N+2}\Pi(c)}{2c^2} \Bigg[C_2(c)\left(N(N+1)+c^2+\frac{4c_f}{c_g}(N+1)\right) \\
&\qquad + C_1(c)\left(c^2+\frac{4c_f}{c_g}(2N+1)-N-N^2\right) + 2N+19+\frac{85}{N+2}+\frac{246}{(N+2)^2}\Bigg] \\
&\leq \frac{c_g\Pi(c)}{c^2} C_3(c) \Bigg[N^{5/2}\left(C_2(c)-C_1(c)\right) + N^{3/2}\left(\left(1+\frac{4c_f}{c_g}\right)C_2(c)+\left(\frac{8c_f}{c_g}-1\right)C_1(c)+2\right) \\
&\qquad + \sqrt{N}\left((C_1(c)+C_2(c))\left(c^2+\frac{4c_f}{c_g}\right)+19\right) + \frac{85}{(H(c)+2)^{1/2}} + \frac{246}{(H(c)+2)^{3/2}}\Bigg].
\end{aligned}$$

For the third term in (3.117), using (3.109), the triangle inequality, and (7.74) in Osipov et al. (2013) for the first inequality and using $|\mu_m^c| / |\mu_m^c + \mu_k^c| \leq 1$ for the second, we obtain

$$\left|\int_{-1}^{1} x\frac{\partial\psi_N^c}{\partial c}(x)\frac{\partial\psi_k^c}{\partial x}(x)dx\right| \le \frac{4\left|\psi_N^c(1)\right|\left|\psi_k^c(1)\right|}{|c|}\sum_{m\neq N,\ m\equiv N[2]}\frac{|\mu_N^c|\,|\mu_m^c|\,|\psi_m^c(1)|^2}{\left|(\mu_m^c)^2-(\mu_N^c)^2\right|}\frac{|\mu_m^c|}{|\mu_m^c+\mu_k^c|}$$
$$\le \frac{4\left|\psi_N^c(1)\right|\left|\psi_k^c(1)\right|}{|c|}\mathcal{C}(f,N,c),$$

hence, using (3.111) for the first inequality and (3.106) and (3.111) replacing c_f by c_h for the third,

$$\sum_{k\equiv N[2],\ k\neq N}\frac{2\,|\mu_k^c|}{c\,|\mu_N^c-\mu_k^c|}\left|\int_{-1}^{1} x\frac{\partial\psi_N^c}{\partial c}(x)\frac{\partial\psi_k^c}{\partial x}(x)dx\right|\left\|\psi_k^c\right\|_{L^\infty([-1,1])}$$
$$\le 4c_f\sqrt{4N+2}(C_1(c)+C_2(c))\frac{\Pi(c)}{c^2}\sum_{k\equiv N[2],\ k\neq N}\frac{|\mu_k^c|}{|\mu_N^c-\mu_k^c|}\left|\psi_k^c(1)\right|\left\|\psi_k^c\right\|_{L^\infty([-1,1])}$$
$$\le 4c_f\sqrt{4N+2}(C_1(c)+C_2(c))\frac{\Pi(c)}{c^2}\mathcal{C}(h,N,c)$$
$$\le 4c_hc_f\sqrt{4N+2}(C_1(c)+C_2(c))^2\frac{\Pi(c)^2}{c^2}\le 8c_hc_fC_3(c)(C_1(c)+C_2(c))^2\frac{\Pi(c)^2}{c^2}\sqrt{N}.\quad\square$$

Lemma 18. *For all $u,x_0,R>0$, $t\in\mathbb{R}$, $q\in\{1,\infty\}$, λ from (3.36) and $N(Rx_0U)$ and $\widetilde{\boldsymbol{N}}(q)$ from (3.37), for all $N\geq N(Rx_0U)$, we have*

$$\sup_{\boldsymbol{b}\in[-R,R]^p}\left|\frac{\partial^2}{\partial t^2}\left(\left(\frac{Rx_0t}{2\pi}\right)^{p/2}\lambda(t)\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}\left(\frac{\boldsymbol{b}}{R}\right)\right)\right|\le \mathbb{1}\{U/2\le|t|\le U\}C_8(Rx_0U,p,U)N^{2+k_q/2},$$
$$C_8(Rx_0U,p,U):=\left(\frac{Rx_0U}{\pi}\right)^{p/2}\frac{C_3(Rx_0U)^pN(Rx_0U)^{(p-1)/(2q)}}{N(Rx_0U)^2}\left(\frac{p|p-2|}{U^2}+C_9(U)\frac{2p}{U}+C_{10}(U)\right.$$
$$\left.+\left(\frac{2p}{U}+2C_9(U)\right)\frac{pC_{16}(Rx_0U)}{N(Rx_0U)^2}+\frac{p(p-1)C_{16}(Rx_0U)}{N(Rx_0U)^2}+pC_{11}(Rx_0U)\right),$$
$$C_9(U):=\sup_{t\in[U/2,U]}\left|\lambda'(t)\right|,\ C_{10}(U):=\sup_{t\in[U/2,U]}\left|\lambda''(t)\right|,$$
$$C_{11}(Rx_0U):=\frac{(Rx_0)^2\,\Pi(Rx_0U)}{(Rx_0U)^2}\left(C_4(Rx_0U)+\frac{C_5(Rx_0U)}{N(Rx_0U)}+\frac{C_6(Rx_0U)}{N(Rx_0U)^2}+\frac{C_7(Rx_0U)}{N(Rx_0U)^{5/2}}\right),$$
$$C_{16}(Rx_0U):=2c_fRx_0\left(C_1(Rx_0U)+C_2(Rx_0U)\right)C_3(Rx_0U)\frac{\Pi(Rx_0U)}{Rx_0U}.$$

Proof. Let $q=1$. By symmetry, we take $t\in[U/2,U]$, $\boldsymbol{b}\in[-R,R]^p$, and $c>0$. We have

$$R(t,\boldsymbol{b}) := \left|\frac{\partial^2}{\partial t^2}\left(\left(\frac{Rx_0t}{2\pi}\right)^{p/2}\lambda(t)\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}\left(\frac{\boldsymbol{b}}{R}\right)\right)\right|$$

$$\leq \left(\frac{Rx_0}{2\pi}\right)^{p/2} t^{p/2}\left[\left(\frac{p\,|p-2|}{4t^2}\lambda(t)+\frac{p}{t}\,|\lambda'(t)|+|\lambda''(t)|\right)\left|\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}\left(\frac{\boldsymbol{b}}{R}\right)\right|\right.$$

$$\left.+\,Rx_0\left(\frac{p}{t}\lambda(t)+2|\lambda'(t)|\right)\left|\left.\frac{\partial\psi^c_{\widetilde{\boldsymbol{N}}(q)}}{\partial c}\right|_{c=Rx_0t}\left(\frac{\boldsymbol{b}}{R}\right)\right|+(Rx_0)^2\lambda(t)\left|\left.\frac{\partial^2\psi^c_{\widetilde{\boldsymbol{N}}(q)}}{\partial c^2}\right|_{c=Rx_0t}\left(\frac{\boldsymbol{b}}{R}\right)\right|\right],$$

$$\frac{\partial\psi^c_{\widetilde{\boldsymbol{N}}(q)}}{\partial c}\left(\frac{\boldsymbol{b}}{R}\right)=\sum_{j=2}^{p}\psi^c_N\left(\frac{\boldsymbol{b}_1}{R}\right)\frac{\partial\psi^c_{N(Rx_0U)}}{\partial c}\left(\frac{\boldsymbol{b}_j}{R}\right)\prod_{\substack{l=2\\ l\neq j}}^{p}\psi^c_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right)$$

$$+\frac{\partial\psi^c_N}{\partial c}\left(\frac{\boldsymbol{b}_1}{R}\right)\prod_{l=2}^{p}\psi^c_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right),$$

$$\frac{\partial^2\psi^c_{\widetilde{\boldsymbol{N}}(q)}}{\partial c^2}\left(\frac{\boldsymbol{b}}{R}\right)=2\sum_{j=2}^{p}\frac{\partial\psi^c_{N(Rx_0U)}}{\partial c}\left(\frac{\boldsymbol{b}_j}{R}\right)\frac{\partial\psi^c_N}{\partial c}\left(\frac{\boldsymbol{b}_1}{R}\right)\prod_{\substack{l=2\\ l\neq j}}^{p}\psi^c_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right)$$

$$+\sum_{k=2}^{p}\sum_{\substack{j=2\\ j\neq k}}^{p}\psi^c_N\left(\frac{\boldsymbol{b}_1}{R}\right)\frac{\partial\psi^c_{N(Rx_0U)}}{\partial c}\left(\frac{\boldsymbol{b}_j}{R}\right)\frac{\partial\psi^c_{N(Rx_0U)}}{\partial c}\left(\frac{\boldsymbol{b}_k}{R}\right)\prod_{\substack{l=2\\ l\neq j, l\neq k}}^{p}\psi^{x_0t}_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right)$$

$$+\frac{\partial^2\psi^c_N}{\partial c^2}\left(\frac{\boldsymbol{b}_1}{R}\right)\prod_{l=2}^{p}\psi^c_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right)+\sum_{j=2}^{p}\psi^c_N\left(\frac{\boldsymbol{b}_l}{R}\right)\frac{\partial^2\psi^c_{N(Rx_0U)}}{\partial c^2}\left(\frac{\boldsymbol{b}_j}{R}\right)\prod_{\substack{l=2\\ l\neq j}}^{p}\psi^c_{N(Rx_0U)}\left(\frac{\boldsymbol{b}_l}{R}\right).$$

We conclude using $N \geq Rx_0U$ (by the discussion before Lemma 14), the third assertion of Lemma 15, and Lemma 17. The case $q=\infty$ is obtained with $N(Rx_0U)=N$. □

Lemma 19. *For all $R, x_0 > 0$, $2\sigma > k_q + 4$, $q \in \{1,\infty\}$, $2\tau \geq \left(3e^{1/2}Rx_0/4\right)\vee 1$, we have*

$$\int_{\mathbb{R}} e^{-2\tau|t|}\sum_{\boldsymbol{m}\in\mathbb{N}_0^p}|\boldsymbol{m}|_q^{2\sigma}\left(c^P_{\boldsymbol{m}}(t)\right)^2 dt \leq \frac{C_{12}(\sigma,p)}{\tau p^{2\sigma/q}},$$

$$C_{12}(\sigma,p) := \Gamma(2\sigma+p+1/2)\left(\frac{2^{p-1}p}{2\sigma+p}\left(\frac{8}{3e^{1/2}}\right)^{2\sigma+p}+\frac{\pi e^3 p 2^p\sqrt{3}}{9}\right).$$

Proof. When $q=1$, we use $|\boldsymbol{m}|_1 \leq p\,|\boldsymbol{m}|_\infty$. Let $q=\infty$, R, x_0, σ, τ as in the lemma. Because $P_0 = \mathbb{1}\{|\cdot|_\infty \leq 1\}/2^{p/2}$, for all $m \in \mathbb{N}_0$, $\left|\langle P_0, \psi^c_m\rangle_{L^2([-1,1])}\right| \leq 1$, and, for all

$m > |c|$, $\left|\langle P_0, \psi_m^c\rangle_{L^2([-1,1])}\right| \le |\mu_m^c|/\sqrt{2}$ (see Proposition 3 and (13) in Bonami and Karoui (2014a)) we obtain, for all $t \neq 0$,

$$\sum_{\boldsymbol{m}\in\mathbb{N}_0^p} |\boldsymbol{m}|_\infty^{2\sigma}\left(c_{\boldsymbol{m}}^P(t)\right)^2 \le \sum_{|\boldsymbol{m}|_\infty \le Rx_0|t|} |\boldsymbol{m}|_\infty^{2\sigma} \mathbb{1}\{Rx_0|t|\ge 1\} + \sum_{|\boldsymbol{m}|_\infty > Rx_0|t|} \frac{|\boldsymbol{m}|_\infty^{2\sigma}\left|\mu_{|\boldsymbol{m}|_\infty}^{Rx_0t}\right|^2}{2}. \tag{3.118}$$

Using (3.118), Lemma 12, and $\sum_{|\boldsymbol{m}|_\infty=k} 1 \le p(k+1)^{p-1}$ for the first inequality, $m+1 \le 2m$ when $m \ge 1$ for the second, and $2m+1 \le 3m$, $(Rx_0t+1)^{2\sigma+p} \le (2Rx_0t)^{2\sigma+p}$ when $m, Rx_0t \ge 1$, and (1.3) in Mortici and Chen (2011) for the third, we have

$$\begin{aligned}
&\int_{\mathbb{R}} e^{-2\tau|t|} \sum_{\boldsymbol{m}\in\mathbb{N}_0^p} |\boldsymbol{m}|_\infty^{2\sigma}\left(c_{\boldsymbol{m}}^P(t)\right)^2 dt \\
&\le \int_0^\infty 2pe^{-2\tau t}\left(\sum_{m\le Rx_0t}(m+1)^{p-1}m^{2\sigma}\mathbb{1}\{Rx_0t\ge 1\} + \frac{\pi e^3}{9}\sum_{m>Rx_0t}(m+1)^{p-1}m^{2\sigma}\left(\frac{eRx_0t}{4m}\right)^{2m}\right)dt \\
&\le \int_0^\infty 2^p pe^{-2\tau t}\int_1^{Rx_0t+1} u^{2\sigma+p-1}du\mathbb{1}\{Rx_0t\ge 1\}dt + \frac{\pi e^3 p2^p}{9}\sum_{m\ge 1} m^{2\sigma+p-1}\left(\frac{eRx_0}{4m}\right)^{2m}\int_0^\infty \frac{t^{2m}}{e^{2\tau t}}dt \\
&\le \frac{2^{2(\sigma+p)}p}{2\sigma+p}\int_{1/(Rx_0)}^\infty e^{-2\tau t}(Rx_0t)^{2\sigma+p}dt + \frac{\pi e^3p2^p\sqrt{3}}{9\tau}\sum_{m\ge 1} m^{2\sigma+p-1/2}e^{2m\ln(3Rx_0/(8\tau))} \\
&\le \frac{2^{p-1}p\Gamma(2\sigma+p+1)}{(2\sigma+p)\tau}\left(\frac{8}{3e^{1/2}}\right)^{2\sigma+p} + \frac{\pi e^3p2^p\sqrt{3}}{9\tau}\int_0^\infty e^{-t}t^{2\sigma+p-1/2}dt \le \frac{C_{12}(\sigma,p)}{\tau p^{2\sigma/q}}. \quad \square
\end{aligned}$$

Lemma 20. *For all $N \ge H(Rx_0U)$, $R, U > 0$, $q \in \{1,\infty\}$, and F from (3.35), we have*

$$I_1 := \int_{[-1,1]^p}\int_{\mathbb{R}} \left|\partial_t \mathcal{F}[F](t, x_0t\boldsymbol{x})\right|^2 d\boldsymbol{x}dt \le R^pC_{17}(Rx_0U,p,U)N^2\rho_{\widetilde{\boldsymbol{N}}(q)}^{i_{[-1,1]},Rx_0U} \tag{3.119}$$

$$C_{17}(Rx_0U,p,U) := C_{15}(Rx_0U,p,U) + \frac{2pUC_{16}(Rx_0U)^2}{N(Rx_0U)},$$

$$C_{15}(Rx_0U,p,U) := \frac{25p^2}{8U}\left(1+\frac{2(Rx_0U)^2}{3^{3/2}}\right)^4 + \frac{UC_9(U)^2}{N(Rx_0U)^2} + \frac{5pC_9(U)\ln(2)}{2N(Rx_0U)}\left(1+\frac{2(Rx_0U)^2}{3^{3/2}}\right)^2,$$

$$I_2 := \int_{[-1,1]^p}\int_{\mathbb{R}} \left|\mathcal{F}[F](t, x_0t\boldsymbol{x})\right|^2 d\boldsymbol{x}dt \le R^pU\rho_{\widetilde{\boldsymbol{N}}(q)}^{i_{[-1,1]},Rx_0U}. \tag{3.120}$$

Proof. Let $N \geq H(Rx_0U) \geq 2$. For simplicity of notations, we omit $i_{[-1,1]}$ from ρ. We have

$$\begin{aligned}\mathcal{F}[F](t, x_0t\boldsymbol{x}) &= \left(\frac{Rx_0|t|}{2\pi}\right)^{p/2} R^{p/2}\lambda(t)\mathcal{F}_{Rx_0t}\left[\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}\right](\boldsymbol{x}) \\ &= R^{p/2}i^{\left|\widetilde{\boldsymbol{N}}(q)\right|_1}\lambda(t)\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}(\boldsymbol{x}) \;\; \left(\text{because } \mu^{Rx_0t}_m = i^m\left(\frac{2\pi}{Rx_0|t|}\right)^{1/2}\sqrt{\rho^{Rx_0t}_m}\right).\end{aligned} \tag{3.121}$$

This yields

$$I_1 \leq R^p\int_{\mathbb{R}}\int_{[-1,1]^p}\left(\left(\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}}{\mathrm{d}t}\lambda(t) + \lambda'(t)\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}\right)\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}(\boldsymbol{x}) + \sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}\lambda(t)\frac{\partial\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}(\boldsymbol{x})}{\partial t}\right)^2 dtd\boldsymbol{x}.$$

Using (7.114) in Osipov et al. (2013), cross-products terms are zero and $I_1 \leq R^p(I_{11} + I_{12})$, where

$$I_{11} = \int_{\mathbb{R}}\left(\lambda(t)^2\left(\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}}{\mathrm{d}t}\right)^2 + \lambda'(t)^2\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)} + 2\lambda(t)|\lambda'(t)|\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}}{\mathrm{d}t}\right)dt,$$

$$I_{12} = \int_{\mathbb{R}}\lambda(t)^2\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}\left(\int_{[-1,1]^p}\left(\frac{\partial\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}(\boldsymbol{x})}{\partial t}\right)^2 d\boldsymbol{x}\right)dt.$$

Then, using (7.100) in Osipov et al. (2013) for the second equality yields, for all $t \neq 0$,

$$\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_N}}{\mathrm{d}t} = \frac{x_0R}{2\sqrt{\rho^{Rx_0t}_N}}\left.\frac{\mathrm{d}\rho^c_N}{\mathrm{d}c}\right|_{c=Rx_0t} = \frac{\sqrt{\rho^{Rx_0t}_N}}{|t|}\left(\psi^{Rx_0t}_N(1)\right)^2,$$

in particular $\rho^{Rx_0t}_N$ is increasing in t and, by the first assertion of Lemma 15,

$$\forall U/2 \leq |t| \leq U,\; \frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_N}}{\mathrm{d}t} \leq \frac{(N+1/2)\sqrt{\rho^{Rx_0t}_N}}{|t|}\left(1 + \frac{2(Rx_0U)^2}{3^{3/2}}\right)^2. \tag{3.122}$$

When $q = 1$, using $N + 1/2 \leq 5N/4$ for all $N \geq 2$ and

$$\begin{aligned}\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}}{\mathrm{d}t} =&(p-1)\left(\sqrt{\rho^{Rx_0t}_{N(Rx_0U)}}\right)^{p-2}\sqrt{\rho^{Rx_0t}_{N}}\left(\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{N(Rx_0U)}}}{\mathrm{d}t}\right)\\ &+\left(\sqrt{\rho^{Rx_0t}_{N(Rx_0U)}}\right)^{p-1}\left(\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{N}}}{\mathrm{d}t}\right),\end{aligned}$$

we have

$$\frac{\mathrm{d}\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}}{\mathrm{d}t} \leq \frac{5pN}{4\,|t|}\left(1+\frac{2(Rx_0U)^2}{3^{3/2}}\right)^2\sqrt{\rho^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}}. \tag{3.123}$$

The same inequality holds for $q = \infty$ (there $N = N(Rx_0U)$). This yields, for all $q \in \{1, \infty\}$,

$$\begin{aligned}&I_{11}\\ &\leq\left(\frac{25p^2N^2}{8}\int_{U/2}^{U}\frac{dt}{t^2}\left(1+\frac{2(Rx_0U)^2}{3^{3/2}}\right)^4+UC_9(U)^2\right)\rho^{Rx_0U}_{\widetilde{\boldsymbol{N}}(q)}\\ &+\frac{5pNC_9(U)}{2}\left(1+\frac{2(Rx_0U)^2}{3^{3/2}}\right)^2\int_{U/2}^{U}\frac{dt}{t}\rho^{Rx_0U}_{\widetilde{\boldsymbol{N}}(q)}\\ \leq& C_{15}(Rx_0U,p,U)N^2\rho^{Rx_0U}_{\widetilde{\boldsymbol{N}}(q)}.\end{aligned}$$

Then, by (7.114) in Osipov et al. (2013) and Lemma 16, we have, for all $U/2 \leq |t| \leq U$,

$$\begin{aligned}\int_{[-1,1]^p}\left(\frac{\partial\psi^{Rx_0t}_{\widetilde{\boldsymbol{N}}(q)}(\boldsymbol{x})}{\partial t}\right)^2 d\boldsymbol{x} =&(Rx_0)^2\int_{[-1,1]}(p-1)\left(\left.\frac{\partial\psi^{c}_{N(Rx_0U)}(x)}{\partial c}\right|_{c=Rx_0t}\right)^2 dx\\ &+(Rx_0)^2\int_{[-1,1]}\left(\left.\frac{\partial\psi^{c}_{N}(x)}{\partial c}\right|_{c=Rx_0t}\right)^2 dx\\ \leq& 2p\left(C_{16}(Rx_0U)\right)^2 N \quad (\text{using } N \geq N(Rx_0U)).\end{aligned}$$

The same holds for $q = \infty$ (there $N = N(Rx_0U)$). This and $N \geq N(Rx_0U)$ yield (3.119).

(3.120) follows from (3.121) and the fact that $c \in (0,\infty) \mapsto \rho^c_m$ is nondecreasing. □

3.7.3 Estimation of the marginal f_{β}

For all $(\omega_m)_{m\in\mathbb{N}_0}$ increasing, $\omega_0=1$, $l,M>0$, $q\in\{1,\infty\}$, consider

$$\mathcal{H}^{q,\omega}_{w,W}(l,M):=\left\{f:\|f\|_{L^2(w\otimes W^{\otimes p})}\le M,\ \sum_{k\in\mathbb{N}_0}\omega_k^2\left\|\theta_{q,k}\right\|^2_{L^2(\mathbb{R})}\le 2\pi l^2\right\}.$$

For brevity, we present the slow rates and the estimator $\widehat{f}_{\beta}^{q,N,\epsilon}:=\sum_{|\boldsymbol{m}|_q\le N(\epsilon)}\widehat{c}_{\boldsymbol{m}}(\epsilon)\varphi_{\boldsymbol{m}}^{W,\epsilon x_0}/\sigma_{\boldsymbol{m}}^{W,\epsilon x_0}$. It is based on $f_{\boldsymbol{\beta}}=\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](0,\cdot_2)$.

Proposition 6. *Let $W=i_{[-R,R]}$. For all $q\in\{1,\infty\}$, $l,M,R>0$, $\sigma>2$, $\mathbb{S}_\beta\subseteq[-R,R]^p$, $\underline{N}$ solution of $2(1+\sigma)k_q(\underline{N}+k_q'/2)\ln(\underline{N})+p(1-\sigma)\ln(\underline{N})+\ln(\omega_{\underline{N}}^2)=\ln(n_e)$, $\epsilon=\theta/\omega_{\underline{N}}$, $(\omega_k)_{k\in\mathbb{N}_0}=(k^\sigma)_{k\in\mathbb{N}_0}$, and w such that $\int_{\mathbb{R}}a^2/w(a)da<\infty$, we have*

$$\left(\frac{\ln(n_e)}{\ln_2(n_e)}\right)^{2\sigma}\sup_{f_{\boldsymbol{\beta}}\in\mathcal{H}^{q,\omega}_{w,W}(l,M)\cap D,\ f_{\boldsymbol{X}|\mathcal{X}}\in\mathcal{E}}\mathbb{E}\left[\left\|\widehat{f}_{\boldsymbol{\beta}}^{q,N,\epsilon}-f_{\boldsymbol{\beta}}\right\|^2_{L^2(\mathbb{R}^p)}\right]=O_p(1).$$

Proof. We assume $f_{\boldsymbol{X}|\mathcal{X}}$ is known. The general case can be handled like in the proof of (T2.1). Use $f_{\boldsymbol{\beta}}^{\epsilon}:=\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](\epsilon,\cdot)$ and define $f_{\boldsymbol{\beta}}^{q,\epsilon,N}$ like $\widehat{f}_{\boldsymbol{\beta}}^{q,\epsilon,N}$ with $\widetilde{c}_{\boldsymbol{m}}(t)$ (see Lemma 2) instead of $\widehat{c}_{\boldsymbol{m}}(t)$. Use $\left\|\widehat{f}_{\boldsymbol{\beta}}^{q,N,\epsilon}-f_{\boldsymbol{\beta}}\right\|^2_{L^2(\mathbb{R}^p)}\le 3\sum_{j=1}^3\|R_j\|^2_{L^2(\mathbb{R}^p)}$, where $R_1:=\widehat{f}_{\boldsymbol{\beta}}-f_{\boldsymbol{\beta}}^{q,N,\epsilon}$, $R_2:=f_{\boldsymbol{\beta}}^{q,N,\epsilon}-f_{\boldsymbol{\beta}}^{\epsilon}$, and $R_3:=f_{\boldsymbol{\beta}}^{\epsilon}-f_{\boldsymbol{\beta}}$. Let $n\ge e^e$ large enough so that $N\ge 1\bigvee((\sigma-1)p-\sigma)/(2k_q)$. By similar arguments from (3.59), (3.57), $N\le\underline{N}$, and $(\underline{N}+1)^{2k_q\underline{N}+p}\le e^{2k_q}2^p\underline{N}^{2k_q\underline{N}+p}$, we have

$$\mathbb{E}\left[\|R_1\|^2_{L^2(\mathbb{R}^p)}\right]\le\frac{Q_qc_{\boldsymbol{X}}e^{2k_q}2^p}{\pi^pn}\epsilon^p\underline{N}^p\left(1\bigvee\frac{\theta\underline{N}}{\epsilon}\right)^{2k_q\underline{N}+p}=\frac{Q_qc_{\boldsymbol{X}}e^{2k_q}2^p\theta^p}{\pi^pn}\underline{N}^{p(1-\sigma)+2(1+\sigma)k_q(\underline{N}+k_q'/2)}.\tag{3.124}$$

We also obtain $\|R_2\|^2_{L^2(\mathbb{R}^p)}\le 2\pi l^2/\omega_{\underline{N}}^2$ and

$$\begin{aligned}\|R_3\|^2_{L^2(\mathbb{R}^p)}\le&\int_{[-R,R]^p}\left(\int_{\mathbb{R}}\left|e^{i\epsilon a}-1\right|f_{\alpha,\boldsymbol{\beta}}(a,\boldsymbol{b})da\right)^2d\boldsymbol{b}\\ \le&\epsilon^2\int_{[-R,R]^p}\left(\int_{\mathbb{R}}|a|\,f_{\alpha,\boldsymbol{\beta}}(a,\boldsymbol{b})da\right)^2d\boldsymbol{b}\le\frac{\theta^2M^2}{\omega_{\underline{N}}^2}\int_{\mathbb{R}}\frac{a^2}{w(a)}da<\infty.\end{aligned}\tag{3.125}$$

Then, using $\ln(n)=2(1+\sigma)k_q(\underline{N}+k_q'/2)\ln(\underline{N})+p(1-\sigma)\ln(\underline{N})+\ln(\omega_{\underline{N}}^2)\ge 2\sigma k_q\underline{N}\ln(\underline{N})$

and $\mathcal{W}(x) \leq \ln(x+1)$, we have $\underline{N} \leq \ln(n)/(2\sigma k_q \ln_2(n_e)(1+\ln(1+2\sigma k_q/e))$. The result follows from the definition of $\underline{N}$, (3.124), and (3.125). □

Similar ideas apply for the estimation of $f_{\boldsymbol{\beta}_j}$ for $j=1,\dots,p$.

3.7.4 Talagrand inequality for complex functions

Lemma 21. *Let $X_1,\dots,X_n$ n independent random vectors, $\Lambda := (\sqrt{1+\cdot}-1)\wedge 1$, $\mathcal{U}$ a countable set of complex measurable functions, and, for all $u\in\mathcal{U}$, $\nu_n(u) := \sum_{i=1}^n \left(u(X_i)-\mathbb{E}\left[u(X_i)\right]\right)/n$. If there exist $M,H,v>0$ such that*

$$\sup_{u\in\mathcal{U}}\|u\|_{L^\infty(\mathbb{R}^p)}\leq M,\ \mathbb{E}\left[\sup_{u\in\mathcal{U}}|\nu_n(u)|\right]\leq H,\ \sup_{u\in\mathcal{U}}\frac{1}{n}\sum_{i=1}^n Var\left(\Re(u(X_i))\right)\bigvee Var\left(\Im(u(X_i))\right)\leq v,$$

then, for all $\eta>0$,

$$\mathbb{E}\left[\left(\sup_{u\in\mathcal{U}}|\nu_n(u)|^2-4(1+2\eta)H^2\right)_+\right]\leq 48\left(\frac{v}{n}e^{-\eta\frac{nH^2}{6v}}+\frac{294M^2}{\Lambda(\eta)^2n^2}e^{-\frac{\sqrt{2}\Lambda(\eta)\sqrt{\eta}}{42}\frac{nH}{M}}\right).$$

Proof. The result follows from Theorem 7.3 in Comte and Genon-Catalot (2018) and

$$\mathbb{E}\left[\left(\sup_{u\in\mathcal{U}}|\nu_n(u)|^2-4(1+2\eta)H^2\right)_+\right]\leq\mathbb{E}\left[\left(\sup_{u\in\mathcal{U}}\Re(\nu_n(u))^2+\sup_{u\in\mathcal{U}}\Im(\nu_n(u))^2-4(1+2\eta)H^2\right)_+\right]$$
$$\leq\mathbb{E}\left[\left(\sup_{u\in\mathcal{U}}\Re(\nu_n(u))^2-2(1+2\eta)H^2\right)_+\right]+\mathbb{E}\left[\left(\sup_{u\in\mathcal{U}}\Im(\nu_n(u))^2-2(1+2\eta)H^2\right)_+\right].$$ □

3.7.5 Approximation by PSWF in Sobolev ellipsoids.

For all $\sigma,s,l>0$ and $q\in\{1,\infty\}$, denote by $\left(\phi_{\boldsymbol{m}}\left(\cdot/R\right)\right)_{\boldsymbol{m}\in\mathbb{Z}^p} := \left(e^{i\pi\boldsymbol{m}^\top\cdot/R}/(2R)^{p/2}\right)_{\boldsymbol{m}\in\mathbb{Z}^p}$, $\mathcal{F}[f](\star,\boldsymbol{k}) := \int_{\mathbb{R}}e^{i\star a}\int_{[-R,R]^p}e^{i\pi\boldsymbol{k}^\top\boldsymbol{b}/R}\ f(a,\boldsymbol{b})da d\boldsymbol{b}/(2R)^{p/2}$, and

$$H^{q,s,\sigma}(l) := \left\{f:\int_{\mathbb{R}}\sum_{\boldsymbol{k}\in\mathbb{Z}^p}|\mathcal{F}[f](t,\boldsymbol{k})|^2\left(1\vee t^{2s}\right)dt\bigvee\int_{\mathbb{R}}\sum_{\boldsymbol{k}\in\mathbb{Z}^p}|\mathcal{F}[f](t,\boldsymbol{k})|^2\left(1\vee|\boldsymbol{k}|_q^{2\sigma}\right)dt\leq 2\pi l^2\right\}.$$

Denote, for all $N \in \mathbb{N}$ and $c \neq 0$, by P_c^N (resp. $\mathcal{E}^N$) the projector in $L^2\left(i_{[-R,R]}^{\otimes p}\right)$ onto the vector space spanned by $\left(\psi_{\boldsymbol{m}}^c(\cdot/R)/R^{p/2}\right)_{|\boldsymbol{m}|_\infty<N}$ (resp. $\left(\phi_{\boldsymbol{m}}(\cdot/R)\right)_{|\boldsymbol{m}|_\infty<N}$). For all $t \neq 0$ and $\left(n,m,N,\widetilde{N}\right) \in \mathbb{N}_0^4$, denote by $\varphi^t := \mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2)$, $\beta_n^m(t) := \langle \psi_m^t, P_n\rangle_{L^2([-1,1])}$, J_j the Bessel function of the first kind and order $j > -1$, $K_t^{N,\widetilde{N}} := \left\|\mathcal{E}^{\widetilde{N}}\varphi^t - P_{x_0t}^N\mathcal{E}^{\widetilde{N}}\varphi^t\right\|^2$, and $I_{N,\widetilde{N}} := \sum_{\boldsymbol{k}\in\mathbb{Z}^p:\ |\boldsymbol{k}|_\infty<\widetilde{N}} \sum_{|\boldsymbol{m}|_\infty\geq N} |\langle \phi_{\boldsymbol{k}}(\cdot/R), \psi_{\boldsymbol{m}}^{x_0\star}(\cdot/R)\rangle|^2$.

Proposition 7. *For all $\sigma, l, M, R > 0$, $q \in \{1,\infty\}$, and $s \geq \sigma + p/2$, we have, for all $N \geq 10$,*

$$\int_{\mathbb{R}} \left\|\mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2) - P_{x_0t}^N\mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2)\right\|^2 dt \leq \frac{2\pi A l^2}{N^{2\sigma}}, \tag{3.126}$$

$$\begin{aligned}
A :=& 2\left(\left(\frac{1}{1/(\pi e)-1/10}\right)^{2\sigma} + c\left(\frac{p+2\sigma}{be}\right)^{p+2\sigma} + \left(\frac{2eRx_0}{\pi}\right)^p (e^2x_0)^{2\sigma}\right),\\
a :=& \frac{\sqrt{5e^3}(e^2+1/e^2)^{5/8}}{3(\ln(2)+2)2^{11/4}}, b := p\left(\frac{5}{8}\ln\left(\frac{21}{10}\right) - \frac{1}{e}\right),\\
c :=& \frac{p(4R^2/(\pi e))^p}{2R}\left(a^{2p}\frac{8^p(p-1)^{p-1}}{(3p)^p e^{p-1}} + \left(\frac{(2p-1)8}{5pe}\right)^{2p-1}\frac{5^{p-1}8}{p16^p\ln(21/10)}\right).
\end{aligned}$$

Proposition 7 is an analogue of (3.63) with N constant and $(\omega_k)_{k\in\mathbb{N}_0} = (k^\sigma)_{k\in\mathbb{N}_0}$. It shows that the approximation error when we use a truncated series expansion in the PSWF basis is of order $N^{-2\sigma}$ whether we work on the class $H^{q,s,\sigma}(l)$ or $\mathcal{H}_{w,i_{[-R,R]}}^{q,\phi,\omega}(l,M)$ with $\phi = 1 \vee |\cdot|^s$. (3.64) can be obtained using the first inequality in the definition of $H^{q,s,\sigma}(l)$ and, for all $t \neq 0$, $\sum_{\boldsymbol{k}\in\mathbb{Z}^p}|\mathcal{F}[f](t,\boldsymbol{k})|^2 = \|\mathcal{F}_{1\mathrm{st}}[f](t,\cdot_2)\|^2_{L^2\left(i_{[-R,R]}^{\otimes p}\right)} = \sum_{\boldsymbol{m}\in\mathbb{N}_0^p}|b_{\boldsymbol{m}}(t)|^2$. Thus (T2.1) also holds for functions in the intersection of

$$H^{q,s,\sigma}(l)\bigcap\left\{f:\ \|f\|_{L^2\left(w\otimes i_{[-R,R]}^{\otimes p}\right)} \leq M\right\}.$$

The proof below uses techniques from the proof of Lemma 11 in Bonami and Karoui (2014*a*).

Proof. In this proof, $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote the scalar product and norm in $L^2([-R,R]^p)$.

Take $f \in H^{q',s,\sigma'}(l)$. Let $N \geq 10$ and $\widetilde{N} := \lfloor \tau N \rfloor$, where $\tau := 1/(\pi e)$. We have

$$\begin{aligned}\left\|\varphi^t - P_{x_0t}^N \varphi^t\right\|^2 &\leq 2\left(\left\|\varphi^t - \mathcal{E}^{\widetilde{N}}\varphi^t - P_{x_0t}^N\left(\varphi^t - \mathcal{E}^{\widetilde{N}}\varphi^t\right)\right\|^2 + K_t^{N,\widetilde{N}}\right)\\ &\leq 2\left(\left\|\varphi^t - \mathcal{E}^{\widetilde{N}}\varphi^t\right\|^2 + K_t^{N,\widetilde{N}}\right). \qquad (3.127)\end{aligned}$$

Using that $\left(\psi_{\boldsymbol{m}}^{x_0t}(\cdot/R)/R^{p/2}\right)_{\boldsymbol{m}\in\mathbb{N}_0^p}$ are orthonormal in $L^2([-R,R]^p)$ and the Cauchy-Schwarz inequality in the second display yield

$$\begin{aligned}K_t^{N,\widetilde{N}} &= \left\|\sum_{\boldsymbol{k}\in\mathbb{Z}^p:\ |\boldsymbol{k}|_\infty<\widetilde{N}}\left\langle\varphi^t,\phi_{\boldsymbol{k}}\left(\frac{\cdot}{R}\right)\right\rangle\left(\sum_{|\boldsymbol{m}|_\infty\geq N}\left\langle\phi_{\boldsymbol{k}}\left(\frac{\cdot}{R}\right),\psi_{\boldsymbol{m}}^{x_0t}\left(\frac{\cdot}{R}\right)\right\rangle\psi_{\boldsymbol{m}}^{x_0t}\left(\frac{\star}{R}\right)\frac{1}{R^p}\right)\right\|^2\\ &\leq \sum_{\boldsymbol{k}\in\mathbb{Z}^p:\ |\boldsymbol{k}|_\infty<\widetilde{N}}\left|\left\langle\varphi^t,\phi_{\boldsymbol{k}}\left(\frac{\cdot}{R}\right)\right\rangle\right|^2 I_{N,\widetilde{N}}(t) \leq \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^p}|\mathcal{F}[f](t,\boldsymbol{k})|^2\right) I_{N,\widetilde{N}}(t).\end{aligned}$$

We have, using (18.17.19) in Olver et al. (2010) for the first equality and for all $k \in \mathbb{Z}$ and $m \in \mathbb{N}_0$,

$$\begin{aligned}&\left|\left\langle\phi_k\left(\frac{\cdot}{R}\right),\psi_m^{x_0t}\left(\frac{\cdot}{R}\right)\right\rangle\right|^2 = \frac{R}{2}\left|I_{m,k}+O_{m,k}\right|^2 \leq R\left(|I_{m,k}|^2+|O_{m,k}|^2\right),\\ &I_{m,k} := \sum_{n=0}^{\lfloor 5m/8\rfloor-1}\beta_n^m(x_0t)\left\langle e^{i\pi k\cdot},P_n\right\rangle_{L^2([-1,1])},\\ &O_{m,k} := \sum_{n\geq\lfloor 5m/8\rfloor}\beta_n^m(x_0t)i^n\sqrt{\frac{2}{|k|}}\sqrt{n+\frac{1}{2}}J_{n+1/2}(|k|\pi).\end{aligned}$$

Using, for all $k \in \mathbb{Z}$, $\left|\left\langle e^{i\pi k\cdot},P_n\right\rangle\right| \leq \sqrt{2}$, Proposition 3 in Bonami and Karoui (2014a), and integral test for convergence (indeed, by (3.4) page 34 in Osipov et al. (2013), for all $m \geq 2\vee(e^2x_0|t|)$, $2\sqrt{\chi_m^{x_0t}}/(x_0|t|) \geq 2e^2 > 1$), we obtain, for all $m \geq 2\vee(e^2x_0|t|)$,

$$\begin{aligned}|I_{m,k}| &\leq \sqrt{\frac{5}{2\pi}}\int_0^{\lfloor 5m/8\rfloor}\left(\frac{2\sqrt{\chi_m^{x_0t}}}{x_0|t|}\right)^x dx\left|\mu_m^{x_0t}\right|\\ &\leq \frac{\sqrt{5/(2\pi)}}{\ln\left(2\sqrt{\chi_m^{x_0t}}/(x_0|t|)\right)}\left(\frac{2\sqrt{\chi_m^{x_0t}}}{x_0|t|}\right)^{\lfloor 5m/8\rfloor}\left|\mu_m^{x_0t}\right|.\end{aligned}$$

Let $m \geq 2 \vee (e^2 x_0 |t|)$. Using Lemma 12 for the first inequality, we obtain

$$\begin{aligned}
|I_{m,k}| &\leq \frac{\sqrt{5e^3}}{3} \frac{1}{\ln(2)+2} \left(\frac{2\sqrt{m(m+1)+x_0^2 t^2}}{x_0 |t|} \right)^{5m/8} \left(\frac{e x_0 |t|}{4(m+3/2)} \right)^m \\
&\leq \frac{\sqrt{5e^3}}{3} \frac{1}{\ln(2)+2} \left(\frac{\sqrt{e^2+1/e^2}(m+1)}{2^{11/5}(m+3/2)} \right)^{5m/8} \exp\left(-\frac{3m}{8} \ln\left(\frac{m}{e x_0 |t|} \right) \right) \\
&\leq a \exp\left(-\frac{3m}{8} \ln\left(\frac{m}{e x_0 |t|} \right) \right).
\end{aligned}$$

Using, for all $j > -1/2$, $x \in \mathbb{R}$, and $n \in \mathbb{N}_0$, $|J_j(x)| \leq |x|^j / (2^j \Gamma(j+1))$ (see 9.1.20 in Abramowitz and Stegun (1965)), $|\beta_n^m(x_0 t)| \leq 1$, and $\sqrt{n+1/2} < \Gamma(n+3/2)/n!$ (see (5.6.4) in Olver et al. (2010)) for the first inequality and $m > 8/5$ and $n! \geq (n/e)^n \sqrt{2\pi n}$ for the third, we obtain, for all $k \in \mathbb{Z}$,

$$\begin{aligned}
|O_{m,k}| &\leq \sum_{n \geq \lfloor 5m/8 \rfloor} \frac{\sqrt{\pi}}{n!} \left(\frac{|k|\pi}{2} \right)^n \\
&\leq \frac{\sqrt{\pi}}{\lfloor 5m/8 \rfloor!} \left(\frac{|k|\pi}{2} \right)^{\lfloor 5m/8 \rfloor} \exp\left(\frac{|k|\pi}{2} \right) \leq \sqrt{\frac{5m}{16}} \left(\frac{|k|\pi e}{2(5m/8-1)} \right)^{5m/8} \exp\left(\frac{|k|\pi}{2} \right).
\end{aligned}$$

Using $\left|\langle \phi_k(\cdot/R), \psi_m^{x_0 t}(\cdot/R) \rangle\right|^2 \leq R$ for all $(k,m) \in \mathbb{N}_0^2$ for the first inequality, $\sum_{|\boldsymbol{m}|_\infty = j} 1 \leq p(j+1)^{p-1}$ for the second, (3.57) and the convexity of $x \mapsto x^p$ for the fourth inequality, we have, for all t such that $N \geq e^2 x_0 |t|$,

$$\begin{aligned}
I_{N,\widetilde{N}}(t) &\leq R^{p-1} \sum_{\boldsymbol{k} \in \mathbb{Z}^p:\ |\boldsymbol{k}|_\infty < \widetilde{N}} \sum_{j=N}^{\infty} \sum_{|\boldsymbol{m}|_\infty = j} \prod_{l=1}^{p} \left| \left\langle \phi_{\boldsymbol{k}_l}\left(\frac{\cdot}{R}\right), \psi_j^{x_0 t}\left(\frac{\cdot}{R}\right) \right\rangle \right|^2 \\
&\leq R^{2p-1} \sum_{\boldsymbol{k} \in \mathbb{Z}^p:\ |\boldsymbol{k}|_\infty < \widetilde{N}} \sum_{j=N}^{\infty} p(j+1)^{p-1} \prod_{l=1}^{p} \left(|I_{j,\boldsymbol{k}_l}|^2 + |O_{j,\boldsymbol{k}_l}|^2 \right) \\
&\leq pR^{2p-1} \sum_{\boldsymbol{k} \in \mathbb{Z}^p:\ |\boldsymbol{k}|_\infty < \widetilde{N}} \sum_{j=N}^{\infty} (j+1)^{p-1} \left(a^2 \left(\frac{e x_0 |t|}{j} \right)^{3j/4} + \frac{5j e^{|\boldsymbol{k}_l|\pi}}{16} \left(\frac{|\boldsymbol{k}_l|\pi e}{2(5j/8-1)} \right)^{5j/4} \right) \\
&\leq \frac{p(4R^2\tau N)^p}{2R} \sum_{j=N}^{\infty} j^{p-1} a^{2p} \left(\frac{e x_0 |t|}{j} \right)^{3pj/4} + \left(\frac{5 e^{\widetilde{N}\pi} j^2}{16} \right)^p \left(\frac{\widetilde{N}\pi e}{2(5j/8-1)} \right)^{5pj/4} \frac{1}{j}.
\end{aligned}$$

Using $\kappa(t) := -3\ln(e x_0 |t|/N)/8$, $\kappa(t) \geq 3/8$ for $N \geq 2 \vee (e^2 x_0 |t|)$, and $\sup_{j \geq 1} j^{p-1} e^{-p\kappa(t) j} =$

$(1-1/p)^{p-1}/(\kappa(t)e)^{p-1}$ for the second inequality, we obtain, for all $N \geq e^2 x_0 |t|$,

$$\sum_{j=N}^{\infty} j^{p-1} \left(\frac{ex_0 |t|}{j}\right)^{3pj/4} \leq \frac{(1-1/p)^{p-1}}{(\kappa(t)e)^{p-1}} \int_N^\infty e^{-p\kappa(t)j} dj \leq \frac{(p-1)^{p-1}}{(p\kappa(t))^p e^{p-1}} e^{-p\kappa(t)N}.$$

Using $1 - 8/(5N) \geq 1/5$, that for $j \geq N$, $\widetilde{N}\pi e/(2(5j/8-1)) \leq 10\tau\pi e/21 = 10/21$, and $\sup_{j\geq 1} j^{2p-1} e^{-5pj/8} = ((2p-1)8/(5pe))^{2p-1}$ for the first inequality, we obtain

$$\begin{aligned}\sum_{j=N}^{\infty} j^{2p-1} \left(\frac{\widetilde{N}\pi e}{2(5j/8-1)}\right)^{5pj/4} &\leq \left(\frac{(2p-1)8}{5pe}\right)^{2p-1} \int_N^\infty e^{-5pj\ln(21/10)/8} dj \\ &\leq \left(\frac{(2p-1)8}{5pe}\right)^{2p-1} \frac{8}{5p\ln(21/10)} e^{-5pN\ln(21/10)/8}.\end{aligned}$$

Using (3.127) for the first display, using $\sup_{t:\ |t| \leq N/(e^2x_0)} I_{N,\widetilde{N}}(t) \leq cN^p e^{-bN}$ (because $5\ln(21/10)/8 - \tau\pi < 3/8 \leq \kappa(t)$), using $s \geq \sigma + p/2$, $f \in H^{q,s,\sigma}(l)$, and, for all $|t| > N/(e^2x_0)$, $I_{N,\widetilde{N}}(t) \leq R^p \sum_{\boldsymbol{k}\in\mathbb{Z}^p:\ |\boldsymbol{k}|_\infty < \widetilde{N}} \|\phi_{\boldsymbol{k}}\|^2_{L^2([-1,1]^p)} \leq (2\tau NR)^p$ for the second, we have

$$\begin{aligned}&\int_{\mathbb{R}} \left\|\mathcal{F}_{1\text{st}}[f](t,\cdot_2) - P^N_{x_0t}\mathcal{F}_{1\text{st}}[f](t,\cdot_2)\right\|^2 dt \\ &\leq 2\left\|\mathcal{F}_{1\text{st}}[f] - \mathcal{E}^{\widetilde{N}}\mathcal{F}_{1\text{st}}[f]\right\|^2_{L^2(\mathbb{R}\times[-R,R]^p)} \\ &\quad + 2\int_{-N/(e^2x_0)}^{N/(e^2x_0)} \sum_{\boldsymbol{k}\in\mathbb{Z}^p} |\mathcal{F}_{1\text{st}}[f](t,\boldsymbol{k})|^2 dt \sup_{t:\ |t|\leq N/(e^2x_0)} I_{N,\widetilde{N}}(t) \\ &\quad + \frac{2(2\tau NR)^p}{1 \vee (N/(e^2x_0))^{2s}} \int_{|t|>N/(e^2x_0)} \sum_{\boldsymbol{k}\in\mathbb{Z}^p} |\mathcal{F}_{1\text{st}}[f](t,\boldsymbol{k})|^2 (1 \vee t^{2s}) dt \\ &\leq \frac{4\pi l^2}{(\tau N-1)^{2\sigma}} + \frac{4\pi l^2 cN^p}{e^{bN}} + \frac{4\pi l^2 (2\tau Re^2x_0)^p (e^2x_0)^{2\sigma}}{N^{2\sigma}}.\end{aligned}$$

Using $\tau - 1/10 > 0$ and (3.29) yield the result. □

Chapter 4

Nonparametric Ecological Inference with an Application to Electoral Studies

Abstract

This paper considers a nonparametric framework for the prediction of contingency tables for given values of the marginal distributions of two discrete variables from a dataset of margins in different groups. My methods are evaluated on a real dataset to predict the probability to vote according to race. It is related to the prediction of the random coefficients given values of the outcomes and regressors in a system of linear random coefficients equations with bounded regressors. I characterize the identified set without further assumptions and show that the *no contextual effects* (NCE) assumption leads to new constructive point identification. Then, I develop a minimax adaptive estimator. I evaluate the performances of the methods on real datasets where the true value of the parameters is known. This shows that the NCE assumption is reasonable here. Finally, I apply my methods to estimate the effect of door-to-door visits on vote shares among two categories of voters, based on past votes, in the randomized experiment of Pons (2018). My results suggest that canvassing is

especially effective through persuasion of undecided voters, rather than mobilization of convinced ones.

Keywords: Ecological Inference, Random Coefficients, Data Combination, Partial Identification, Voting Experiments.

Introduction

A common empirical issue is to observe a sample of marginal distributions of two individual discrete variables $C \in \{1, \ldots, d_C\}$ and $R \in \{1, \ldots, d_R\}$ over the same groups of individuals g, while the distributions of C conditional on R for the various groups remain unknown. A simple yet striking illustration is the probability to go to the polls, where $d_C = 2$, according to race R for given precincts g. In this example, the precincts correspond to the groups and the conditional probabilities are usually unobserved. Nevertheless, one can combine the margins through the precincts. Here, the margins are the turnout rates and the racial composition of each precinct, respectively

$$\boldsymbol{Y}_g := \begin{pmatrix} \mathbb{P}_g(C=1) \\ \vdots \\ \mathbb{P}_g(C=d_C) \end{pmatrix} \in [0,1]^{d_C} \quad \text{and} \quad \boldsymbol{X}_g := \begin{pmatrix} \mathbb{P}_g(R=1) \\ \vdots \\ \mathbb{P}_g(R=d_R) \end{pmatrix} \in [0,1]^{d_R},$$

where the coordinates of $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$ sum to 1. The former is provided by the election returns while the later is provided from the census. The conditional distributions $\boldsymbol{B}_g$, or equivalently - as the margins are known - the contingency tables, are matrices with d_R rows and d_C columns which contains the probabilities to vote or not conditional on race, namely $\boldsymbol{B}_{r,c,g} := \mathbb{P}_g(C = c|R = r)$. This problem is also encountered in other contexts such as combining market level data with census data to perform demand analysis. It is also the case for randomized experiments with protected variables. The sheer amount of aggregate public data available, the public awareness of the need to protect data privacy, and the enormous cost of obtaining individual data at large scale all stress the importance of econometric tools for making this type of *ecological* inference, which uses data at the group level (see, *e.g.*, Robinson, 1950; King, 1997).

The coefficients of the matrix $\boldsymbol{B}_g$ are the outcome probabilities conditional on the covariate for group g, hence have to satisfy the law of total probability, for all g,

$$\forall c = 1, \dots, d_C, \quad \mathbb{P}_g\left(C = c\right) = \sum_{r=1}^{d_R} \mathbb{P}_g\left(C = c | R = r\right) \mathbb{P}_g\left(R = r\right), \tag{4.1}$$

which can be rewritten

$$\boldsymbol{Y}_g = \boldsymbol{B}_g^\top \boldsymbol{X}_g.$$

However, without additional restrictions, for each group, there are many possible tables compatible with the observed margins $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$. This is the fundamental indeterminacy of the ecological inference. There are two ways to handle this heterogeneity in the literature, which are related.

These tables and margins are heterogeneous across groups hence the point of view which I follow in this paper treats the observed sample of margins for the groups, together with the unobserved and heterogeneous conditional distributions, as random vectors and matrices drawn from a sampling distribution

$$(\boldsymbol{B}_g, \boldsymbol{X}_g, \boldsymbol{Y}_g) \sim \mathbb{P}_{\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y}},$$

where $\mathbb{P}_{\boldsymbol{Z}}$ denotes the law of a random vector $\boldsymbol{Z}$ (see also, *e.g.*, King, 1997; Wakefield, 2004; Imai et al., 2008, in the Statistics and Political Science literatures). This paper then considers inference on the conditional expectation

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}) \in [0,1]^{d_R \times d_C} \mapsto \mathbb{E}\left[\boldsymbol{B} | (\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y})\right] \in \mathcal{M}_{d_R, d_C}([0,1]), \tag{4.2}$$

where $\mathcal{M}_{d_R, d_C}([0,1])$ are $d_R \times d_C$ matrices with elements in $[0,1]$. This gives the best prediction of the heterogeneous conditional distribution $\boldsymbol{B}$ for given values of the margins $(\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y})$.[1] Another way to interpert this parameter $\boldsymbol{m}$ is that, if $\boldsymbol{B}$

[1]The posterior average effects studied in a Bayesian context in Bonhomme and Weidner (2019) shares some similarities with $\boldsymbol{m}$.

were observed, $\boldsymbol{m}$ would be a local average of $\boldsymbol{B}$ over the groups whose margins are close to $(\boldsymbol{x}, \boldsymbol{y})$. This translates in the example as the prediction of the probability to vote at an election for people of a certain race given the turnout rate and racial composition. A second point of view, used in Econometrics (see, *e.g.*, Cross and Manski, 2002*a*; Molinari and Peski, 2006*a*; Fan et al., 2014, 2016; Manski, 2018; Jiang et al., 2020; Gaillac, 2020), rather looks for all the conditional matrices $\boldsymbol{B}_g$ which are compatible with given margins $\boldsymbol{Y}_g$ and $\boldsymbol{X}_g$. However, this often yields identified sets which are too large for the practitioners.

This paper first studies identification of (4.2) and shows how the two above points of view are related. The law of total probability (4.1), together with the constraints on the margins, yield that $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y})$ satisfies exactly the linear system of random coefficients equations

$$\forall c = 1, \ldots, d_C, \ \boldsymbol{Y}_c = \sum_{r=1}^{d_R} \boldsymbol{B}_{r,c} \boldsymbol{X}_r, \quad \forall r = 1, \ldots, d_R, \quad \sum_{c=1}^{d_C} \boldsymbol{B}_{r,c} = 1 \tag{4.3}$$

$$\forall c = 1, \ldots, d_C, \forall r = 1, \ldots, d_R, \quad \boldsymbol{B}_{r,c} \geq 0, \quad \boldsymbol{X}_{r,c} \geq 0, \quad \sum_{r=1}^{d_R} \boldsymbol{X}_r = 1 \quad . \tag{4.4}$$

Without further assumptions, the identified set for $\boldsymbol{m}$, *i.e.*, the set of parameters compatible with the distribution of the data, is related to the one considered in Cross and Manski (2002*a*) and is exactly the one studied in Manski (2018) and Gaillac (2020). Then, I make the nonparametric assumption of independence between the random coefficients and the random margin of the covariate $\boldsymbol{B} \perp \boldsymbol{X}$ - *i.e. no contextual effects* (NCE) in the ecological inference literature. This can be viewed as the exogeneity of the regressor $\boldsymbol{X}$. In the voting example, this exogeneity assumption means that the probability that an individual of a given race go to the polls is independent of the racial composition of the precincts. Estimation under NCE and a degenerated matrix $\boldsymbol{B}$ has been studied in Goodman (1959), then under parametric assumptions and 2×2 tables in King (1997), in Rosen et al. (2001) with d_R categories and d_C choices, using a Bayesian hierarchical model in Wakefield (2004), and Bayesian nonparametric estimation method in Imai et al. (2008).[2] Being frequentist nonparametric is important, as misspecification or the choice of prior could drive the results (see, *e.g.*, Heckman and Singer (1984)).

My main contribution is to obtain new nonparametric constructive point identification of the prediction with two possibilities for the outcome and an arbitrary number of covariate categories. With more than two possibilities, I show how additional nonparametric restrictions on the dimension of the unobserved heterogeneity restore point identification in the linear system of random coefficients equations model. This complements the nonidentification result of Masten (2017). The NCE assumption might not be credible in some contexts. In the example, assume that the researcher does not observe income categories composition of the precincts while the share of Black people being higher in low-income precincts. Then, the NCE assumption fails if *ceteris paribus*, low-income people vote less than the others. More generally, this assumption is restrictive in the sense that it amounts to assume that there is no omitted variable bias. When additional group level variables $\boldsymbol{Z}$ can be observed, such as the precincts compositions in terms of levels of income or education, this allows to assume NCE conditional on these variables, *i.e.*, $\boldsymbol{B} \perp \boldsymbol{X}|\boldsymbol{Z}$, and to perform inference on

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \mapsto \mathbb{E}\left[\boldsymbol{B}|(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\right].$$

This extension have been considered in a parametric way in the literature (see, *e.g.*, King, 1997; Imai et al., 2008) whereas my method is the first to incorporate the additional variables $\boldsymbol{Z}$ nonparametrically. Another extension proposes an alternative to the NCE assumption when the researcher observes an instrument, adapting the control function approach of Masten and Torgovitsky (2013).

I then perform nonparametric estimation of the conditional expectation of the random coefficients satisfying a linear system of equations with two possibilities ($d_C = 2$),

[2]Recent developments use optimal transport as in Frogner and Poggio (2019), with a maximum at posteriori under Dirichlet prior, imposing independence between the rows of the table, and in Muzellec et al. (2017), selecting a solution minimizing a distance, *apriori* reflecting the ideological proximity.

where regressors are bounded - as the covariate margins $\boldsymbol{X}_g$ are vectors of probabilities. Thus, this paper also contributes to the large literature about random coefficients models (see, among many others, Beran and Hall, 1992; Beran and Millar, 1994; Beran et al., 1996a; Hoderlein et al., 2010; Gautier and Kitamura, 2013; Masten, 2017; Hoderlein et al., 2017; Dunker et al., 2017; Newey and Stouli, 2018; Gaillac and Gautier, 2019c,b; Breunig, 2020). It extends Gaillac and Gautier (2019c), which perform estimation of the density of the random coefficients in the difficult case where regressors are bounded, as it considers a functional. Thus, under classical assumptions on the smoothness of the underlying random coefficients distribution, my minimax lower bounds exhibit faster rates of convergence than in Gaillac and Gautier (2019c). My estimator is based on boundary corrected wavelets adapted to the geometry of the support of the regressors. It is optimal in the minimax sense, up to logarithmic factors in the rates of convergence, and adaptive, *i.e.*, its tuning parameter is automatically selected from the data. I also show asymptotic normality. My method performs well, outperforming those making parametric assumptions when the latter are violated, on Monte-Carlo simulations and on the real precincts data about turnout and race mentioned earlier. I provide a R package RobustEI.

This paper concludes showing that a direct plug-in of the above estimator allows the estimation of the effect a treatment T_g on the choice probabilities for categories in a clustered experiment by group g where the individual outcomes C and covariates R are protected. Consider the potential outcome model of Rubin (1974), where we denote by $\boldsymbol{B}_g(0)$ the potential outcome for group g if not treated, $\boldsymbol{B}_g(1)$ the potential outcome if treated, and we observe only the treatment status $T_g \in \{0,1\}$ and relalized outcome $\boldsymbol{B}_g(T_g)$. When the groups g are treated $T_g = 1$ randomly conditionally on the variables $\boldsymbol{W}_g$, *i.e.*, the *unconfoundedness* assumption at the group level g,

$$\boldsymbol{B}_g(1), \boldsymbol{B}_g(0) \perp T_g \quad | \boldsymbol{W}_g,$$

the treatment effect $\gamma_{r,1} = \mathbb{E}\left[\boldsymbol{B}_{r,1}(1) - \boldsymbol{B}_{r,1}(0)\right]$ on the probability to choose $C = 1$ among individuals of category $R = r$, namely $\boldsymbol{B}_{r,1}$, can be expressed as a function of

m,

$$\gamma_{r,1} = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{B}_{r,1}|\boldsymbol{W}, T=1\right] - \mathbb{E}\left[\boldsymbol{B}_{r,1}|\boldsymbol{W}, T=0\right]\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{m}_{r,1}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, T)|\boldsymbol{W}, T=1\right] - \mathbb{E}\left[\boldsymbol{m}_{r,1}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, T)|\boldsymbol{W}, T=0\right]\right], \tag{4.5}$$

where T and $\boldsymbol{W}$ are considered as additional variables $\boldsymbol{Z} := (\boldsymbol{W}, T)$. This paves the way for many applications in Political Science, where this is a common situation.

I consider an application to the effect of door-to-door visits on vote shares among different categories of voters, based on past votes, in the experiment of Pons (2018). The experiment consists in sending campaigners visit households in randomly selected precincts to convince them to vote for the left-wing candidate at the 2012 French Presidential election. The secret ballot prevents from observing directly the joint distribution of the probability of voting for a candidate and individual characteristics in the precincts. Pons (2018) shows that there is an impact at the precinct level on the left candidate's vote shares. He justifies that motivation of left-wing nonvoters is not the main channel. He argues that the main mechanism is the persuasion of undecided active voters. However, due to the data, he can not directly estimate the impact on this category of voters. I provide quantitative evidence for this heterogeneity. The strategy is to leverage the two rounds of the French Presidential election, where only two candidates qualify for the second round. Hence, the group of individuals whose candidate they voted for at the first round of the election did not qualify for the second round proxies the group of undecided active voters. I observe, for a sample of precincts, the second round vote shares for the left-wing candidate and the precincts composition based on past votes and use my estimator of $\boldsymbol{m}$. My main result validates the insights of Pons (2018), quantifying the impact on the left-wing candidate's vote shares among the category of undecided active voters. There is no effect on the vote shares among the others. Finally, I use a second decomposition with respect to the level of education. It shows that the persuasion effect of the visits is stronger on less-educated voters.

The remainder of the paper is organized as follows. Section 4.1 studies identification. Then, Section 4.2 considers inference with $d_C = 2$, with theoretical asymptotic results and Monte-Carlo simulations. Finally, Section 4.3 performs estimation of the treatment effect on the vote shares given categories of voters. Appendix 4.5.1 provides identification when $d_C > 2$ and Appendix 4.5.2 gathers complements and proofs of the main theorems.

4.1 Identification

4.1.1 Identification without assuming no contextual effects

In this section, I assume that the researcher observes a distribution $\mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}}$ across groups. For a random vector $\boldsymbol{X}$, denote by $\mathbb{S}_{\boldsymbol{X}}$ its support.

Let $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$. The set of matrices $\boldsymbol{B} \in \mathcal{M}_{d_R,d_C}([0,1])$ satisfying (4.1) can be characterized, from Cross and Manski (2002*a*) (see Proposition 1 in Gaillac, 2020), as the set of matrices with elements in $[0,1]$ such that (4.3) holds, namely

$$\mathcal{I}(\boldsymbol{x}, \boldsymbol{y}) = \left\{\boldsymbol{B} \in \mathcal{M}_{d_R,d_C}([0,1]),\ \boldsymbol{B}\mathbf{1} = \mathbf{1},\ \boldsymbol{y} = \boldsymbol{B}^\top \boldsymbol{x}\right\}. \tag{4.6}$$

$\mathcal{I}(\boldsymbol{x}, \boldsymbol{y})$ is nonempty,[3] bounded and defined by a finite number of equality and inequality constraints. It is therefore a closed and convex polytope (see, *e.g.*, Brualdi, 2006, p.337). The set of associated joint distributions is known as the transportation polytope (see, *e.g.*, Cuturi, 2012).

Proposition 1 below simply shows that, without restricting the dependence between $\boldsymbol{B}$ and $\boldsymbol{X}$ - even with the knowledge of the distribution of the margins - the prediction $\boldsymbol{m}$ is only partially identified. The best we can assert for given margins $\boldsymbol{x}, \boldsymbol{y}$ is that the prediction $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y})$ belongs the set $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y})$ obtained without structure. This relates identification of the prediction $\boldsymbol{m}$ to the partial identification literature in econometrics following Cross and Manski (2002*a*).

[3] The independent distribution $\boldsymbol{B} = \mathbf{1}\boldsymbol{y}^\top$, which corresponds to the case where choices probabilities do not depend on the covariate, always belongs to $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y})$.

Definition of the identified set for $\boldsymbol{m}$. I explicit here useful elements of nonparametric identification (see, *e.g.*, Matzkin, 2007*b*). The distribution of the observables is $\mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}}$, while the distribution of the observables generated by $\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}}$ and the system (4.1) is $\mathbb{P}^{gen}(\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}})$. $\mathcal{R}$ is a set of restrictions defined accordingly, like satisfying the independence restriction $\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}} = \mathbb{P}_{\boldsymbol{B}} \otimes \mathbb{P}_{\boldsymbol{X}}$. The functional of interest is (4.2) and satisfies $\boldsymbol{m} = \Gamma(\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}}, \mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}})$ for a certain deterministic function Γ.[4] The identified set for $\boldsymbol{m}$ is the set of matrix valued functionals such that there exists a unobserved associated distribution $\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}}$ which generates observations compatible with the distribution of the data,

$$\mathcal{J}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma,\mathcal{R}) := \{\boldsymbol{m}: \ \exists\, \mathbb{P}_{\boldsymbol{B},\boldsymbol{X}} \in \mathcal{R}, \quad \mathbb{P}^{gen}(\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}}) = \mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}}, \quad \Gamma(\mathbb{P}_{\boldsymbol{B},\boldsymbol{X}}, \mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}}) = \boldsymbol{m}\}\,.$$

Proposition 1 (Partial identification without assuming NCE)**.** *Let the distribution of $(\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y})$ satisfy (4.1) and $\mathcal{R}$ be unrestricted. The identified set for $\boldsymbol{m}$ is*

$$\mathcal{J}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma,\mathcal{R}) = \left\{\boldsymbol{m} : [0,1]^{d_R\times d_C} \to \mathcal{M}_{d_R,d_C}([0,1]),\ \forall(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}},\ \boldsymbol{m}(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{I}(\boldsymbol{x},\boldsymbol{y})\right\},$$

where $\mathcal{I}(\boldsymbol{x},\boldsymbol{y})$ is defined in (4.6).

4.1.2 Identification assuming no contextual effects

I now consider the following exogeneity assumption which constrains the dependence between the regressor and the random coefficients.

Assumption 1 ("No contextual effects" (NCE))**.** *Assume that the heterogeneous conditional probabilities are independent of the shares of the different categories across groups, namely:*

$$\boldsymbol{B} \perp \boldsymbol{X}.$$

Assumption 1 is classical both in the random coefficients and in the ecological inference litteratures. This nonparametric assumption is however strong for some

[4] It is detailed in the proofs, using Bayes' theorem and that the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{B},\boldsymbol{X}$ is fixed by (4.1).

applications (see, *e.g.*, Tam Cho, 1998) hence the need to perform sensitivity analysis to the predictions obtained under this assumption. In assumptions 4 or 5, I consider alternative assumptions when other covariates are available. For a vector $\boldsymbol{r}$ of size d, denote by $\underline{\boldsymbol{r}}$ the vector of size $d-1$, containing the first $d-1$ entries of $\boldsymbol{r}$.

Assumption 2. *The support of $\underline{\boldsymbol{X}}$ has nonempty interior.*

I maintain Assumption 2 for simplicity. Note that, because $\underline{\boldsymbol{X}}$ are probabilities, this latter assumption is not restrictive in most applications. Support conditions on the regressors in this context are relaxed in Theorem 5 in Gaillac and Gautier (2019*b*) and we could allow for discrete regressors whose support is countably infinite.

The system (4.1) is a particular type of seemingly unrelated regressions (SUR) with random coefficients which contain a common regressor, with additional constraints $\boldsymbol{X}^\top \mathbf{1} = \boldsymbol{Y}^\top \mathbf{1} = 1$. It is shown in Corollary 1 in Masten (2017) in the context of SUR that the joint distribution of $\boldsymbol{B}$ is necessarily not point identified (see Proposition (P2.a) below). Proposition (P2.b) below is new and shows that, with more than two choices, even the conditional expectation of the random coefficients is not identified without additional assumptions on the random matrix. When $d_C = 2$, because in our model the distribution of $\boldsymbol{B}$ is compactly supported $\mathbb{S}_{\boldsymbol{B}_{\cdot,1}} \subseteq [0,1]^{d_R}$, then Proposition (P2.2) below is Proposition 2.2 in Beran and Millar (1994) and (P2.1) is a direct consequence of it.

Proposition 2 (Identification without contextual effects)**.** *Let the distribution of* $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y})$ *satisfy* (4.1) *and Assumption 1. We have, for all* $d_R \geq 2$*, when* $d_C = 2$*,*

(*P2.1*) $\boldsymbol{m}$ *is identified under Assumption 2;*

(*P2.2*) *the distributions of* $\boldsymbol{B}$ *and of* $\boldsymbol{B}$ *conditional on* $(\boldsymbol{X}, \boldsymbol{Y})$ *are identified under Assumption 2;*

and, when $d_C > 2$*,*

(*P2.a*) *the distribution of* $\boldsymbol{B}$ *is not identified under Assumption 2;*

(P2.b) $\boldsymbol{m}$ *is not identified under Assumption 2.*

In Proposition 2 and under Assumption 2, I use the fact that the support of $\boldsymbol{B}$ is compact, hence the distribution is determined by its moments. Theorem 1 below goes further than the nonidentification result of (P2.b) with partial identification results and also showsn nonparametric constructive point identification in the case $d_C = 2$. The notations $\cdot, \star$ are used to denote a variable in a function. For two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\mathbb{P}_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$, $f_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$, and $F_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$ denote the conditional probability, density, and cumulative distribution.

Assumption 3. *The conditional density* $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ *exists and, for all* $j = 1, \ldots, d_R - 1$ *and* $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$*, its partial derivatives* $\partial_{\boldsymbol{x}_j} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\cdot|\boldsymbol{x})$ *are integrable and square integrable on* $\mathbb{R}^{d_C-1}$.

I give in Section 4.2 sufficient conditions for Assumption 3 in terms of minimal smoothness of the density of $\boldsymbol{B}$. Note that many classical parametric distributions of $\boldsymbol{B}$ yield that Assumption 3 holds, such as the uniform distribution, the truncated normal used by King (1997), the beta or the Dirichlet distributions with parameter strictly greater than one or the logit-normal distribution.

Theorem 1 (Constructive identification when $d_C = 2$)**.** *Let* $d_C = 2$*, the distribution of* $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y})$ *satisfy* (4.1) *and define the restriction* $\mathcal{R}$ *corresponding to assumptions 1, 2, and 3. Then, the prediction* $\boldsymbol{m}$ *is point identified and satisfies, for all* $r = 1, \ldots, d_R$ *and* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$*,*

$$\boldsymbol{m}_{r,1}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}_1 + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \frac{\partial_{\boldsymbol{x}_l} F_{\boldsymbol{Y}_1|\underline{\boldsymbol{X}}}(\boldsymbol{y}_1|\underline{\boldsymbol{x}})}{f_{\boldsymbol{Y}_1|\underline{\boldsymbol{X}}}(\boldsymbol{y}_1|\underline{\boldsymbol{x}})} \tag{4.7}$$

and $\boldsymbol{m}_{r,2} = 1 - \boldsymbol{m}_{r,1}$.

Partial identification and point identification when $d_C > 2$. Proposition 7 in Appendix 4.5.1 studies partial identification with more than two outcomes possibilities, $d_C > 2$, assuming NCE. It shows that the elements of $\boldsymbol{m}$ are solutions of a system of coupled transport partial differential equations. When one limits the dimension of the unobserved heterogeneity, Appendix 4.5.1 then provides a way to solve this system and to restore point identification when $d_C = 3$, which can be extended to $d_C > 3$. Specifically, I assume that some random coefficients are linearly dependent of the others. I show that $\boldsymbol{m}$ can be expressed as a linear combination of the quantities $\partial_{\boldsymbol{x}_l} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$.

4.1.3 Identification with additional variables or instruments

Using conditioning. Additional variables $\boldsymbol{Z}$ of dimension p_Z can be used to allow for some contextual effects, performing the analysis conditional on $\boldsymbol{Z}$.

Assumption 4. $\boldsymbol{B} \perp \boldsymbol{X} | \boldsymbol{Z}$.

Under Assumption 4, the parameter of interest becomes the expectation of $\boldsymbol{B}$ conditional on the observed quantities, *i.e.*, given values of the margins $\boldsymbol{X}, \boldsymbol{Y}$ and the additional variables $\boldsymbol{Z}$:

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}} \mapsto \mathbb{E}\left[\boldsymbol{B} | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = \boldsymbol{z}\right] \in \mathcal{M}_{d_R, d_C}([0,1]). \tag{4.8}$$

Identification under Assumption 4 is the parallel of Theorem 1 (see Proposition 12 in the Appendix) and states that the same type of formula can be obtained for (4.8) in the case $d_C = 2$, simply conditioning on $\boldsymbol{Z}$, for all $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}}$ and $r = 1, \ldots, d_R$,

$$\boldsymbol{m}_{r,1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{y}_1 + \sum_{l=1}^{d_R - 1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \frac{\partial_{\boldsymbol{x}_l} F_{\boldsymbol{Y}_1 | \underline{\boldsymbol{X}}, \boldsymbol{Z}}(\boldsymbol{y}_1 | \underline{\boldsymbol{x}}, \boldsymbol{z})}{f_{\boldsymbol{Y}_1 | \underline{\boldsymbol{X}}, \boldsymbol{Z}}(\boldsymbol{y}_1 | \underline{\boldsymbol{x}}, \boldsymbol{z})} \tag{4.9}$$

and $\boldsymbol{m}_{r,2} = 1 - \boldsymbol{m}_{r,1}$.

Using the control function approach. An alternative is to use the control function approach used in, *e.g.*, Florens et al. (2008); Masten and Torgovitsky (2013); Newey and Stouli (2020), when an instrument $\boldsymbol{W}$ is available.

Assumption 5. *1.* **(First stage equation)** *For each $r = 1, \dots, d_R - 1$, there exists a scalar random variable $\boldsymbol{V}_r$ and a possibly unknown function $\boldsymbol{h}_r$ that is strictly increasing in its second argument, for which $\boldsymbol{X}_r = \boldsymbol{h}_r(\boldsymbol{W}, \boldsymbol{V}_r)$.*

2. **(Instrument exogeneity)** $(\boldsymbol{B}, \boldsymbol{V}) \perp \boldsymbol{W}$.

Assumption 5 is an alternative to the NCE assumption and places restrictions on the dependence between $\boldsymbol{X}$ and $\boldsymbol{B}$. Namely, it implies that most of the correlation between $\boldsymbol{X}$ and $\boldsymbol{B}$ occurs through $\boldsymbol{V}$. This can be structurally motivated in some applications. Define $\boldsymbol{Z}_r := F_{\boldsymbol{X}_r|\boldsymbol{W}}(\boldsymbol{X}_r|\boldsymbol{W})$ for $r = 1, \dots, d_R - 1$. Proposition 1 in Masten and Torgovitsky (2013) ensures that $(\boldsymbol{Z}, \boldsymbol{B}) \perp \boldsymbol{W}$ and that $\boldsymbol{X} \perp \boldsymbol{B}|\boldsymbol{Z}$, which yields identification in Theorem 1 under Assumption 5 rather than Assumption 1.

Proposition 3 (Identification using the control function when $d_C = 2$)**.** *Let the distribution of $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W})$ satisfy the assumptions 2 and 5. Define $\boldsymbol{Z}$ by, for $r = 1, \dots, d_R - 1$, $\boldsymbol{Z}_r := F_{\boldsymbol{X}_r|\boldsymbol{W}}(\boldsymbol{X}_r|\boldsymbol{W})$. Then*

$$\boldsymbol{m} : (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}} \mapsto \mathbb{E}\left[\boldsymbol{B}|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = \boldsymbol{z}\right]$$

satisfies (4.9).

4.2 Inference

This section, for simplicity, considers only the case $d_R \times 2$, where heterogeneity does not have to be restricted further than Assumption 1 to get point identification. My plug-in strategy below can be directly adapted to the case $d_C = 3$, using Proposition 8. When $d_C = 2$, using the constraint $\boldsymbol{Y}^\top \mathbf{1} = 1$, the system (4.1) can be treated using a single equation, without loss of generality. Because of the constraints $\boldsymbol{X}^\top \mathbf{1} = \boldsymbol{Y}^\top \mathbf{1} = 1$, which create redundancies, I use hereafter a slight abuse of notations and, for $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$, use the notation $\boldsymbol{m}(\boldsymbol{x}, y)$ instead of $\boldsymbol{m}((\boldsymbol{x}, 1 - \boldsymbol{x}^\top \mathbf{1})^\top, (y, 1 - y)^\top)$. When $\mathcal{S}$ is measurable, $L^2(\mathcal{S})$ denotes the space of complex-valued square integrable functions equipped with the inner product $\langle f, g \rangle_{L^2(\mathcal{S})} = \int_{\mathcal{S}} f(\boldsymbol{x}) \overline{g}(\boldsymbol{x}) d\boldsymbol{x}$.

Asymptotic analysis with the minimax risk. This section characterizes the asymptotic properties of estimators of $\boldsymbol{m}$ in the minimax context, which I explain here. Based on a sample of margins $(\boldsymbol{X}_g, \boldsymbol{Y}_g)_{g=1}^{G}$ for G groups, I define the expected error of an estimator $\widehat{\boldsymbol{m}}_{r,1}$ of $\boldsymbol{m}_{r,1}$, for $r = 1, \ldots, d_R$,

$$\mathcal{R}^q\left(\widehat{\boldsymbol{m}}_{r,1}, \boldsymbol{m}_{r,1}\right) := \mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}_{r,1} - \boldsymbol{m}_{r,1}\right\|_{L^q(\mathcal{S})}\right]$$

in L^q norm on $\mathcal{S}$, which is a subset of $[0,1]^{d_R}$ defined later in Assumption (Est.3), for $q \in \{2, \infty\}$. First, for my particular estimator $\widehat{\boldsymbol{m}}_{r,1}^{j_0}$, where j_0 is the tuning parameter, I show an upper bound on the maximum risk, which the worst error estimating $\boldsymbol{m}$ associated to a density $f_{\boldsymbol{B}_{\cdot,1}}$ - assuming that it exists - in the space $\mathcal{H}^{s+1}(l)$ defined later, for $r = 1, \ldots, d_R - 1$,

$$\frac{1}{r(G)} \underbrace{\sup_{f_{\boldsymbol{B}_{\cdot,1}} \in \mathcal{H}^{s+1}(l)} \mathcal{R}^q\left(\widehat{\boldsymbol{m}}_{r,1}^{j_0}, \boldsymbol{m}_{r,1}\right)}_{\text{Maximum risk}} = O(1), \tag{4.10}$$

where $r(G)$ is thus a rate of convergence for my estimator. $\mathcal{H}^{s+1}(l)$ characterizes the smoothness of the distributions $f_{\boldsymbol{B}_{\cdot,1}}$ and is indexed by two parameters s and l. Thus, controlling the maximum risk for an estimator shows the uniformity of its performances with respect to all the distributions in the class $\mathcal{H}^{s+1}(l)$. Second, I turn to the question of the optimality of this estimator. The performance measure I consider is the minimax risk, *i.e.*, the minimum of the maximum risk that an estimator $\widehat{\boldsymbol{m}}_{r,1}$ can achieve,

$$\mathcal{R}_G^{q,*} := \inf_{\widehat{\boldsymbol{m}}_{r,1}} \sup_{f_{\boldsymbol{B}_{\cdot,1}} \in \mathcal{H}^{s+1}(l)} \mathcal{R}^q\left(\widehat{\boldsymbol{m}}_{r,1}, \boldsymbol{m}_{r,1}\right). \tag{4.11}$$

I show a lower bound r on the latter which takes the form, for all $r = 1, \ldots, d_R$,

$$\exists \nu > 0 : \ \underline{\lim}_{G \to \infty} \frac{1}{r(G)} \mathcal{R}_G^{q,*} \geq \nu. \tag{4.12}$$

Of course, the aim is to establish the sharpest lower bound possible and to obtain a rate for my estimator in (4.10) that is the closest to the rate that is achievable for this statistical problem in (4.12). Note that, (4.10) also gives an upper bound on the minimax risk (4.11), as I consider a specific estimator. Theorem 2 below shows that my estimator achieves the best rate when $q = 2$ and that it is close to it, up to logarithmic factors, for $q = \infty$. However, the tuning parameter j_0 has to be chosen as a function of the smoothness parameter s, which is unobserved. Hence the last step is to choose the tuning parameter $\widehat{j}_0$, using only the data, while keeping a rate close to the case where the smoothness parameter is known. Namely, I show that my estimator is adaptive proving

$$\frac{1}{r(G)} \sup_{f_{\boldsymbol{B}_{\cdot,1}} \in \mathcal{H}^{s+1}(l)} \mathcal{R}^2\left(\widehat{\boldsymbol{m}}_{r,1}^{\widehat{j}_0}, \boldsymbol{m}_{r,1}\right) = O(1), \tag{4.13}$$

where the rate r is the one in (4.10) up to a logarithmic term. Table (4.1) below presents a summary of the rates obtained with my estimator in L^2 norm.

Smoothness and sampling assumptions. For $d \geq 1$, denote the Fourier transform of $f \in L^1\left(\mathbb{R}^d\right) \cap L^2\left(\mathbb{R}^d\right)$ by $\mathcal{F}[f](\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{b}^\top \boldsymbol{x}} f(\boldsymbol{b}) d\boldsymbol{b}$.

Assumption 6 (Smoothness assumption, Sobolev ellipsoid)**.** *Let $l \in (0, \infty)$, $s > (d_R - 1)/2$, and assume that $f_{\boldsymbol{B}_{\cdot,1}}$ exists and belongs to*

$$\mathcal{H}^s(l) := \left\{ f_{\boldsymbol{B}} : \quad \int_{\mathbb{R}^{d_R}} \left(1 \vee |\boldsymbol{\xi}|_2\right)^{2s} \left|\mathcal{F}[f_{\boldsymbol{B}_{\cdot,1}}](\boldsymbol{\xi})\right|^2 d\boldsymbol{\xi} \leq l^2 \right\}.$$

The key proposition that links this Sobolev type smoothness to the model is Proposition 9, which is of independent interest. Note that, contrary to Assumption 3, the uniform distribution or the truncated normal used by King (1997), do not satisfy Assumption 6. This is due to the discontinuity at the boundary of the support. Hence smooth approximations of the uniform distribution or the truncated normal at the boundary satisfy Assumption 6. More importantly, the beta and the Dirichlet distributions with parameter strictly greater than one, or the logit-normal distribution, which are usual parametric distributions to represent probabilities (see, *e.g.*, Katz and King, 1999; Imai et al., 2008), satisfy Assumption 6.

The following assumptions are introduced to be able to derive convergence rates.

Assumption 7. *Assume that:*

($Est.1$) *there exist densities* $f_{\underline{\boldsymbol{X}}}$ *and* $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$*, which are considered known for simplicity in the body of this paper and estimated under Assumption 12 in the Appendix;*

($Est.2$) *we observe an i.i.d sample* $(\boldsymbol{X}_g, \boldsymbol{Y}_g)_{g=1}^G$*;*

($Est.3$) *For* $c_{\boldsymbol{X}}, c_{\boldsymbol{X},\boldsymbol{Y}} \in (0,\infty)$, $\|1/f_{\underline{\boldsymbol{X}}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq c_{\boldsymbol{X}}$, $\|f_{\underline{\boldsymbol{X}}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq C_{\boldsymbol{X}}$*, and there exists a strict subset* $\mathcal{S}$ *of* $\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$ *such that* $\|1/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\|_{L^\infty(\mathcal{S})} \leq c_{\boldsymbol{X},\boldsymbol{Y}}$.

Table 4.1: Summary of the minimax $L^2(\mathcal{S})$ risk rates of convergence in $\mathcal{H}^{s+1}(l)$

	Lower bound, "best" est.	Est. (4.15), s known	Est. (4.15)-(4.18), data-driven
Rate, $r(G)$	$G^{-\frac{2s}{2s+d_R+1}}$	$\forall N,\ G^{-\frac{2s_N}{2s_N+d_R+1}}$ $s_N \to_{N\to\infty} s$	$\forall N,\ \left(\frac{G}{\ln(G)}\right)^{-\frac{2s_N}{2s_N+d_R+1}}$ $s_N \to_{N\to\infty} s$
Statement	(4.12), Proposition 4	(4.10), Theorem 2	(4.13), Proposition 5

Notes: The asymptotic is in the number of groups G. est. means estimator. N is a arbitrarily fixed parameter indexing the smoothness of the wavelets functions used my estimator (4.15).

4.2.1 Upper bounds

Assumption 8 (Assumption on the support of the covariate margins)**.** *Assume that either*

($RC.1$) $\mathbb{S}_{\underline{\boldsymbol{X}}} = \prod_{l=1}^{d_R-1}[\widetilde{\boldsymbol{x}}_l, \widetilde{\boldsymbol{x}}_l + x_0]$*, where* $\widetilde{\boldsymbol{x}} \in [0,1]^{d_R-1}$ *and* $x_0 > 0$.

($RC.2$) *or* $\mathbb{S}_{\underline{\boldsymbol{X}}} = \left\{\boldsymbol{x} \in [0,1]^{d_R-1};\ \sum_{l=1}^{d_R-1} \boldsymbol{x}_l \leq 1\right\}$ *which is a triangle;*

Assumption (RC.2) is more natural as we work with shares, but more difficult to handle than (RC.1). The proof of Theorem 1 is constructive and my estimator is based on a plug-in approach of an estimator of $\left(\partial_{\boldsymbol{x}_l} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)_{l=1}^{d_R-1}$. The strategy implies having first-step estimators of $f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}$ and $f_{\underline{\mathbf{X}}}$ as described in Appendix 4.5.2. As these are quantities more classically estimated in the statistical literature and for the sake of simplicity, I make high level assumptions on their rates of estimation and refer to, *e.g.*, Giné and Nickl (2016), for estimators based on wavelets. In the body of the paper here I assume that $f_{\underline{\mathbf{X}}}$ and $f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}$ are known for simplicity, but all results are proved in Appendix in the more general case where they are estimated. They are implemented in a data-driven way in the R package RobustEI using Legendre polynomials or wavelets.

My estimator uses the vaguelet-wavelet decomposition that I now describe. Appendix 4.5.4 also provides a simpler yet non optimal estimator based on Legendre polynomials. In practice, the latter has good finite sample properties hence it is used in our application in Section 4.3.

The vaguelet-wavelet decomposition. Let $J, N \in \mathbb{N}$, $2^J \geq 2N$. The estimator uses wavelets based on the Daubechies scaling function (see, *e.g.*, Daubechies, 1992) corrected at the boundaries according to the support of $\underline{\mathbf{X}}$, which constitute an orthonormal basis of $L^2(\mathbb{S}_{\underline{\mathbf{X}}})$,

$$\mathcal{B} = \left\{ \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}, j \in \mathbb{N} \setminus \{0, \dots, J-1\},\ \boldsymbol{k} \in \Lambda_j,\ \boldsymbol{w} \in \mathcal{W}_j \right\},$$

where Λ_j is a set of cardinal $O\left(2^{j(d_R-1)}\right)$ (see Appendix 4.5.2 for the precise definition according to the geometry fo the support of $\underline{\mathbf{X}}$), $\widetilde{\mathcal{W}}$ is the set of $2^{d_R-1}-1$ sequences $\boldsymbol{w}$ of zeros and ones, excluding the $d_R - 1$ dimensional vector of zeros $\mathbf{0}$, for $j \neq J$, $\mathcal{W}_j = \widetilde{\mathcal{W}}$, $\mathcal{W}_J = \mathbf{0}$, and $\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}$ are based on dilatations and translations of ϕ and ψ,[5] which are the initial scaling and wavelet functions from the Daubechies family with $N \geq 1$ (see, *e.g.*, Härdle et al., 2012) followed by a correction at the boundary, adapted to the support of $\underline{\mathbf{X}}$. Appendix 4.5.2 presents the different types of boundary corrections adapted to Assumption 8, where I use the boundary corrected wavelets

[5] $\Psi^{\mathbf{0}}_{j,\boldsymbol{k}} = \Phi_{J,\boldsymbol{k}}$, $\boldsymbol{k} \in \Lambda_J$ which corresponds to dilatation of scaling function ϕ only, up to the boundary correction. See Appendix 4.5.2 for a more precise description of these functions.

introduced in Cohen et al. (1993) and adapt the correction introduced in Ajmi et al. (2011).

Estimator. To deal with the statistical problem, I use

$$\widehat{\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}^{j_0}(\star|\cdot) := \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\mathcal{W}_j} \widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(\star) 2^j \Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\cdot),$$

where $\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}} := \partial_l \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}/2^j$ are called the vaguelets and, for all $y \in [0,1]$,

$$\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) := \frac{1}{G}\sum_{g=1}^{G} \frac{\mathbb{1}\{\underline{\boldsymbol{Y}}_g \le y\}}{f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)} \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\left(\underline{\boldsymbol{X}}_g\right) \tag{4.14}$$

is an estimator of $d_{j,\boldsymbol{k},\boldsymbol{w}}(y) := \left\langle \mathbb{E}\left[\mathbb{1}\{\underline{\boldsymbol{Y}} \le y\}|\underline{\boldsymbol{X}} = \cdot\right], \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}} \right\rangle_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}$. This yields my estimator of $\boldsymbol{m}_{r,1}$, for $r = 1, \dots, d_R$ and $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$,

$$\widehat{\boldsymbol{m}}^{j_0}_{r,1}(\boldsymbol{x}, y) = y + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \frac{\widehat{\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}^{j_0}(y|\boldsymbol{x})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}. \tag{4.15}$$

Theorem 2 (L^∞ and L^2 convergence rates)**.** *Let $d_C = 2$, $s \ge (d_R - 3)/2$, $l > 0$, $N \in \mathbb{N}$, $N \ge 1 + (d_R + 1)/s$, $J \in \mathbb{N}$ such that $2^J \ge 2N$, $s_N := s(\widetilde{N} + (d_R + 1)(\widetilde{N} - 1)/(2s)) \to_{N\to\infty} s$, $\widetilde{N} = (2N + d_R - 1)/(2N + 2(s+1) + d_R) \to_{N\to\infty} 1$, $j_0 = \lfloor \widetilde{j} \rfloor$, where $\widetilde{j}$ is solution of $2^{\widetilde{j}(2s_N + d_R + 1)} \widetilde{j}^{3\mathbb{1}\{q=\infty\}} = G_e$. Make assumptions 1-3, 7 and 8, then*

(T2.1) (4.10) holds with $q = \infty$ and $r(G) = (G/\ln(G)^3)^{-s_N/(2s_N + d_R + 1)}$;

(T2.2) (4.10) holds with $q = 2$ and $r(G) = G^{-s_N/(2s_N + d_R + 1)}$.

We choose the Daubechies wavelets because they are compactly supported and have Fourier transform decaying at least polynomially at infinity, with index limited by N.[6] This yields that s is replaced by s_N, which can be taken arbitrarily close to s as $N \to \infty$, in the rates of Theorem 2. Note that the condition (4.64) on the preliminary estimators of $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$, based on preliminary samples, requires a

[6] I also use Legendre functions, see Appendix 4.5.4, and could also use Prolate Spheroidal Wave functions (see, *e.g.*, Osipov et al. (2013)), but their oscillations at the boundaries of [-1,1] prevents from obtaining optimal rates, see Proposition 13.

larger independent sample not to impact the convergence rates in Theorem 2 as the dimension of $\underline{\boldsymbol{X}}$ increases.

4.2.2 Lower bounds

Proposition 4 (Minimax lower bounds)**.** *Make assumptions 1-3. For* $0 < l < \infty$*,* $d_C = 2$*,* $s \geq (d_R - 3)/2$*,* $\|f_{\underline{\boldsymbol{X}}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq C_{\boldsymbol{X}} < \infty$*, and assume that* $\mathbb{S}_{\underline{\boldsymbol{X}}_r}$*,* $r = 1, \dots, d_R - 1$*,* $\mathbb{S}_{\underline{\boldsymbol{Y}}}$ *are strict subsets of* $[0,1]$*. Then,*

(P4.a) when $q = \infty$*, (4.12) holds with* $r(G) = (G/\ln(G))^{-s/(2s+d_R+1)}$*;*

(P4.b) when $q = 2$ *and* $d_R = 2$*, (4.12) holds with* $r(G) = G^{-s/(2s+d_R+1)}$*.*

I compare two related inverse problems where regressors have limited variation, in the model (4.3)-(4.4) with $d_C = 2$ and $d_R = 2$:

1. estimation of the density $f_{\boldsymbol{B}_{\cdot,1}}$,
2. estimation of the functional $\boldsymbol{m}$.

Because the equations in (4.3)-(4.4) are linked, I consider estimation of $\boldsymbol{m}_{\cdot,1}$ only. Estimation of the density $f_{\boldsymbol{B}_{1,1},\boldsymbol{B}_{2,1}}$ when regressors have limited variation is an inverse problem addressed in Gaillac and Gautier (2019*c*). There, we decompose the problem using the truncated Fourier operator $\mathcal{F}_c : L^2\left(W_{[-1,1]}\right) \to L^2([-1,1])$, where $W_{[-1,1]} = \mathbb{1}\{[-1,1]\} + \infty\ \mathbb{1}\{[-1,1]^c\}$ and $L^2\left(W_{[-1,1]}\right) = \left\{f \in L^2\left(\mathbb{R}^d\right) :\ \mathrm{supp}(f) \subseteq [-1,1]\right\}$, $\mathcal{F}_c[f] = \mathcal{F}[f](c\ \cdot)$ and show that, for all $t \in \mathbb{R}$, in $L^2([-1,1])$,

$$\mathcal{F}_{tx_0}\left[\mathcal{F}_{1\mathrm{st}}\left[f_{\boldsymbol{B}_{2,1},\boldsymbol{B}_{1,1}-\boldsymbol{B}_{2,1}}\right](t,\cdot_2)\right](\star) = \mathbb{E}\left[e^{it\boldsymbol{Y}_1}|\underline{\boldsymbol{X}} = x_0\star\right],$$

where $\mathcal{F}_{1\mathrm{st}}$ denotes the Fourier transform with respect to the first variable. We show that $\mathcal{F}_c$ admits a singular value decomposition, and that the singular values decay sub-exponentially with k as $e^{-2k\ln(7e\pi(k+1)/c)}$ (see, *e.g.*, Lemma B.5. in Gaillac and Gautier, 2019*c*). Thus, this is a severely ill-posed problem and lower bounds for the L^2 risk in Theorem 1 in Gaillac and Gautier (2019*c*) give $(\ln(G)/\ln_2(G))^{-s}$, where s is a Sobolev type regularity (see, *e.g.*, Appendix B.5. in Gaillac and Gautier, 2019*c*).

A plug-in approach of this estimate of the density to estimate $\boldsymbol{m}$ leads to slower rates than my direct approach.

Estimation of the functional $\boldsymbol{m}$ is a simpler problem, hence achieves faster rates. Minimax convergence rates for the L^∞ risk in nonparametric estimation of the l-th derivative of a regression function with d dimensional covariates assuming that it belongs to a classical Sobolev space indexed by s are $(G/\ln(G))^{-s/(2s+d+2l)}$ (see, *e.g.*, Theorem 6.3.7 in Giné and Nickl, 2016). Our setting, where the dependent variable $\boldsymbol{B}$ in the nonparametric regression is not observed and the dimension of the covariates are $d_R - 1$ thus compares to the case $d = d_R - 1$ (which is the dimension of $\boldsymbol{X}$) and $l = 1$. This can be seen through, for all $y \in \mathbb{S}_{\boldsymbol{Y}_1}$ and in $L^2(\mathbb{S}_{\boldsymbol{X}_1})$,

$$(\boldsymbol{m}_{r,1}(\cdot, y) - y) f_{\boldsymbol{Y}_1|\boldsymbol{X}_1}(y|\cdot) = (\cdot - \mathbb{1}\{r = 1\}) \partial_x F_{\boldsymbol{Y}_1|\boldsymbol{X}_1}(y|\cdot), \tag{4.16}$$

(using Theorem 1 when $d_R = 2$), which relates $\boldsymbol{m}$ to the first derivative of a nonparametric regression function. This shows that this problem is as difficult as estimating the right-hand-side, *i.e.*, the first derivative of a nonparametric regression function on a bounded interval. To fully understand the rates, we have to relate with the smoothness Assumption 6. The difficulty arises from the natural smoothness condition 6, which is linked to model (4.3)-(4.4) and requires the Fourier transform of the compactly supported distribution $\mathbb{P}_{\boldsymbol{B}_{\cdot,1}}$ to decay fast at infinity. However, in relation to the Heisenberg principle (see, *e.g.*, Theorem 2.6 in Mallat, 1999), a function can not be arbitrarily localized in the initial domain and in the Fourier domain.

4.2.3 Data-driven estimation

Data-driven estimation adapts the Goldenshluger-Lepski method (see, *e.g.*, Goldenshluger and Lepski, 2014; Lacour and Massart, 2016) to choose j_0 automatically from the data, while losing only logarithmic terms in the convergence rates. I focus on adaptation with the L^2 risk.[7] Let $p_G := \theta \ln(G)$, $\theta > 6$ and, for all $j_0 \in \mathbb{N}^{\mathbb{R}}$,

[7]The case L^∞ can be treated similarly, see, *e.g.*, Theorem 8.2.8 in Giné and Nickl (2016).

$j_{\max} = \lfloor \widetilde{j} \rfloor$, where $\widetilde{j}$ is solution of $2^{\widetilde{j}} = G^{1/(d_R+1)}$. Let also

$$\beta\left(y, j_0\right) := \max_{j_0+1 \le j' \le j_{\max}} \left(\sum_{j=j_0+1}^{j'} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} 2^{2j} \left| \widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) \right|^2 - \Sigma\left(j'\right) \right)_+,$$

$$\Sigma\left(j_0\right) := \frac{24 \overline{A} \widetilde{A} c_{\boldsymbol{X}} |\widetilde{\mathcal{W}}|}{1 - 2^{-(d_R+1)}} \frac{(1 + 2p_G) 2^{j_0(d_R+1)}}{G}, \tag{4.17}$$

where $\overline{A}$ and $\widetilde{A}$ are respectively the maximum over $l = 1, \dots, d_R - 1$ of the upper Riesz constants A_l and the inverse of the lower Riesz constants a_l of the system $\left(\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}\right)$, see (4.66). Take $\widehat{j}_0$ as

$$\forall y \in \mathbb{S}_{\underline{\boldsymbol{Y}}}, \quad \widehat{j}_0(y) \in \underset{J \le j \le j_{\max}}{\operatorname{argmin}} \left(\beta(y, j) + \Sigma(j)\right) \tag{4.18}$$

hence as the minimizer of an objective that replicates the bias variance trade-off we face when the parameter is deterministic (see, *e.g.*, Gaillac and Gautier, 2019*c*; Lacour and Massart, 2016, for more intuition). Note that Σ does not depend on y because it is not required to obtain Proposition 5 but with the knowledge of $k(y) := \sup_{\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}} \mathbb{E}\left[\mathbb{1}\{\underline{\boldsymbol{Y}} \le y\} | \underline{\boldsymbol{X}} = \boldsymbol{x}\right]$, then one could use $\Sigma(j, y) := \Sigma(j) k(y)$ instead of $\Sigma(j)$ in (4.18), with the same asymptotic properties but maybe better finite sample ones.

Proposition 5 (Data-driven convergence rates for the L^2 risk)**.** *Let $d_C = 2$, $l > 0$, $N \in \mathbb{N}$. Make assumptions 1-3, 8 and 7, then, for $r = 1, \dots, d_R - 1$,* (4.13) *holds with $r(G) = (G/\ln(G))^{-s_N/(2s_N+d_R+1)}$, where s_N is defined in Proposition 4.*

Proposition 5 shows that choosing adaptively the parameter j_0 only yields a logarithmic penalty in the convergence rates compared to the optimal choice in (T1.1b). Note that an alternative would be to use wavelet thresholding (see, *e.g.*, Donoho and Johnstone, 1995; Donoho et al., 1995) to handle the selection of the parameter. This would yield a nonlinear estimator, different from the one studied in Section 4.2.4.

4.2.4 Asymptotic normality

Under Assumption 9 below on the parameter choice and the convergence rates of the preliminary estimators, I obtain asymptotic normality for the prediction $\boldsymbol{m}_{r,1}(\boldsymbol{x}, y)$

for given value of the margins.

Assumption 9. *Let (Asn.1)* $2^{j_0 d_R} j_0/\sqrt{G} \longrightarrow 0$*; (Asn.2)* $G/2^{j_0(2s_N+d_R+1)} \underset{G\to\infty.}{\longrightarrow} 0$*; (Asn.3)* $\forall \boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$, $\underline{c}_{\boldsymbol{X},\boldsymbol{Y}} := \inf_{y\in\mathbb{S}_{\underline{\boldsymbol{Y}}}} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) > 0$.

Let $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$. We have, when $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{X}}}$ are known,

$$\widehat{\boldsymbol{m}}_{r,1}^{j_0}(\boldsymbol{x}, y) - y = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x}, y), \tag{4.19}$$

where

$$\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x}, y) := \sum_{l=1}^{d_R-1} \frac{(\boldsymbol{x}_l - \mathbb{1}\{l = r\})\mathbb{1}\{\underline{\boldsymbol{Y}}_g \le y\}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)} \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\mathcal{W}_j} \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\underline{\boldsymbol{X}}_g\right) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x}).$$

Proposition 6 (Asymptotic normality)**.** *Let* $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$*,* $d_C = 2$*,* $s \ge (d_R - 3)/2$*. Let* $r = 1, \dots, d_R - 1$*,* $j_0 \in \mathbb{N}$*, and* $\boldsymbol{v}_r^{j_0}(\boldsymbol{x}, y) := Var\left(\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x}, y)\right)$*. Make assumptions 1-3, 7, 6, and 9, then we have,*

$$\sqrt{\frac{G}{\boldsymbol{v}_r^{j_0}(\boldsymbol{x}, y)}}\left(\widehat{\boldsymbol{m}}_{r,1}^{j_0}(\boldsymbol{x}, y) - \boldsymbol{m}_{r,1}(\boldsymbol{x}, y)\right) \underset{G\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, 1).$$

Proposition 6 shows that the impact of estimating $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{X}}}$ under Assumption 7 is negligible. Note that condition (Asn.1) is satisfied by the optimal choices of parameters in Theorem 2 when $s > (d_R - 1)/2$. However, as usual in the literature, Proposition 6 does not apply to data-driven selected parameters as in Section 4.2.3, as these are random quantities. Variance computation and confidence intervals are integrated to the R package RobustEI.

4.2.5 Monte-Carlo simulations and real data validation

Monte-Carlo simulations without contextual effects

Consider the case $d_R = 2$, $d_C = 2$, and two setups where the NCE holds and the conditional expectation $\boldsymbol{m}$ can be computed.

First, consider a data generating process (DGP) satisfying King (1997) assumptions. Thus, take $\boldsymbol{B}_{\cdot,1}$ following a truncated normal distribution in $[0,1]^2$, which I denote by $\mathcal{N}_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $X \sim \mathcal{N}_T(0.65, 0.07)$. Take $\boldsymbol{\mu} = (0.1, 0.9)^\top$, $\boldsymbol{\Sigma}_{11} = 0.4$, $\boldsymbol{\Sigma}_{2,2} = 0.1$, and $\boldsymbol{\Sigma}_{2,1} = 0.1$. I compare estimators to 1) the true value of $\boldsymbol{m}$ or 2) the value of $\boldsymbol{B}_g$ (it is the usual reference in the ecological inference literature). The first part of Table 4.2 shows the results. It compares my estimator and the parametric estimator of King (1997), which is in a favorable case.[8] As expected in this setting, the parametric estimator performs better as its assumptions are satisfied. However, my estimator converges well in that case, although at a nonparametric rate.

Second, consider a DGP where the assumptions of the parametric model of King (1997) are not satisfied. Specifically, I take $\boldsymbol{B}$ distributed as a logit mixture of normal distributions with mixing probability $(0.6, 0.4)$. The first distribution is normal with mean $(-1.4, 1.4)$, variance $(0.1, 0.1)$, and covariance 0. I take $X^* \sim \mathcal{N}(0.8, 0.001)$, which yields a very peaked distribution and then apply the logit inverse transformation to obtain X with values in $[0,1]$. This setting is close to Imai et al. (2008), Simulation II. Results are presented in the second part of Table 4.2. In this case, the nonparametric estimator outperforms the parametric one, which does not even converge.[9]

In both cases my estimator uses preliminary adaptive estimators of $f_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{X}}}$ based on Legendre polynomials and which are implemented in my package RobustEI.[10] As standard in the literature (see, *e.g.*, Comte et al., 2013; Dion, 2014), the multiplicative constant appearing in (4.17) is in practice calibrated from a simulation study.

[8] I wanted also to include the nonparametric Bayesian estimator of Imai et al. (2008), but the associated R package **eco** is no longer available and validated on CRAN and the archives codes crash when the Gibbs sampling starts.

[9] The table comparing $\mathbb{E}[\|\widehat{m}_{r,1} - m_{r,1}\|^2_{L^2([0.05,0.95]^2)}]$ and $\mathbb{E}[\|\widehat{m}_{r,1} - m_{r,1}\|^2_{L^1([0.05,0.95]^2)}]$ instead of sample weigthed errors for the different estimators gives similar conclusions and is available upon request.

[10] Note that, similarly to Li and Racine (2008), one could also use a smoothed version of the indicator function $\mathbb{1}\{\underline{\boldsymbol{Y}} \leq y\}$ in the estimator, namely $\Phi((y - \underline{\boldsymbol{Y}})/h_M)$, where Φ is the cumulative distribution function of the standard normal and h_M which goes to zero with M such that it does not impact the rates of convergence. I observe that it has better finite sample properties in some cases.

Table 4.2: In-sample errors without contextual effects

With truncated normal (*i.e.*, corresponding to King (1997) parametric assumptions)

	l^1 error				l^2 error					
	$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		Comp. time	
Sample size	1000	5000	1000	5000	1000	5000	1000	5000	1000	5000
Regression	0.175	0.175	0.088	0.087	0.218	0.217	0.125	0.125	0.001	0.003
King (1997)	**0.014**	**0.011**	**0.013**	**0.010**	**0.020**	**0.016**	**0.016**	**0.013**	36.1	6.3
Non. para.	0.060	0.052	0.067	0.059	0.086	0.075	0.109	0.095	1.7	2.7

With logit-mixture of normals

	l^1 error				l^2 error					
	$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		Comp. time	
Sample size	1000	5000	1000	5000	1000	5000	1000	5000	1000	5000
Regression	0.078	0.067	0.211	0.165	0.097	0.087	0.250	0.197	0.002	0.002
King (1997)	0.049	0.053	0.194	0.212	0.057	0.061	0.231	0.243	26.3	6.4
Non. para.	**0.032**	**0.021**	**0.110**	**0.060**	**0.052**	**0.050**	**0.118**	**0.079**	3.2	7.9

Notes: in this 2×2 case, the in-sampled l^1 error is computed as $\sum_{g=1}^{G} |\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g) - \mathbb{E}[\boldsymbol{B}_{r,1}|\boldsymbol{X} = \boldsymbol{X}_g, \boldsymbol{Y} = \boldsymbol{Y}_g]|/G$ and the l^2 error as $(\sum_{g=1}^{G}(\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g) - \mathbb{E}[\boldsymbol{B}_{r,1}|\boldsymbol{X} = \boldsymbol{X}_g, \boldsymbol{Y} = \boldsymbol{Y}_g])^2/G)^{1/2}$, where $\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g)$ are the different estimators. “Comp. time" refers to computational time for estimation for one simulation. I use the implementation of King (1997) in the R package EiCompare. The Monte-Carlo experiment uses 300 simulations.

Monte-Carlo simulations with contextual effects

Consider a DGP where $\boldsymbol{B}$ and $\boldsymbol{X}$ are related through an additional variable Z. Specifically, I use a DGP close to the one in Imai et al. (2008). Namely, I consider $Z^* \sim \mathcal{N}(-0.85, 0.5)$ then

$$\boldsymbol{B}^*_{\cdot,1} = 0.85 \left(\begin{smallmatrix} Z^* \\ -Z^* \end{smallmatrix} \right) + \epsilon_1, \ \epsilon_1 \sim \mathcal{N}\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} \right), \left(\begin{smallmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{smallmatrix} \right) \right)$$
$$X^* = 0.5(Z^*)^2 + \epsilon_2, \ \epsilon_2 \sim \mathcal{N}(0, 0.5).$$

This yields the observed data $(\boldsymbol{X}, \boldsymbol{Y}, Z)$ using the inverse-logit transformation, where $\boldsymbol{B}_{1,1}$ and $\boldsymbol{X}_1$ have a sample correlation of -0.34. Results are displayed in Table 4.3. Without Z, both the parametric and the nonparametric estimators still achieve good performances, but only the nonparametric one seems to improve as the sample size augments. The nonparametric estimators with Z achieve smaller errors, in particular with large sample sizes. In this 2×2 example, the computational times seem comparable.

Table 4.3: In-sample errors with contextual effects

	l^1 error				l^2 error					
	$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		$\boldsymbol{B}_{1,1}$		$\boldsymbol{B}_{2,1}$		Comp. time	
	1000	5000	1000	5000	1000	5000	1000	5000	1000	5000
					Without Z					
Regression	0.144	0.145	0.130	0.131	0.182	0.184	0.167	0.167	0.004	0.004
King (1997)	0.072	0.070	**0.100**	0.100	0.107	0.107	**0.130**	0.129	23.64	4.67
Non. para.	0.084	0.062	0.137	0.105	0.109	0.083	0.184	0.139	9.47	10.23
					With Z					
King (1997)	0.071	0.071	**0.100**	0.100	0.108	0.107	**0.130**	0.130	6.79	21.19
Non. para.	**0.064**	**0.053**	0.109	**0.089**	**0.081**	**0.076**	0.137	**0.122**	11.76	23.3

Notes: in this 2×2 case, the in-sampled l^1 error is computed as $\sum_{g=1}^{G} |\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g) - \mathbb{E}\left[\boldsymbol{B}_{r,1} | \boldsymbol{X} = \boldsymbol{X}_g, \boldsymbol{Y} = \boldsymbol{Y}_g, Z = Z_g\right]|/G$ and the l^2 error as $(\sum_{g=1}^{G} (\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g) - \mathbb{E}\left[\boldsymbol{B}_{r,1} | \boldsymbol{X} = \boldsymbol{X}_g, \boldsymbol{Y} = \boldsymbol{Y}_g, Z = Z_g\right])^2/G)^{1/2}$, where $\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g)$ are the different estimators. See the Appendix for non-sampled results and comparison to the true value of $\boldsymbol{B}$. "Comp. time" refers to computational time for estimation for one simulation. I use the implementation of King (1997) in the R package EiCompare. The Monte-Carlo experiment uses 300 simulations.

Comparison with ground truth in 3×2 case: turnout by race

This section illustrates the approach on an example where we know the true value of the contingency tables for each precinct.[11] It uses the dataset studied in Imai and Khanna (2016), which includes approximately ten million of individual records from L2, a leading nonpartisan firm which supplies voter data. The file makes available, among others, the race of the individuals, their precinct, and turnout history. I aggregate data about race and turnout at the precinct level which allows to compare to the truth. I continue with the same illustration as in the introduction, based on the binary decision to vote or not at the 2008 general election and on three categories for race: White people, Black people, and others. Thus, we have $d_C = 2$ and $d_R = 3$.

[11] Numerous such datasets have been released to evaluate the performance of ecological inference methods in real situations (see, *e.g.*, Jiang et al., 2020).

In this example, I view the 2008 election as a sample from $\mathbb{P}_{\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y}}$, with heterogeneity accross precincts. I run the analysis at the precinct level (8,843 observations). Table 4.4 compares my estimator and the estimator of Rosen et al. (2001). More precisely, it considers three types of estimators 1) assuming NCE (Assumption 1); 2) using Assumption 4 conditioning on Z_1 which is the share of individuals registered as Democrat at the district level 3) same as 2) but conditioning on Z_2 the share of individuals registered as Other. In case 2), we thus control for local activism that could create aggregation effects at the precinct level. Note that, as we consider a national election, aggregation bias coming from local stakes is less probable.[12] Table 4.4 presents the results First, note that in this example computational time is really reduced using our nonparametric estimator rather than methods based on simulations. Second, there does not appear to be much difference in this case between the estimators using the additional variable Z or not. Third, my nonparametric estimator achieves the best performances in prediction.

Table 4.4: In-sample errors in turnout by race in Florida

	MAE			RMSE			Time (s.)
	$\boldsymbol{B}_{1,1}$	$\boldsymbol{B}_{2,1}$	$\boldsymbol{B}_{3,1}$	$\boldsymbol{B}_{1,1}$	$\boldsymbol{B}_{2,1}$	$\boldsymbol{B}_{3,1}$	
Rosen et al. (2001), without controls	0.048	0.131	0.096	0.099	0.174	0.133	>3600
Rosen et al. (2001), with Z_1	0.043	0.143	0.097	0.092	0.188	0.141	>3600
Rosen et al. (2001), with Z_2	0.057	0.193	0.090	0.122	0.231	0.136	>3600
Non. para., without controls	0.026	**0.097**	0.095	0.056	**0.153**	0.144	12.0
Non. para., with Z_1	**0.023**	0.108	**0.086**	**0.054**	0.165	**0.127**	13.3
Non. para., with Z_2	0.027	0.105	0.119	0.056	0.167	0.192	41.6

Notes: in this 3×2 case, the in-sampled MAE is computed as $\sum_{g=1}^{G} |\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g) - \boldsymbol{B}_{r,1,g}|/G$ and the $RMSE$ as $(\sum_{g=1}^{G}(\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g) - \boldsymbol{B}_{r,1,g})^2/G)^{1/2}$, where $\widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, Z_g)$ are the different estimators. $\boldsymbol{B}_{1,1}$ is probability to vote conditional on being White, $\boldsymbol{B}_{2,1}$ is probability to vote conditional on being Black, and $\boldsymbol{B}_{3,1}$ is probability to vote conditional on being neither White nor Black. I use as Z_1 the share of individuals registered as democrats in the precinct and Z_2 the share of individuals whose party is neither Democrat nor Republican. "Comp. time" refers to computational time for estimation for one simulation. I use the implementation of King (1997) in the R package EiCompare.

[12] However, due to the join vote at the House and for the Presidential election, individuals could decide to participate based on unobservable district level stakes for the House election.

4.3 Application: estimation of the treatment effect on the vote shares given categories of voters

4.3.1 Motivation and context

The context of experiments where either the outcome is protected or a covariate of interest might only be available for several groups - for example villages or classes - in another dataset that can not be merged is very common. For example, one experiment might be targeted to evaluate the impact on some specific outcome like income, but might generate some side effects on electoral behavior or health outcomes not included in the survey - for example in the PROGRESA experiment (see, *e.g.*, Imai et al., 2020). I consider here a particular large-scale experiment where the outcome and the covariate of interest are both protected, and where ecological inference helps describing and quantifying the different mechanisms.

The importance of interpersonal discussions in the political decision process and the persuasion of voters has been largely documented (see, *e.g.*, Lazarsfeld et al., 1944; Gerber and Green, 2000; DellaVigna and Gentzkow, 2010; DellaVigna et al., 2016). This justifies door-to-door campaigning. However, the precise estimation the heterogeneity of the impact of interpersonal discussions on the vote is limited by the secret ballot. Several experiments document precisely the impact of discussions on turnout (see, *e.g.*, Gerber and Green, 2000; Imai, 2005; Pons and Liegey, 2019; Green and Gerber, 2019) either using voters registration database or surveys. However, data requirements are huge for the former while sample selection, reporting biases, and costs limit the latter. Hence the importance of analyses based on actual electoral results.

Pons (2018) estimates the impact of door-to-door visits using a large scale experiment embedded the 2012 French presidential left-wing main candidate campaign, *i.e.* François Hollande. The experiment being clustered at the precinct level, he analyzes the impact of assigning a precinct to the treatment group and effectively allocating canvassers to precincts on the turnout and vote shares. He shows, among others, that there is no significant impact on turnout but that there is one on the left-wing candidate's vote shares at both rounds of the election.

Explaining these results with such aggregate data is a complicated task as these depend on the heterogeneity of the voters behaviors, which can not be directly inferred from precinct level data. Specifically, Pons (2018) focuses on two main competing explanations for the impact on the left-wing candidate's vote shares: a persuasive effect on undecided active voters (see, *e.g.*, DellaVigna and Gentzkow, 2010) or a mobilization effect of left-wing nonvoters (potentially joined with a demobilization of right-wing voters). Pons (2018) forms several convincing arguments in favor of the persuasive effect, which could occur mostly through a change in the perceptions of left-wing candidate after the discussions. Pons (2018) provides quantitative arguments justifying that the mobilization effect is negligible using the impact on the difference between the left-wing candidate's vote shares and voter turnout, which remains significant. Our method complements his study and allows to estimate the impact on the vote shares among different types of voters according to their past electoral behavior. Thus, this quantifies precisely the importance of the persuasion effect at the individual level. Our methodology could also be used to estimate the differential impact of the campaign on other subgroups based, *e.g.*, on age, income categories, or working status. Specifically, Appendix 4.5.3 discussed below shows the impact of the visits on vote shares according to the level of education.

4.3.2 Experimental design, data, and method

Experimental design. The experiment started 11 weeks before the first round of the 2012 French presidential election and finished at the second round. It consisted in sending 80,000 left-wing activists knocking on estimated 5 million doors with principal goal to encourage people to vote for the left-wing candidate of the *Parti Socialiste* (PS) which was the mainstream left party in France in 2012. The French presidential election is a two rounds contest, where the two candidates achieving the highest vote shares in the first round qualify for the second. The repartition of the vote shares and the turnout at this election is displayed in Figure 4-2 and I refer to Pons (2018) for a more detailed exposition of the context.

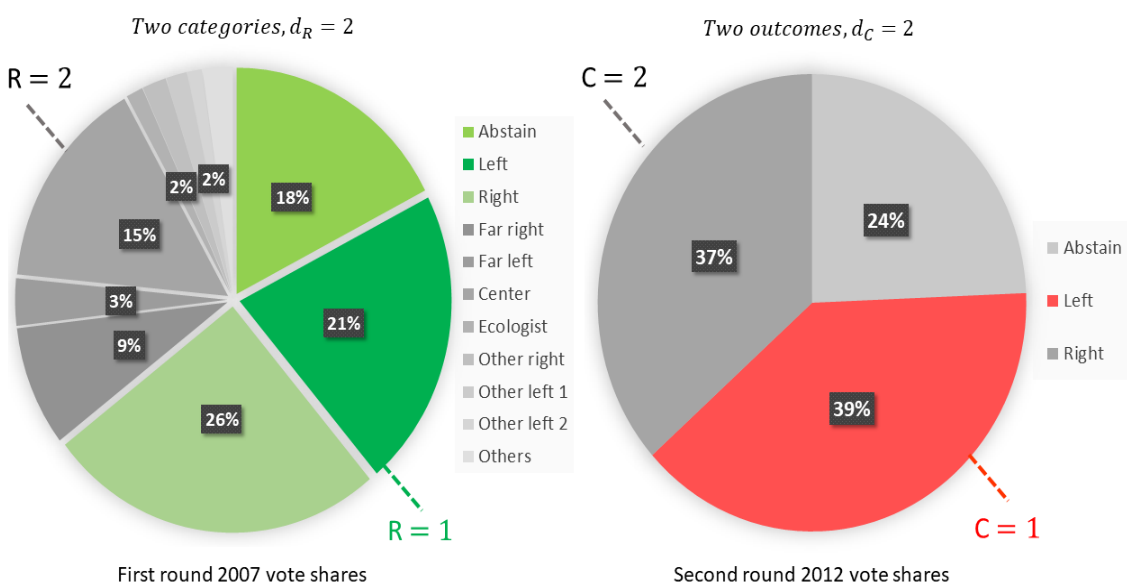
Pons (2018) designed the experiment in order to be able to evaluate the impact of the visits and to maximise their impact, hence adopting a specific stratified randomization. First, the entire country is divided into territories. Then, he defines stratum of 5 precincts ordered by highest $\widetilde{PO}$, which is a computed proxy of the potential to win votes, defined as the fraction of nonvoters multiplied by the left-wing vote share among the active voters at the second round of the 2007 Presidential election. In each stratum, exactly four precincts where randomly assigned to the treatment group and one to the control group. Finally, due to human resources, not all treatment precincts have been allocated to canvassers hence there is imperfect compliance to the allocation prescribed. Thus, to evaluate the effect of allocating a precinct to canvassers, the assignment randomization dummy is used as an instrument. I adopt the same sample of precincts as in Pons (2018) to perform the analysis, which gives 2,665 precincts.

My aim is study the persuasion hypothesis of Pons (2018) empirically. Hence, I focus on the heterogeneity of the impact on the left-wing candidate's 2nd round vote shares ($d_C = 2$) according to whether individuals' first choice at the first round of the 2007 Presidential election qualified ($R = 1$, $d_R = 2$) or not ($R = 2$) for the second round of 2012 (see Figure 4-1 below). As these latter individuals ($R = 2$) have to change their vote in the second round (or to abstain) with respect to their first choice in 2007, they are good proxies for undecided active voters.[13] Thus, the covariate R takes value 1 if they voted for the main left party, the main right party, or abstained at the first round of the 2007 Presidential election, and 2 otherwise.[14] I consider only two categories to keep statistical power, but the point estimates give similar conclusions with three categories. The precincts assignment status T and potential to win votes $\widetilde{PO}$ are used as additional variables in the prediction, see Section 4.1.3

[13] I use groups based 2007 first round choice rather than on 2012 first round because the treatment might have impacted the composition of the groups between treatment and control. The structure and candidates for the main French parties in the elections in 2007 and 2012 are quite close.

[14] By main left party, I refer to *Parti Socialiste* (PS) and by the main right one to *Union pour un mouvement populaire* (UMP).

and below.



Notes: Shares among registered voters, null votes are included in abstention.

Figure 4-1: Electoral results of the first round of 2007 and second round of 2012 Presidential elections, and the two categories and two outcome possibilities used in the first decomposition.

Table 4.5: Description of the variables in $\boldsymbol{m}_{r,1}(\cdot) = \mathbb{E}\left[\boldsymbol{B}_{r,1} | (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \cdot\right]$

Variables	$d_C = 2, d_R = 2$	Description
$\boldsymbol{Y}$	$c = 1$	Left-wing vote shares at the second round
	$c = 2$	Right-wing vote shares plus abstention at the 2012 second round
$\boldsymbol{X}$	$r = 1$	Share of voters whose 2007 first round choice is at the 2012 second
	$r = 2$	Share of voters who have to make a different second round choice
$\boldsymbol{B}$	$r = 1$, $c = 1$	Left candidate vote shares among the first category of voters
	$r = 2$, $c = 1$	Left candidate vote shares among "undecided active voters"
Add. var. $\boldsymbol{Z}$	T	Precinct assigment dummy
	A	Precinct allocation to canvassers dummy (treatment)
	$\boldsymbol{W}$	Controls, including $\widetilde{PO}$, potential to win votes

Notes: The individual level variables are C, the decision to vote for the left-wing candidate or not (2 possibilities $d_C = 2$) and R, which index two categories of voters based on past votes at the first round, see Figure 4-1. Variables in blue are unobserved and in red are observed.

$\mathbb{P}_g(C = 1) = \boldsymbol{Y}_{1,g}$ are the left-wing candidate's vote shares in precinct g, which are observed and used as dependent variable in Pons (2018). T_g denotes a dummy equal to 1 if the precinct was assigned to the treatment group and to 0 if it was assigned to the control group while A_g is a dummy which takes value 1 if the precinct was allocated to the canvassers and 0 otherwise. $\boldsymbol{W}_g$ is a possible vector of controls, always including $\widetilde{PO}_g$. Contrary to Pons (2018), I focus on the impact of the assignment on the votes of sub-groups of the population, *i.e.* on the unobserved $\mathbb{P}_g(C = 1|R = r) = \boldsymbol{B}_{r,1,g}$, which are formed by past voting behavior. Table 4.5 provides a summary of the notations coherent with the previous sections.

Estimation method. I use a direct estimation of the effect of the assignment on $\boldsymbol{B}_{r,1}$ for the two categories $r = 1, 2$,[15] namely, using that the assignment is random conditionally on $\widetilde{PO}$ which is included in $\boldsymbol{W}$, we obtain (4.5) using $\boldsymbol{m}_{r,1}$: $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, t) \mapsto \mathbb{E}\left[\boldsymbol{B}_{r,1} | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{W} = \boldsymbol{w}, T = t\right]$, controlling for $\boldsymbol{W}_g$ nonpara-

[15] Results using an estimation method closer to Pons (2018) in a two stages approach are available upon request and are very close.

metrically. A natural estimator of $\gamma_{r,1}$ is

$$\widehat{\gamma}_{r,1} = \sum_{g \text{ s.t. } T_g=1} \widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, \boldsymbol{W}_g, T_g) - \sum_{g \text{ s.t. } T_g=0} \widehat{\boldsymbol{m}}_{r,1}(\boldsymbol{X}_g, \boldsymbol{Y}_g, \boldsymbol{W}_g, T_g), \qquad (4.20)$$

where $\widehat{\boldsymbol{m}}_{r,1}$ is the estimator of Section 4.2.1, obtained under Assumption 4. Results are stable in our application to the use of different sets of additional variables $\boldsymbol{W}$.

The randomization by stratum based on $\widetilde{PO}$ in territories, which is very specific to this application, generates the need to control also for stratum fixed effects. However, due to dimensionality issues, this can not be included as control in the estimation of $(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, t) \mapsto \mathbb{E}[\boldsymbol{B}_{r,1} | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{W} = \boldsymbol{w}, T = t]$, hence the above estimation does not take it into account otherwise than controlling for $\widetilde{PO}$ in $\boldsymbol{W}$.[16] To take into account this randomization by stratum in the standard errors associated to the estimator (4.20), I compute the latter using bootstrap clustered at the stratum level.

To estimate the effect of the allocation on $\boldsymbol{B}_{r,1,g}$, I consider

$$\boldsymbol{B}_{r,1,g} = g_r(A_g, \boldsymbol{W}_g) + \epsilon_{r,g}, \quad \mathbb{E}[\epsilon_{r,g} | T_g] = 0, \qquad (4.21)$$

for an unknown function g_r, where the effect of the allocation on vote shares among the sub-population of type $R = r$ is

$$\gamma_{r,2} = \mathbb{E}[g_r(1, \boldsymbol{W}_g) - g_r(0, \boldsymbol{W}_g)].$$

This allows to control for $\boldsymbol{W}_g$ nonparametrically handling the endogeneity of A_g. To ensure the coherency of (4.21) with Assumption 4, I impose the additional restrictions $A_g \perp \boldsymbol{X}_g | T_g, \boldsymbol{W}_g$ and $\epsilon_{r,g} \perp \boldsymbol{X}_g | T_g, \boldsymbol{W}_g$. This yields, using the notation $a_{i,j} = \mathbb{P}(A_g = i | T_g = j)$ and $\boldsymbol{W}_g \perp T_g$, for $j = 0, 1$,

$$\mathbb{E}[\boldsymbol{B}_{r,1,g} | T_g = j] = \mathbb{E}[g_r(1, \boldsymbol{W}_g)] a_{1,j} + \mathbb{E}[g_r(0, \boldsymbol{W}_g)] a_{0,j}.$$

[16] Again, one can also consider an estimator of the effect of the assignment which handles stratum fixed effects. However it uses a two stages approach which is valid under stronger parametric assumptions on the controls.

Solving this system yields that $\gamma_{r,2}$ satisfies $\gamma_{r,2} = \gamma_{r,1}/(a_{0,0}a_{1,1} - a_{1,0}a_{0,1})$. Thus, with the empirical counterparts of $a_{i,j}$, $i, j \in \{0, 1\}$ and (4.20), I obtain an estimator for $\gamma_{r,2}$, the effect of the allocation on vote shares among the sub-population of type $R = r$.

4.3.3 Results

Table 4.6 presents the main estimation results. In Table 4.6, individuals in category 2 are proxies for the undecided active voters. The main finding is that Table 4.6 provides suggestive evidence of a persuasive effect on these undecided active voters as the coefficients for the category 2 for the intention-to-treat (ITT) or instrumental variable estimation are significantly positive whereas the ones for the "qualified parties" voters - *i.e.*, category 1 - are not significant. The estimate of $a_{0,0}a_{1,1} - a_{1,0}a_{0,1}$ is 0.552. Results using the alternative two-stages approach which consider stratum fixed effects give estimates close to Table 4.6.

Table 4.6: Impact on the left-wing candidate's 2nd round vote shares among different categories of voters

Category	$R = 1$	$R = 2$, **"Undecided active voters"**
ITT estimation	0.0076	0.0174
	[-0.0379,0.0215]	[0.0040,0.0510]
Instrumental variable estimation	0.0138	0.0316
	[-0.0687,0.0389]	[0.0073,0.0925]

Notes: Intention-to-treat (ITT) estimation shows the effect of a precinct being assigned to the treatment group (ITT results from (1)) on the two different types of individuals: category 1 are individuals who voted for left (Royal), right (Sarkozy), or abstained at the 2007 election while category 2 are the others, which proxy undecided active voters. Instrumental variable estimation shows the effect of a precinct being allocated to canvassers using the assignment dummy T as instrument. All the results use $W = \widetilde{PO}$ as control, which enters nonparametrically. The unit of observation is the unit of randomization (precinct or municipality) and each regression is based on 2,665 observations. 95% bootstrap confidence intervals are in parentheses, computed via 300 stratum-clustered bootstrap. The outcome variables are estimated using our main adaptive estimator based on Legendre functions. Category 1 constitutes 65% of the registered voters in a precinct in average, with a minimum of 28.5% and a maximum of 93%.

If we assume that the visits have zero impact on the right-wing candidate supporters and that the effect on the left-wing voters is positive, the fact that there is no effect on category 1 in Table 4.6 gives an upper bound on the mobilization impact of left-wing nonvoters. Thus, there might not be any mobilization effect in this experiment. However, Pons (2018) describes the alternative explanation that the visits could have demobilized right-wing voters and simultaneously mobilized left-wing nonactive ones, resulting in the observed zero effect. He argues that it is not likely to be the case. The results of Decomposition 2 in Table 4.7 in Appendix 4.5.3 go in his direction. Here, category 1 are the left and right wing voters in 2007 first round. Hence, under this alternative demobilization hypothesis, the effect in category 1 should be negative and the effect on category 2 remain positive. As estimates of Table 4.7 for category 1 are positive and non-significant, there might not be any mobilization or demobilization effects in this experiment.

Finally, Appendix 4.5.3 considers a different decomposition of the electorate, in two categories according to the level education. Specifically, I distinguish voters according to whether they graduated from high school or not.[17] There are two possible predictions in the literature for the differential impact of political communication according to education. On the one hand, classical models of Bayesian updating (see, *e.g.*, Zaller, 1992; Box and Tiao, 1973) predict that less knowledgeable voters are going to react more to credible information. On the other hand, these voters are less likely to be reached by classical political communication channels, hence are less prone to change their mind (see, *e.g.*, Prior, 2006; Le Pennec and Pons, 2019). Reaching individuals without distinction, the door-to-door visits of this experiment allow to concentrate on testing the former explanation. I keep the same vector of controls as previously. Results about the impact on the left-wing candidate vote shares at the second round on these two categories are displayed in Table 4.8. It shows that there is a positive and significant impact of the visits on less-educated voters, with a 0.0101 [0.0010,0.0146] estimate for ITT and 0.0183 [0.0016,0.0265] for instrumental variable estimation. There is negative but not significant effect on the more educated ones. Results are robust to the use of the first or second round 2007 left-wing vote shares as additional controls. These results suggest that beliefs might be more affected by the information brought by personal visits among those with less education. This underlines the importance of prior knowledge in the persuasion mechanism. This is coherent with Bayesian models of beliefs formation and results in Le Pennec and Pons (2019) on the differential impact of campaigns on consistency between vote intention and vote choice according to education. Results of Cantoni and Pons (2018) also quantify the higher importance of the local context for younger voters, which have generally less political knowledge.

[17] More specifically, according to whether they have their *baccalauréat*.

4.4 Conclusion

This paper provides an ecological inference approach in a nonparametric framework. I improve on the literature assumptions to perform inference and provide tools to perform sensitivity analysis to the assumptions which yield point identification. First, without adding more structure, I relate the partial identification of the prediction to the identified set described in the literature with one group, *e.g.*, Cross and Manski (2002*a*). A new characterisation it is provided in Gaillac (2020). Second, when the researcher considers credible the nonparametric assumption of no contextual effects, my approach provides a constructive identification strategy and my plug-in nonparametric estimator is minimax adaptive. I provide evidence on simulations that the method performs well in finite sample, even with more than two categories and with some contextual effects.

One limitation is the fact that my estimator is formulated as a ratio, which then requires trimming of the density of the regressors, which can be difficult to calibrate. I let for future research the inclusion of tools such as warped bases (see, *e.g.*, Chagny, 2015) which could overcome these issues. Other avenues of research will use my constructive identification to perform semi-parametric estimation and to allow for the combination of my estimator with individual survey data (like in, *e.g.*, Wakefield, 2004).

Finally, nonparametric ecological inference can also be used in other contexts. First, this method can be adapted to demand estimation with market level data, taking a different direction from the current nonparametric extensions of the BLP model (see, *e.g.*, Compiani, 2019). Second, my methodology can be used in experiments where either the outcome is protected or a covariate of interest might only be available for several groups in another dataset that can not be merged. Finally, closer to my application is the analysis of split-ticket voting (see, *e.g.*, Burden and Kimball, 1998). Theoretically, split-ticket exemplifies the desire for a divided government (see, among others, Alesina and Rosenthal, 1995) which sees voters allocating their vote differently at the House, Senate, Governor, or Presidential election to counterbalance the power of either representative. Many models analyzing the desire for a divided government are based on spatial ideological models (see, *e.g.*, Knight, 2014; Mebane, 2000) and rather use surveys than ecological inference.

4.5 Appendix

This section contains several complements and proofs for the main results of the article. Several auxiliary derivations are in a Supplementary Appendix, along with additional complementary results. Specifically, this appendix presents additional results on identification with $d_C = 3$ (Appendix 4.5.1), the complements and proofs of the main theorems (Appendix 4.5.2).

Notations

The notations $\cdot$, $\star$ are used to denote a variable in a function. and $\mathbb{1}\{A\}$ for the indicator function of a set A. For all $r \in \mathbb{R}$, $\boldsymbol{r}$ is the vector, which dimension will be clear from the text, where each entry is r. For a vector $\boldsymbol{X} \in \mathbb{R}^{d_R}$, we use $\operatorname{diag}(\boldsymbol{X})$ to denote the diagonal matrix in $\mathcal{M}_{d_R,d_R}(\mathbb{R})$ which diagonal elements are the elements of $\boldsymbol{X}$. For all $(f_g)_{m\in\mathbb{N}_0}$ functions with values in $\mathbb{C}$, and $\boldsymbol{m} \in \mathbb{N}_0^d$, denote by $f_{\boldsymbol{m}} = \prod_{k=1}^d f_{\boldsymbol{m}_k}$.

4.5.1 Identification when $d_C > 2$

Partial identification when $d_C > 2$ without contextual effects

Proposition 7 (Partial identification, $d_C > 2$)**.** *Let the distribution of* $(\boldsymbol{B}, \boldsymbol{X}, \boldsymbol{Y})$ *satisfy* (4.1) *and define the restriction* $\mathcal{R}_0$ *corresponding to assumptions 1, 2, and 3. Let* $d_C > 2$*, then* $\mathcal{J}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma, \mathcal{R}_0)$ *is included into the set of functions of the form* $\boldsymbol{m} = \boldsymbol{M}/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$*, where* $\boldsymbol{M}_{r,c} : \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}} \mapsto [0,1]$ *for* $r = 1, \ldots, d_R$ *and* $c = 1, \ldots, d_C$ *are continuous functions which admit a continuous derivative with respect to* $\boldsymbol{y}_c$*, for* $c = 1, \ldots, d_C - 1$*,* $\boldsymbol{M}_{r,C} = 1 - \sum_{c=1}^{d_C-1} \boldsymbol{M}_{r,c}$*, and, for all* $r = 1, \ldots, d_R$*,* $c = 1, \ldots, d_C - 1$*,* *and* $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$,

$$\sum_{r=1}^{d_R-1} \boldsymbol{x}_r \boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}) + (1-\boldsymbol{x}^\top \mathbf{1})\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}), \tag{4.22}$$

$$\sum_{c=1}^{d_C-1} \partial_{\boldsymbol{y}_c} \boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}) = \sum_{c=1}^{d_C-1} \partial_{\boldsymbol{y}_c} \boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l=r\}) \partial_{\boldsymbol{x}_l} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{x}), \tag{4.23}$$

where $\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) := f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{x})\boldsymbol{y}_c$. *Moreover, for all* $c = 1,\dots,d_C-1$ *and* $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}$, $\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}_1,\dots,\boldsymbol{y}_c = 0,\dots,\boldsymbol{y}_{d_C-1}) = 0$.

Proposition 7 shows that, when $d_C > 2$, the parameter of interest satisfies a system of coupled partial differential equations. However, the solutions are in general not unique nor explicit.

Identification when $d_C = 3$ when restricting the dimension of the unobserved heterogeneity

I consider the case where the researcher assumes that some random coefficients are linearly dependent of the others. This reduces the dimension of the unobserved heterogeneity, hence reducing the size of the identified set when we have more than two choices. I consider the case $d_C = 3$ and describe in Remark 3 the set of assumptions that one would make to handle higher dimensional cases.

Assumption 10 (Restricted heterogeneity, $d_C = 3$)**.** *Let* $\boldsymbol{\omega} := \big((r,1)_{r\in\{1,\dots,d_R\}},(d_R,2)\big)$ *which is a sequence of length* $d := d_R + d_C - 2 = d_R + 1$ *of indexes. Let the* d *coefficients* $(\boldsymbol{B}_{\boldsymbol{\omega}_k})_{k=1,\dots,d}$ *be the latent unobserved heterogeneity, that I denote by* $\boldsymbol{U} := (\boldsymbol{U}_1,\dots,\boldsymbol{U}_d)$, *hence*

$$\boldsymbol{U}_l := \boldsymbol{B}_{\boldsymbol{\omega}_l}, \quad l = 1,\dots,d.$$

The (d_R-1) *other random coefficients can be expressed as*

$$\boldsymbol{B}_{r,2} = \sum_{k=1}^{d} \boldsymbol{a}_{r,k} \boldsymbol{U}_k, \quad r = 1,\dots,d_R-1,$$

where $\boldsymbol{a} \in \mathcal{M}_{d_R-1,d}(\mathbb{R})$ are fixed coefficients.

Remark 1 (More general formulation)**.** *A slightly more general formulation would assume instead of Assumption 10 that these are d latent sources of random unobserved heterogeneity $(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_d) = \boldsymbol{U}$, and that the coefficients $\boldsymbol{B}_{r,c}$ depend linearly of $\boldsymbol{U}$. However, the simplified set up that I consider is more transparent, facilitates testing and estimation of $\boldsymbol{a}$, and amounts to the same type of assumptions.*

Note that in the case of $d_C = 2$, Assumption 10 is not a restriction as $(d_R - 1)(d_C - 2) = 0$, which is in line with Theorem 1. This yields for $d_C = 3$,

$$\boldsymbol{Y}_1 = \sum_{r=1}^{d_R} \boldsymbol{U}_r \boldsymbol{X}_r \tag{4.24}$$

$$\boldsymbol{Y}_2 = \sum_{r=1}^{d_R-1} \sum_{k=1}^{D} \boldsymbol{a}_{r,2,k} \boldsymbol{U}_k \boldsymbol{X}_r + \boldsymbol{U}_{d_R+1} \boldsymbol{X}_{d_R}. \tag{4.25}$$

Assumption 1 yields the system of equations

$$\mathbb{E}\left[\boldsymbol{Y}_1 | \boldsymbol{X} = \boldsymbol{x}\right] = \sum_{r=1}^{d_R-1} \left(\mathbb{E}\left[\boldsymbol{U}_r\right] - \mathbb{E}\left[\boldsymbol{U}_{d_R}\right]\right) \boldsymbol{x}_r + \mathbb{E}\left[\boldsymbol{U}_{d_R}\right]$$

$$\mathbb{E}\left[\boldsymbol{Y}_2 | \boldsymbol{X} = \boldsymbol{x}\right] = \sum_{r=1}^{d_R-1} \left(\sum_{k=1}^{d_R+1} \boldsymbol{a}_{r,k} \mathbb{E}\left[\boldsymbol{U}_k\right] - \mathbb{E}\left[\boldsymbol{U}_{d_R+1}\right]\right) \boldsymbol{x}_r + \mathbb{E}\left[\boldsymbol{U}_{d_R+1}\right].$$

This yields using Assumption 2 with $d = d_R + 1$ that $\mathbb{E}\left[\boldsymbol{U}_k\right]$ for $k = 1, \ldots, d_R + 1$ and $\boldsymbol{v}_r := \sum_{k=1}^{d_R+1} \boldsymbol{a}_{r,k} \mathbb{E}\left[\boldsymbol{U}_k\right]$ and $r = 1, \ldots, d_R - 1$ are identified. Thus, I obtain a system of $d_R - 1$ equations and $(d_R - 1)(d_R + 1)$ unknowns coefficients $\boldsymbol{a}_{r,k}$. If $\boldsymbol{a}$ is known, then Proposition 8 below shows point identification in a constructive way, which is the estimation strategy used in Section 4.2 in the case $d_C > 2$. Otherwise, Proposition 8 describes the identified set.

Let me introduce some notations. Under Assumption 10 and (4.24)-(4.25) I obtain, for $c = 1, 2$, ($c = 3$ being redondant with the others due to the constraint $\boldsymbol{Y}^\top \boldsymbol{1} = 1$, I suppress it),

$$\boldsymbol{y}_c = \sum_{k=1}^{d} \boldsymbol{W}_{c,k}(\boldsymbol{x})\mathbb{E}\left[\boldsymbol{U}_k|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\right], \tag{4.26}$$

where $\boldsymbol{W}_{1,k}(\boldsymbol{x}) := \boldsymbol{x}_k \mathbb{1}\{k \leq d_R\}$ for $k = 1, \ldots, d$ and

$$\boldsymbol{W}_{2,k}(\boldsymbol{x}) := \sum_{r=1}^{d_R-1} \boldsymbol{a}_{r,k}\boldsymbol{x}_r + \mathbb{1}\{k = d_R + 1\}\boldsymbol{x}_{d_R}. \tag{4.27}$$

For convenience, I use $\boldsymbol{V}_k : (\boldsymbol{x}, \boldsymbol{y}) \mapsto \mathbb{E}\left[\boldsymbol{U}_k|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\right] f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})$, for $k = 1, \ldots, d$ and $\boldsymbol{M}_{r,c} : (\boldsymbol{x}, \boldsymbol{y}) \mapsto \mathbb{E}\left[\boldsymbol{B}_{r,c}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\underline{\boldsymbol{y}}\right] f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})$, for $r = 1, \ldots, d_R$, $c = 1, 2$.

Identification strategy. Let me now explain the steps of the identification strategy:

(**Step** 1) I can express the coefficients of $\boldsymbol{M}$ in terms of $\boldsymbol{V}$ through

$$\begin{aligned} \boldsymbol{M}_{r,1}(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{V}_r(\boldsymbol{x}, \boldsymbol{y}) \\ \boldsymbol{M}_{r,2}(\boldsymbol{x}, \boldsymbol{y}) &= \sum_{k=1}^{d_R+1} \boldsymbol{a}_{r,k}\boldsymbol{V}_k(\boldsymbol{x}, \boldsymbol{y}), \end{aligned} \tag{4.28}$$

for $r = 1, \ldots, d_R$. Hence the aim is to recover $\boldsymbol{V}$.

(**Step** 2) I express $\boldsymbol{V}_l$ for $l = d_R, d_R + 1$ as function of $\boldsymbol{V}_l$ for $l = 1, \ldots, d_R - 1$. Denote by $\boldsymbol{\rho}_c(\boldsymbol{x}, \boldsymbol{y}) := f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})\boldsymbol{y}_c$, for $c = 1, 2$. Under Assumption 11.1 below and using (4.26), the system with $d - d_R + 1 = 2$ unknowns,

$\boldsymbol{V}_R(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{V}_{d_R+1}(\boldsymbol{x},\boldsymbol{y})$,

$$\boldsymbol{x}_{d_R}\boldsymbol{V}_{d_R}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\rho}_1(\boldsymbol{x},\boldsymbol{y}) - \sum_{k=1}^{d_R-1}\boldsymbol{x}_k\boldsymbol{V}_k(\boldsymbol{x},\boldsymbol{y}) \tag{4.29}$$

$$\boldsymbol{W}_{2,d_R}(\boldsymbol{x})\boldsymbol{V}_{d_R}(\boldsymbol{x},\boldsymbol{y}) + \boldsymbol{W}_{2,d_R+1}(\boldsymbol{x})\boldsymbol{V}_{d_R+1}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\rho}_2(\boldsymbol{x},\boldsymbol{y}) - \sum_{k=1}^{d_R-1}\boldsymbol{W}_{2,k}(\boldsymbol{x})\boldsymbol{V}_k(\boldsymbol{x},\boldsymbol{y}), \tag{4.30}$$

has a unique solution, for $l = d_R, d_R+1$,

$$\boldsymbol{V}_l(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\sigma}_{l-d_R+1}(\boldsymbol{x},\boldsymbol{y}) + \sum_{k=1}^{d_R-1}\boldsymbol{Q}_{l-d_R+1,k}(\boldsymbol{x})\boldsymbol{V}_k(\boldsymbol{x},\boldsymbol{y}), \tag{4.31}$$

where, for $k = 1,\ldots,d_R-1$,

$$\boldsymbol{\sigma}_1(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{\rho}_1(\boldsymbol{x},\boldsymbol{y})}{\boldsymbol{x}_{d_R}},\ \boldsymbol{Q}_{1,k}(\boldsymbol{x}) = -\frac{\boldsymbol{x}_k}{\boldsymbol{x}_{d_R}}, \tag{4.32}$$

$$\boldsymbol{\sigma}_2(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{x}_{d_R}\boldsymbol{\rho}_2(\boldsymbol{x},\boldsymbol{y}) - \boldsymbol{W}_{2,d_R}(\boldsymbol{x})\boldsymbol{\rho}_1(\boldsymbol{x},\boldsymbol{y})}{\boldsymbol{x}_{d_R}\boldsymbol{W}_{2,d_R+1}(\boldsymbol{x})}, \tag{4.33}$$

$$\boldsymbol{Q}_{2,k}(\boldsymbol{x}) = \frac{\boldsymbol{x}_{d_R}\boldsymbol{W}_{2,k}(\boldsymbol{x}) - \boldsymbol{x}_k\boldsymbol{W}_{2,d_R}(\boldsymbol{x})}{\boldsymbol{x}_{d_R}\boldsymbol{W}_{2,d_R+1}(\boldsymbol{x})}. \tag{4.34}$$

(**Step** 3) Then, I identify $\boldsymbol{V}_l$ for $r = 1,\ldots,d_R-1$ as solution of a system of coupled partial transport differential equations, see Appendix 4.1 for details.

Denote by $\widetilde{\boldsymbol{Q}} \in \mathcal{M}_{d_R-1,d_R-1}(\mathbb{R})$ with coefficients $\widetilde{\boldsymbol{Q}}_{r,k}(\boldsymbol{x}) := \boldsymbol{a}_{r,k} + \sum_{l=d_R}^{d_R+1}\boldsymbol{a}_{r,l}\boldsymbol{Q}_{l-d_R+1,k}(\boldsymbol{x})$, for $r = 1,\ldots,d_R-1$ and $k = 1,\ldots,d_R-1$.

Assumption 11. *When $d_C = 3$, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$,*

1. $\boldsymbol{x}_{d_R}\boldsymbol{W}_{2,d_R+1}(\boldsymbol{x}) \neq 0$*;*
2. $\widetilde{\boldsymbol{Q}}(\boldsymbol{x}) \in \mathcal{M}_{d_R-1,d_R-1}(\mathbb{R})$ *is diagonalisable:* $\widetilde{\boldsymbol{Q}}(\boldsymbol{x}) = \boldsymbol{P}^{-1}(\boldsymbol{x})\, diag\,(\boldsymbol{\Lambda}(\boldsymbol{x}))\,\boldsymbol{P}(\boldsymbol{x})$*, where* $diag\,(\boldsymbol{\Lambda}(\boldsymbol{x}))$ *is a diagonal matrix and* $\boldsymbol{P}(\boldsymbol{x})$ *is an orthogonal matrix.*

Proposition 8. *Consider $d_C = 3$. Let the distribution of $(\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y})$ satisfy (4.1) and define the restriction $\mathcal{R}_1$ corresponding to assumptions 1, 2, 3, 10, and 11. Then*

$\mathcal{J}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma, \mathcal{R}_1)$, the identified set for $\boldsymbol{m}$ is included into the set of functions taking the form, for all $r = 1, \ldots, d_R$, $c = 1, 2$, and $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$,

$$\boldsymbol{m}_{r,c}(\boldsymbol{x}, \boldsymbol{y}) = \Pi_{r,c}\left[\boldsymbol{\zeta}, \boldsymbol{\sigma}\right](\boldsymbol{x}, \boldsymbol{y}), \tag{4.35}$$

Π is a linear operator from $\mathcal{M}_{d\times(d_R-1)}\left(l^{\infty}\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right)\right)\times\mathcal{M}_{2\times 1}\left(l^{\infty}\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right)\right)$ to $\mathcal{M}_{d_R,d_C}(l^{\infty}\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right))$,

$$\Pi_{r,1}\left[\boldsymbol{\zeta}, \boldsymbol{\sigma}\right] = \boldsymbol{P}^{-1} Diag\left(\boldsymbol{P}\boldsymbol{K}\boldsymbol{\zeta}\right), \quad r = 1, \ldots, d_R - 1 \tag{4.36}$$

$$\Pi_{d_R,1}\left[\boldsymbol{\zeta}, \boldsymbol{\sigma}\right] = \boldsymbol{Q}_{1,\cdot}^{\top}\boldsymbol{P}^{-1} Diag\left(\boldsymbol{P}\boldsymbol{K}\boldsymbol{\zeta}\right) + \boldsymbol{\sigma}_1, \tag{4.37}$$

$$\Pi_{r,2}\left[\boldsymbol{\zeta}, \boldsymbol{\sigma}\right] = \widetilde{\boldsymbol{Q}}_{r,\cdot}^{\top}\boldsymbol{P}^{-1} Diag\left(\boldsymbol{P}\boldsymbol{K}\boldsymbol{\zeta}\right) + \boldsymbol{a}_{r,d_R}\boldsymbol{\sigma}_1 + \boldsymbol{a}_{r,d_R+1}\boldsymbol{\sigma}_2, \quad r = 1, \ldots, d_R, \tag{4.38}$$

where, $\boldsymbol{\sigma}$ is defined via (4.32)-(4.34),

$$\boldsymbol{K}(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{x}_1 - 1 & \ldots & \boldsymbol{x}_{d_R-1} & 1 & 1 \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 - 1 & \boldsymbol{x}_{d_R-1} & 1 & 1 \\ \vdots & \ddots & & 1 & 1 \\ \boldsymbol{x}_1 & & \boldsymbol{x}_{d_R-1} - 1 & 1 & 1 \end{pmatrix} \tag{4.39}$$

and where $\boldsymbol{\zeta} \in \mathcal{M}_{d\times(d_R-1)}\left(l^{\infty}\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right)\right)$ with $\boldsymbol{\zeta}(\boldsymbol{x}, \boldsymbol{y})/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})$ equals to

$$\begin{pmatrix} \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{x}_1} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_1(\boldsymbol{x})(\boldsymbol{y}_1 - v)|\underline{\boldsymbol{x}})dv & \ldots & \ldots \\ \vdots & & \vdots \\ \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{x}_{R-1}} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_1(\boldsymbol{x})(\boldsymbol{y}_1 - v)|\underline{\boldsymbol{x}})dv & & \vdots \\ f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})\boldsymbol{y}_1 & \ldots & f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})\boldsymbol{y}_1 \\ \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{y}_2}\boldsymbol{\rho}_2(\boldsymbol{x}, v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_1(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv & \ldots & \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{y}_2}\boldsymbol{\rho}_2(\boldsymbol{x}, v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_{R-1}(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv \end{pmatrix}, \tag{4.40}$$

where, $\boldsymbol{\rho}_c(\boldsymbol{x}, \boldsymbol{y}) := f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})\boldsymbol{y}_c$. When $\boldsymbol{a} \in \mathcal{M}_{d_R-1,d}(\mathbb{R})$ in Assumption 10 is known, then this set is reduced to one element.

Remark 2 ($d_C = 2$ as particular case)**.** *Using Proposition 1, the case $d_C = 2$ appears as a particular case where no further assumption has to be made on the random* *coefficients to obtain point identification. When* $d_C = 2$, (4.7) *can be rewritten as*

$$\boldsymbol{m}_{r,1} = \Pi_{r,1}\left[\zeta\right] := (\boldsymbol{K}\boldsymbol{\zeta})_r,$$

where Π *is a linear operator from* $\mathcal{M}_{d\times(d_R-1)}\left(l^\infty\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right)\right)$ *to* $\mathcal{M}_{d_R,d_C}\left(l^\infty\left(\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}\right)\right)$ *and* $\boldsymbol{K}$ *is defined like* (4.39) *with only one column of 1 and, for all* $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$,

$$\boldsymbol{\zeta}(\boldsymbol{x},\boldsymbol{y}) := \left(\frac{\partial_{\boldsymbol{x}_1} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}_1|\underline{\boldsymbol{x}})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})}, \dots, \frac{\partial_{\boldsymbol{x}_{d_R-1}} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}_1|\underline{\boldsymbol{x}})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}_1|\underline{\boldsymbol{x}})}, \boldsymbol{y}_1\right)^\top \tag{4.41}$$

and (4.35) *is* (4.7), *hence this set is also reduced to one element.* 8 *shows that cases* $d_C = 2$ *and* $d_C = 3$ *share a similar structure, where the components needing to be estimated nonparametrically are all the elements of* $\boldsymbol{\zeta}$.

The proof of Proposition 8 is constructive and one can directly adapt the plug-in approach to perform estimation in Section 4.2 using an estimator of $\boldsymbol{\zeta}$ defined in (4.41) for $d_C = 2$, estimating (4.40) for $d_C = 3$. Indeed, Proposition 8 and Remark 2 underline that in cases $d_C = 2$ and $d_C = 3$, one has to nonparametrically estimate the elements of $\boldsymbol{\zeta}$, as my parameter of interest is the image by the linear operator Π (which is also bounded under Assumption 11) of $\boldsymbol{\zeta}$ and $\boldsymbol{\sigma}$. Estimation of

$$(\boldsymbol{x},\boldsymbol{y}) \mapsto \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{x}_l} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_r(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv$$

for $r = 1, \dots, d_R - 1$ and $l = 1, \dots, d_R - 1$, can be done similarly to Section 4.2, while the other components of $\boldsymbol{\zeta}$ in Proposition 8 imply estimating also $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$, $f_{\underline{\boldsymbol{X}}}$, and $\partial_{\boldsymbol{y}_c} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ (for $d_C = 3$ only), when these quantities exist.

Remark 3 (Cases $d_C > 3$)**.** *Using a similar reasoning as in the proof of Proposition* 8, *one could handle nonparametrically the cases* $d_C > 3$, *assuming that the matrices* $\widetilde{\boldsymbol{Q}}^c(\boldsymbol{x})$ *which appear in the system of coupled differential equations all commute by pairs for* $c = 1, \dots, d_C - 1$ *(or equivalently that they are diagonalisable in the same basis), which puts more restrictions on the coefficients* $\boldsymbol{a}$. *I left this for future research.*

4.5.2 Complements and proofs of the main theorems

Notations and Preliminaries. For notational simplicity, we denote the multivariate Fourier transform of measures on the set of matricesby

$$\mathcal{F}\left[\mu\right](\boldsymbol{x}) = \int_{\mathcal{M}_{d_R,d_C}(\mathbb{R})} e^{i<\boldsymbol{b},\boldsymbol{x}>} d\mu(\boldsymbol{b}), \tag{4.42}$$

where $<\boldsymbol{b},\boldsymbol{x}>= \mathrm{Tr}(\boldsymbol{b}^\top \boldsymbol{x}) = \sum_{r=1}^{d_R}\sum_{c=1}^{d_C} \boldsymbol{b}_{r,c}\boldsymbol{x}_{r,c}$ is the inner product between matrices and Tr is the trace operator. This notation is simply a compact way to denote the multivariate Fourier transform, but one could also fix a way to vectorise the matrix and use the usual multivariate Fourier transform. I denote by $\underline{\boldsymbol{B}}$ is the submatrix of $\boldsymbol{B}$ keeping only the (d_C-1) first columns, hence of dimension $d_R\times(d_C-1)$.

Proof of Proposition 1. Consider $\mathbb{E}\left[\boldsymbol{B}|\boldsymbol{X}=\cdot_1,\boldsymbol{Y}=\cdot_2\right]\in\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma,\mathcal{R})$. Taking conditional expectation in (4.1) with respect to $(\boldsymbol{X},\boldsymbol{Y})$, we obtain, a.e. $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$,

$$\forall c=1,\dots,d_C,\quad \boldsymbol{y}_c = \sum_{r=1}^{R}\mathbb{E}\left[\boldsymbol{B}_{r,c}|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\right]\boldsymbol{x}_r.$$

This yields that $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}(\Gamma,\mathcal{R})$ is included in the set on the right-hand-side of the equation of Proposition 1.

Consider now $\boldsymbol{m}:[0,1]^{d_R\times d_C}\mapsto\mathcal{M}_{d_R,d_C}([0,1])$, such that, for all $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$, $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{y})\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})$, Using the constraint $\boldsymbol{Y}\boldsymbol{1}=1$, we consider the first $C-1$ equations in (4.1) as the last one can be deduced from the others. We have, for all $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\boldsymbol{X},\underline{\boldsymbol{Y}}}$ and understanding the integral below as a matrix where we integrate component by component,

$$\mathbb{E}\left[\underline{\boldsymbol{B}}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right] = \int_{\mathbb{R}^{d_R\times(d_C-1)}} \boldsymbol{b}\, d\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\boldsymbol{b}|\boldsymbol{x},\boldsymbol{y}).$$

Thus, we want to show that we can build $\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})$ such that $\mathbb{E}\left[\underline{\boldsymbol{B}}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right]=\boldsymbol{m}(\boldsymbol{x},\boldsymbol{y})$ (using an abuse of notation for $\boldsymbol{m}$). Consider a compactly supported distibution $\mathbb{P}_{\boldsymbol{B}|\boldsymbol{X},\boldsymbol{Y}}$ satisfing (4.3)-(4.4), which is characterized, for all $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\boldsymbol{X},\underline{\boldsymbol{Y}}}$ by its Fourier transform $\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})\right]$, which is analytic. Then, in the Taylor expansion of $\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})\right]$ at 0, we modify the terms related to the first partial derivatives replacing $\partial_{\boldsymbol{t}_{r,c}}\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})\right](0) = i\mathbb{E}\left[\underline{\boldsymbol{B}}_{r,c}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right]$ by $i\boldsymbol{m}_{r,c}(\boldsymbol{x},\boldsymbol{y})$. We obtain this way the Taylor expansion of $\mathcal{F}\left[\widetilde{\mathbb{P}}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})\right]$ which characterizes the distribution $\widetilde{\mathbb{P}}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\cdot|\boldsymbol{x},\boldsymbol{y})$ and which satifies our requirements because $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{I}(\boldsymbol{x},\boldsymbol{y})$. This yields the second inclusion, hence the result. $\square$

Proof of Proposition 2. Start with the proof of (P2.2). Using $\boldsymbol{Y}_1+\boldsymbol{Y}_2=1$, the first part of the proof is Proposition 2.2 in Beran and Millar (1994). The second part of (P2.2) can be deduced from the first one using the Bayes' theorem, (4.3)-(4.4), and Assumption 1 which yield, for all $(\boldsymbol{b},\boldsymbol{x},\boldsymbol{y}) \in \mathcal{M}_{d_R\times(d_C-1)}([0,1])\times\mathbb{S}_{\boldsymbol{X}}\times\mathbb{S}_{\underline{\boldsymbol{Y}}}$, $\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\boldsymbol{b}|\boldsymbol{x},\boldsymbol{y}) = \mathbb{1}\{\boldsymbol{y}=\boldsymbol{b}^\top\boldsymbol{x}\}\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b})/\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$.
The proof of (P2.a) is a consequence of Corollary 1 in Masten (2017) once we have used $\boldsymbol{Y}^\top\mathbf{1}=1$ to consider only equations related to $c=1,\dots,d_C-1$ in (4.3)-(4.4).
Let us now prove (P2.1) and (P2.b). Let $d_C, d_R \geq 2$. Using the constraint $\boldsymbol{Y}\mathbf{1}=1$, we consider the first d_C-1 equations in (4.3)-(4.4) because the last one can be deduced from the others. Hereafter $\underline{\boldsymbol{B}}$ is thus a $d_R\times(d_C-1)$ random matrix with the d_C-1 first columns of $\boldsymbol{B}$. We have, using Bayes' theorem for the second equality, for a.e. $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$ and for all $r=1,\dots,d_R$, $c=1,\dots,d_C-1$,

$$\begin{aligned}
\mathbb{E}\left[\underline{\boldsymbol{B}}_{r,c}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\underline{\boldsymbol{y}}\right] &= \int_{\mathcal{M}_{d_R\times(d_C-1)}(\mathbb{R})} \boldsymbol{b}_{r,c}d\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X},\underline{\boldsymbol{Y}}}(\boldsymbol{b}|\boldsymbol{x},\underline{\boldsymbol{y}}) \\
&= \int_{\mathcal{M}_{d_R\times(d_C-1)}(\mathbb{R})} \boldsymbol{b}_{r,c}\frac{\mathbb{P}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{B}},\boldsymbol{X}}(\underline{\boldsymbol{y}}|\boldsymbol{b},\boldsymbol{x})}{\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\underline{\boldsymbol{y}}|\boldsymbol{x})}d\mathbb{P}_{\underline{\boldsymbol{B}}|\boldsymbol{X}}(\boldsymbol{b}|\boldsymbol{x}) \\
&= \int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} \frac{\boldsymbol{b}_{r,c}}{\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\underline{\boldsymbol{y}}|\boldsymbol{x})}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) \text{ (using Assumption 1)},
\end{aligned} \tag{4.43}$$

where $\mathcal{I}(\boldsymbol{x},\boldsymbol{y})$ is defined like in (4.6). When $d_C=2$, under assumptions 1 and 2, using (P2.2), $\mathbb{P}_{\boldsymbol{B}}$ is identified. Thus, we directly have from (4.43) that $(\boldsymbol{x},\boldsymbol{y}) \mapsto \mathbb{E}\left[\boldsymbol{B}|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\right]$ is also identified.
Consider now the case $d_C>2$. For simplicity, we condider the case $d_C=3$ and $d_R=2$,

as the other cases can be adapted from it. Take $f^1_{\underline{\boldsymbol{B}}}$ as, for all $\boldsymbol{b} \in \mathcal{M}_{2,2}([0,1])$,

$$f^1_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) = \frac{1}{\mathcal{Z}} \prod_{r=1}^{2} \prod_{c=1}^{2} \mathbb{1}\{\boldsymbol{b}_{r,c} \in [0,1]\} \boldsymbol{b}_{r,c},$$

where $\mathcal{Z}$ is a normalisation constant. Consider a second distribution, for all $\boldsymbol{b} \in \mathcal{M}_{2,2}([0,1])$,

$$f^2_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) := f^1_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) + \gamma \mathbb{1}\{\boldsymbol{b} \in \mathcal{M}_{2,2}([0,1])\} \left(\partial_{11}\partial_{22} - \partial_{12}\partial_{21}\right) f^1_{\underline{\boldsymbol{B}}}(\boldsymbol{b}),$$

γ such that $f^2_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) \geq 0$ for all $\boldsymbol{b} \in \mathcal{M}_{2,2}([0,1])$. Note that we have

$$\frac{1}{\mathcal{Z}} \int_{\mathcal{M}_{d_R \times (d_C - 1)}([0,1])} \left(\boldsymbol{b}_{2,1}\boldsymbol{b}_{1,2} - \boldsymbol{b}_{1,1}\boldsymbol{b}_{2,2}\right) d\boldsymbol{b} = 0$$

hence $\int_{\mathcal{M}_{d_R \times (d_C - 1)}([0,1])} f^2_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) d\boldsymbol{b} = 1$. This yields, for all $\boldsymbol{z} \in \mathcal{M}_{2,2}(\mathbb{R})$,

$$\mathcal{F}\left[f^2_{\underline{\boldsymbol{B}}}\right](\boldsymbol{z}) = \left(1 - \gamma(\boldsymbol{z}_{11}\boldsymbol{z}_{22} - \boldsymbol{z}_{12}\boldsymbol{z}_{21})\right) \mathcal{F}\left[f^1_{\underline{\boldsymbol{B}}}\right](\boldsymbol{z}),$$

hence for all $\boldsymbol{t} \in \mathbb{R}^2$, $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$, $\mathcal{F}\left[f^2_{\underline{\boldsymbol{B}}}\right](\boldsymbol{t}\boldsymbol{x}^\top) = \mathcal{F}\left[f^1_{\underline{\boldsymbol{B}}}\right](\boldsymbol{t}\boldsymbol{x}^\top)$. Using Assumption 1, we have,

$$\mathbb{E}\left[e^{i\boldsymbol{t}^\top \underline{\boldsymbol{Y}}} | \boldsymbol{X} = \boldsymbol{x}\right] = \mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right]\left(\boldsymbol{t}\boldsymbol{x}^\top\right)$$

hence $f^1_{\underline{\boldsymbol{B}}}$ and $f^2_{\underline{\boldsymbol{B}}}$ yield the same observables, while being distinct *a.e.*, on $\mathcal{M}_{2,2}([0,1])$. Consider, for example, the coefficient $(1,1)$ of $\underline{\boldsymbol{B}}$. Then, using (4.43), we have, for all

$(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\underline{\boldsymbol{Y}}}$,

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}^1}\left[\boldsymbol{B}_{1,1}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right]-\mathbb{E}_{\mathbb{P}^2}\left[\boldsymbol{B}_{1,1}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right]\\
&=\int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})}\frac{\boldsymbol{b}_{1,1}}{\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}\left(f^1_{\underline{\boldsymbol{B}}}-f^2_{\underline{\boldsymbol{B}}}\right)(\boldsymbol{b})d\boldsymbol{b},\\
&=\frac{\gamma}{\mathcal{Z}\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}\int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})}\boldsymbol{b}_{1,1}\left(\boldsymbol{b}_{2,1}\boldsymbol{b}_{1,2}-\boldsymbol{b}_{1,1}\boldsymbol{b}_{2,2}\right)d\boldsymbol{b}\\
&=\frac{\gamma}{\mathcal{Z}\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}\left(\int_0^1 b\frac{\boldsymbol{y}_1-b\boldsymbol{x}_1}{\boldsymbol{x}_2}db\int_0^1\frac{\boldsymbol{y}_2-b\boldsymbol{x}_2}{\boldsymbol{x}_1}db-\int_0^1\left(\frac{\boldsymbol{y}_1-b\boldsymbol{x}_2}{\boldsymbol{x}_1}\right)^2db\int_0^1\frac{\boldsymbol{y}_2-b\boldsymbol{x}_1}{\boldsymbol{x}_2}db\right)\\
&=\frac{\gamma}{\mathcal{Z}\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})(\boldsymbol{x}_1\boldsymbol{x}_2)^2}\left(\boldsymbol{x}_1\boldsymbol{x}_2\left(\frac{\boldsymbol{y}_1}{2}-\frac{\boldsymbol{x}_1}{3}\right)\left(\boldsymbol{y}_2-\frac{\boldsymbol{x}_2}{2}\right)-\frac{1}{3}\left(\boldsymbol{y}_1^3-(\boldsymbol{y}_1-\boldsymbol{x}_2)^3\right)\left(\boldsymbol{y}_2-\frac{\boldsymbol{x}_1}{2}\right)\right)
\end{aligned}
$$

and using Assumption 2, there exists a subset $\mathcal{S}$ of $\mathbb{S}_{\boldsymbol{X},\underline{\boldsymbol{Y}}}$ with nonempty interior such that the right-hand-side is different from zero *a.e.* $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{S}$, which yields the result (P4.b). □

Lemma 1. *Let $\mathbb{P}_{\boldsymbol{B}}$ be a measure on $\mathcal{M}_{d_R,d_C}(\mathbb{R})$ satisfying (4.3)-(4.4). Then we have, for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$,*

$$
\int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b})=\mathcal{F}^{-1}\left[\mathcal{F}\left[\star\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\boldsymbol{x}^\top\right)\right](\underline{\boldsymbol{y}}),
$$

where the Fourier transform is defined in (4.42).

Proof of Lemma 1. First, using (4.43) we have, for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X},\boldsymbol{Y}}$,

$$
\mathbb{E}\left[\underline{\boldsymbol{B}}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\underline{\boldsymbol{y}}\right]\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\underline{\boldsymbol{y}}|\boldsymbol{x})=\int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b}), \tag{4.44}
$$

and that, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$, $\boldsymbol{y} \in \mathbb{R}^{d_C-1} \mapsto \mathbb{E}\left[\underline{\boldsymbol{B}}|\boldsymbol{X}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y}\right]\mathbb{P}_{\underline{\boldsymbol{Y}}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ is compactly supported in $[0,1]^{d_C-1}$. This yields that $\boldsymbol{y} \in \mathbb{R}^{d_C-1} \mapsto \int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b})$ belongs to $L^1(\mathbb{R}^{d_C-1})\cap L^2(\mathbb{R}^{d_C-1})$ hence its Fourier transform is well defined (see, *e.g.*, Theorem 9.13 in Rudin, 1973). Using the definition of $\mathcal{I}(\boldsymbol{x},\boldsymbol{y})$ for the second equality which yields that $\boldsymbol{b} \in \mathcal{I}(\boldsymbol{x},\boldsymbol{y})$ if and only if $\underline{\boldsymbol{y}} = (\boldsymbol{x}^\top\boldsymbol{b})^\top$ where $\boldsymbol{b} \in \mathcal{M}_{d_R\times(d_C-1)}([0,1])$, that $\boldsymbol{t}^\top(\boldsymbol{x}^\top\boldsymbol{b})^\top = \sum_{c=1}^{d_C-1}\boldsymbol{t}_c(\boldsymbol{x}^\top\boldsymbol{b})_c = \sum_{c=1}^{d_C-1}\boldsymbol{t}_c\sum_{r=1}^{d_R}\boldsymbol{x}_r\boldsymbol{b}_{r,c} = \sum_{c=1}^{d_C-1}\sum_{r=1}^{d_R}(\boldsymbol{t}_c\boldsymbol{x}_r)\boldsymbol{b}_{r,c} =< \boldsymbol{t}\boldsymbol{x}^\top,\boldsymbol{b} >$ for the third equality, and using the definition (4.42) of the Fourier transform, we have, for all $\boldsymbol{t} \in \mathbb{R}^{d_C-1}$,

$$\begin{aligned}\mathcal{F}\left[\int_{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},(\cdot,1-\cdot^\top\mathbf{1}))} \boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b})\right](\boldsymbol{t}) &= \int_{\mathbb{R}^{d_C-1}} e^{i\boldsymbol{t}^\top \underline{\boldsymbol{y}}} \int_{\mathcal{M}_{d_R\times(d_C-1)}([0,1])} \mathbb{1}\{\boldsymbol{b}\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})\}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b})d\underline{\boldsymbol{y}} \\ &= \int_{\mathcal{M}_{d_R\times(d_C-1)}([0,1])} e^{i\boldsymbol{t}^\top(\boldsymbol{x}^\top\boldsymbol{b})^\top}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) \\ &= \int_{\mathcal{M}_{d_R\times(d_C-1)}([0,1])} e^{i<\boldsymbol{t}\boldsymbol{x}^\top,\boldsymbol{b}>}\boldsymbol{b}d\mathbb{P}_{\underline{\boldsymbol{B}}}(\boldsymbol{b}) \\ &= \mathcal{F}\left[\star\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right](\boldsymbol{t}\boldsymbol{x}^\top).\end{aligned}$$

Then, we conclude using Theorem 9.13 d) in Rudin (1973) and taking the Fourier inverse. □

Proof of Proposition 7 and Theorem 1. Let me start with the proof of Proposition 7, then particularize the result to prove Theorem 1. Consider $\mathbb{P}_{\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y}}$ satisfying (4.1) and assumptions 1 and 2. (4.43) and Lemma 1 brings the identification to recovering, for $r = 1,\dots,d_R$ and $c = 1,\dots,d_C-1$, the function $\boldsymbol{t} \in \mathbb{R}^{d_C-1} \mapsto \mathcal{F}\left[\star_{r,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\boldsymbol{t}\boldsymbol{x}^\top\right)$, for all $\boldsymbol{x} \in \mathbb{S}_{\boldsymbol{X}}$. For all $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$, we use the notation $\dot{\boldsymbol{x}} := \left(\boldsymbol{x}^\top, 1-\boldsymbol{x}^\top\mathbf{1}\right)^\top \in \mathbb{S}_{\boldsymbol{X}}$.

Using Assumption 1, we have, for all $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$ and $\boldsymbol{t} \in \mathbb{R}^{d_C-1}$,

$$\varphi(\boldsymbol{x},\boldsymbol{t}) := \mathbb{E}\left[e^{i\boldsymbol{t}^\top\underline{\boldsymbol{Y}}}|\underline{\boldsymbol{X}}=\boldsymbol{x}\right] = \mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}}\right](\boldsymbol{t}\dot{\boldsymbol{x}}^\top). \tag{4.45}$$

Using the dominated convergence theorem, for all $c = 1,\dots,d_C-1$, $r = 1,\dots,d_R-1$, the function φ admits partial derivatives with respect to $\boldsymbol{t}_c$ and $\boldsymbol{x}_r$. Moreover, using that $\mathbb{S}_{\underline{\boldsymbol{X}}}$ has a nonempty interior, the latter derivatives are identified on $\mathbb{S}_{\underline{\boldsymbol{X}}}$, and we have, for all $\boldsymbol{t} \in \mathbb{R}^{d_C-1}$ and $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$,

$$\partial_{\boldsymbol{t}_c}\varphi(\boldsymbol{x},\boldsymbol{t}) = i\dot{\boldsymbol{x}}^\top\mathcal{F}\left[\star_{1:d_R,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\boldsymbol{t}\dot{\boldsymbol{x}}^\top\right), \tag{4.46}$$

$$\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\boldsymbol{t}) = i\boldsymbol{t}^\top\mathcal{F}\left[\star_{r,1:d_C-1}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\boldsymbol{t}\dot{\boldsymbol{x}}^\top\right) - i\boldsymbol{t}^\top\mathcal{F}\left[\star_{d_R,1:d_C-1}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\boldsymbol{t}\dot{\boldsymbol{x}}^\top\right). \tag{4.47}$$

This brings back identification to solving, for all $\boldsymbol{t} \in \mathbb{R}^{d_C-1}$, a system of $d_R \times (d_C-1)$ unknowns $\mathcal{F}\left[\star_{r,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\boldsymbol{t}\dot{\boldsymbol{x}}^\top\right)$, $r = 1\dots,d_R$, $c = 1,\dots,d_C-1$, and d_R+d_C-2

equations. Hence, $\mathbb{E}\left[\boldsymbol{B}|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\right]$ is identified under Assumption 2 when $d_C = 2$.

Using Assumption 2 and the dominated convergence theorem, for all $(\boldsymbol{t},\boldsymbol{x})\in\mathbb{R}^{d_C-1}\times\mathbb{S}_{\underline{\boldsymbol{X}}}$, we have

$$\partial_{\boldsymbol{t}_c}\varphi(\boldsymbol{x},\boldsymbol{t})=\int_{\mathbb{S}_{Y_{-C}}} i\boldsymbol{y}_c e^{i\boldsymbol{t}^\top\boldsymbol{y}} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}=i\mathcal{F}\left[\cdot f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\cdot|\boldsymbol{x})\right](\boldsymbol{t}).$$

Thus, we obtain, for all $\boldsymbol{y}\in\mathbb{S}_{\underline{\boldsymbol{Y}}}$,

$$\mathcal{F}^{-1}\left[\partial_{\boldsymbol{t}_c}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y})=i\boldsymbol{y}_c f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{x})=i\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}). \tag{4.48}$$

Using Assumption 3, which yields that $\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\in L^2(\mathbb{R})$ and (4.46)-(4.47), we obtain

$$\begin{aligned}\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) =&\dot{\boldsymbol{x}}^\top\mathcal{F}^{-1}\left[\mathcal{F}\left[\star_{1:d_R,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y}) \qquad (4.49)\\ \mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y}) =& i\mathcal{F}^{-1}\left[\cdot^\top\mathcal{F}\left[(\star_{r,1:d_C-1}-\star_{d_R,1:d_C-1})\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y}).\end{aligned}$$

Then, using that

$$\sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\mathcal{F}^{-1}\left[\mathcal{F}\left[\star_{r,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y})=-i\mathcal{F}^{-1}\left[\cdot^\top\mathcal{F}\left[\star_{r,1:d_C-1}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y}) \tag{4.50}$$

we obtain, for all $c=1,\dots,d_C-1$, $r=1,\dots,d_R-1$,

$$\begin{aligned}-\mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y}) =& \sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\mathcal{F}^{-1}\left[\mathcal{F}\left[\star_{r,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y})\\ &-\sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\mathcal{F}^{-1}\left[\mathcal{F}\left[\star_{d_R,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\star)\right]\left(\cdot\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y}).\end{aligned} \tag{4.51}$$

Denote by $\boldsymbol{M}_{r,c}:(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\mapsto\mathcal{F}^{-1}\left[\mathcal{F}\left[\bigstar_{r,c}\mathbb{P}_{\underline{\boldsymbol{B}}}(\bigstar)\right]\left(\boldsymbol{t}\dot{\boldsymbol{x}}^\top\right)\right](\boldsymbol{y})$, for $r=1,\dots,d_R$ and $c=1,\dots,d_C-1$, which are continuous functions which admit a continuous derivative with respect to $\boldsymbol{y}_c$. Moreover, from (4.47), we have $\boldsymbol{m}=\boldsymbol{M}/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and the constraint, for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$, $\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}_1,\dots,\boldsymbol{y}_c = 0,\boldsymbol{y}_{d_C-1}) = 0$ holds. Then, using (4.49), we obtain, for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$,

$$\partial_{\boldsymbol{y}_c}\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) = \sum_{r=1}^{d_R-1}\boldsymbol{x}_r\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}) + \partial_{\boldsymbol{y}_c}\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y}) - \sum_{r=1}^{d_R-1}\boldsymbol{x}_r\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y}) \tag{4.52}$$

and summing (4.51) over $r = 1,\dots,d_R-1$,

$$\begin{aligned}
-\sum_{r=1}^{d_R-1}\boldsymbol{x}_r\mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y}) &= \sum_{c=1}^{d_C-1}\left(\sum_{r=1}^{d_R-1}\boldsymbol{x}_r\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}) - \sum_{r=1}^{d_R-1}\boldsymbol{x}_r\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y})\right)\\
&= \sum_{c=1}^{d_C-1}\left(\partial_{\boldsymbol{y}_c}\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) - \partial_{\boldsymbol{y}_c}\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y})\right).
\end{aligned}$$

This yields

$$\sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{d_R,c}(\boldsymbol{x},\boldsymbol{y}) = \sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y}) + \sum_{r=1}^{d_R-1}\boldsymbol{x}_r\mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y}).$$

Then, using Assumption 3 and the dominated convergence theorem for the first equality, then Theorem 9.13 d) in Rudin (1973) for the second, we have

$$\begin{aligned}
\mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\varphi(\boldsymbol{x},\cdot)\right](\boldsymbol{y}) &= \mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_r}\mathcal{F}\left[f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\cdot|\boldsymbol{x})\right]\right](\boldsymbol{y})\\
&= \partial_{\boldsymbol{x}_r}f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{x}).
\end{aligned} \tag{4.53}$$

Using (4.51), we obtain (4.23). This yields that $\boldsymbol{m}$ takes the form described in the statement of Proposition 7.

When $d_C = 2$, integrating (4.23), using $\boldsymbol{M}_{r,c}(\boldsymbol{x},0) = 0$, and $\boldsymbol{\rho}_1(\boldsymbol{x},0) = 0$ for the first equality, and Assumption 3 and the dominated convergence theorem for the second one, we obtain, for all $r = 1, \dots, d_R$ and $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}}$,

$$\begin{aligned}\boldsymbol{M}_{r,1}(\boldsymbol{x}, y) &= \boldsymbol{\rho}_1(\boldsymbol{x}, y) + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \int_0^y \partial_{\boldsymbol{x}_l} f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(v|\boldsymbol{x}) dv \\ &= \boldsymbol{\rho}_1(\boldsymbol{x}, y) + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \partial_{\boldsymbol{x}_l} F_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(y|\boldsymbol{x}).\end{aligned}$$

Using $\boldsymbol{\rho}_1(\boldsymbol{x}, y) = y f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(y|\boldsymbol{x})$ yields the result of Theorem 1. $\square$

Proof of Proposition 8. Denote the right hand side of (4.23) by, for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{S}_{\boldsymbol{X}, \boldsymbol{Y}}$,

$$\boldsymbol{\Theta}_r(\boldsymbol{x}, \boldsymbol{y}) := \sum_{c=1}^{2} \partial_{\boldsymbol{y}_c} \boldsymbol{\rho}_c(\boldsymbol{x}, \boldsymbol{y}) + \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \partial_{\underline{\boldsymbol{x}}_l} f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}}). \tag{4.54}$$

Then, (4.23) can we rewritten as, for $r = 1, \dots, d_R - 1$,

$$\partial_{\boldsymbol{y}_1} \boldsymbol{V}_r(\boldsymbol{x}, \boldsymbol{y}) + \sum_{c=2}^{d_C-1} \partial_{\boldsymbol{y}_c} \boldsymbol{M}_{r,c}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Theta}_r(\boldsymbol{x}, \boldsymbol{y}). \tag{4.55}$$

Using (4.28), we have, for $r = 1, \dots, d_R - 1$

$$\boldsymbol{M}_{r,2}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{d_R-1} \widetilde{\boldsymbol{Q}}_{r,k}(\boldsymbol{x}) \boldsymbol{V}_k(\boldsymbol{x}, \boldsymbol{y}) + \sum_{k=d_R}^{d_R+1} \boldsymbol{a}_{r,k} \boldsymbol{\sigma}_{k-d_R+1}(\boldsymbol{x}, \boldsymbol{y}) \tag{4.56}$$

which yields the system of coupled partial differential equations, for $r = 1, \dots, d_R - 1$:

$$\partial_{\boldsymbol{y}_1} \boldsymbol{V}_r(\boldsymbol{x}, \boldsymbol{y}) + \sum_{k=1}^{d_R-1} \widetilde{\boldsymbol{Q}}_{r,k}(\boldsymbol{x}) \partial_{\boldsymbol{y}_2} \boldsymbol{V}_k(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Theta}_r(\boldsymbol{x}, \boldsymbol{y}), \tag{4.57}$$

with boundary constraints given by $\boldsymbol{V}_r(0, \boldsymbol{y}_2, \boldsymbol{x}) = 0$ for $r = 1, \dots, d_R - 1$. (4.57) is a system of coupled $(d_R - 1) \times (d_R - 1)$ transport partial differential equations that can be put into matrix form

$$\partial_{\boldsymbol{y}_1} \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{y}) + \widetilde{\boldsymbol{Q}}(\boldsymbol{x}) \partial_{\boldsymbol{y}_2} \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Theta}(\boldsymbol{x}, \boldsymbol{y}). \tag{4.58}$$

When $d_C = 3$, using assumption (11.2) yields in (4.58),

$$\partial_{\boldsymbol{y}_1} \widetilde{\boldsymbol{V}}(\boldsymbol{x},\boldsymbol{y}) + \mathrm{diag}\left(\boldsymbol{\Lambda}(\boldsymbol{x})\right) \partial_{\boldsymbol{y}_2} \widetilde{\boldsymbol{V}}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{P}\boldsymbol{\Theta}(\boldsymbol{x},\boldsymbol{y}),$$

where $\widetilde{\boldsymbol{V}} := \boldsymbol{P}\boldsymbol{V}$. Hence we can solve separately these $d_R - 1$ transport differential equations, for $r = 1, \ldots, d_R - 1$,

$$\widetilde{\boldsymbol{V}}_r(\boldsymbol{x},\boldsymbol{y}) = \sum_{k=1}^{d_R-1} \boldsymbol{P}_{r,k}(\boldsymbol{x}) \int_0^{\boldsymbol{y}_1} \boldsymbol{\Theta}_k(\boldsymbol{x}, v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_r(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv, \tag{4.59}$$

using that $\underline{\boldsymbol{V}}_r(0, \boldsymbol{y}_2, \boldsymbol{x}) = 0$ for $r = 1, \ldots, d_R - 1$. Thus, using (4.54), we obtain

$$\begin{aligned}
&\underline{\boldsymbol{V}}_r(\boldsymbol{x},\boldsymbol{y}) \\
&= \sum_{k=1}^{d_R-1} \boldsymbol{P}_{r,k}(\boldsymbol{x}) \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{y}_1} \boldsymbol{\rho}_1(\boldsymbol{x}, v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_r(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv \\
&\ + \sum_{k=1}^{d_R-1} \boldsymbol{P}_{r,k}(\boldsymbol{x}) \int_0^{\boldsymbol{y}_1} \partial_{\boldsymbol{y}_2} \boldsymbol{\rho}_2(\boldsymbol{x}, v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_r(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv \\
&\ + \sum_{k=1}^{d_R-1} \boldsymbol{P}_{r,k}(\boldsymbol{x}) \sum_{l=1}^{d_R-1} (\boldsymbol{x}_l - \mathbb{1}\{l = k\}) \int_0^{\boldsymbol{y}_1} \mathcal{F}^{-1}\left[\partial_{\boldsymbol{x}_l} \varphi(\underline{\boldsymbol{x}}, \cdot)\right](v, \boldsymbol{y}_2 - \boldsymbol{\Lambda}_r(\boldsymbol{x})(\boldsymbol{y}_1 - v))dv \\
&= \mathrm{Diag}(\boldsymbol{P}\boldsymbol{K}\boldsymbol{\zeta})_r(\boldsymbol{x},\boldsymbol{y}) \quad \text{(using (4.53))}.
\end{aligned}$$

This yields the result using (4.31). □

Details on the approximation of $\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot)$ using vaguelet-wavelets.

Let $y \in [0, 1]$. Let me explain how I use the vaguelet-wavelet decomposition to approximate $\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot)$. For any integer $J \in \mathbb{N}$ and assuming that $F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) \in L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$, we have the wavelet expansion

$$F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) = \sum_{j \geq J, \boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} d_{j,\boldsymbol{k},\boldsymbol{w}}(y) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}, \tag{4.60}$$

where $d_{j,\boldsymbol{k},\boldsymbol{w}}(y) := \left\langle \mathbb{E}\left[\mathbb{1}\{\underline{\boldsymbol{Y}} \leq y\} | \underline{\boldsymbol{X}} = \cdot\right], \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}} \right\rangle_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}$. Assume that $F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) \in L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$ admits a square integrable derivative with respect to the $l \in \{1, \ldots, d_R - 1\}$ variable.

Then, the vaguelet-wavelet decomposition of $\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot)$ (see, *e.g.*, Section 2.2 in Cai, 2002) is simply

$$\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) = \sum_{j \geq J, \boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} d_{j,\boldsymbol{k},\boldsymbol{w}}(y) 2^j \Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\cdot), \tag{4.61}$$

where $\left(\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}\right)_{l \geq J, \boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} = \left(\partial_l \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}/2^j\right)_{l \geq J, \boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j}$ are called the vaguelets and form a Riesz basis, which justifies this decomposition (see also the discussion in Appendix 4.5.4). Let $N \geq 2$ and J be fixed and $j_0 \geq J$ a parameter chosen a posteriori as a function of the sample size G. For all $l = 1, \ldots, d_R - 1$, and $y \in [0,1]$ consider the approximation of $\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$

$$\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) := \sum_{j=J}^{j_0} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} d_{j,\boldsymbol{k},\boldsymbol{w}}(y) 2^j \Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\cdot),$$

which yields my approximation of $\boldsymbol{m}_{r,1}$, for $r = 1, \ldots, d_R$ and $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$,

$$\boldsymbol{m}^{j_0}_{r,1}(\boldsymbol{x}, y) = y + \sum_{l=1}^{d_R - 1} (\boldsymbol{x}_l - \mathbb{1}\{l = r\}) \frac{\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}. \tag{4.62}$$

Formulation of the estimator with unknown $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$

Assumption 12 (On the rates of convergence of the preliminary estimators)**.** *Assume that:*

(*Est.1*) *We have estimators $\widehat{f_{\underline{\boldsymbol{X}}}}$ based on a preliminary sample $\mathcal{P}_{G_0} = (\boldsymbol{X}_g)^0_{g=-G_0+1}$ independent of $(\boldsymbol{X}_g, \boldsymbol{Y}_g)^G_{g=1}$ and $\widehat{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}$ based on a second preliminary sample $\mathcal{P}_{G_1} = (\boldsymbol{X}_g)^{-G_0}_{g=-(G_1+G_0)+1}$ independent of $(\boldsymbol{X}_g, \boldsymbol{Y}_g)^G_{g=-G_0}$;*

(*Est.2*) *$\mathcal{E}$ and $\mathcal{E}'$ are sets of densities and conditional densities on $\mathbb{S}_{\underline{\boldsymbol{X}}}$ and $\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$ such that, for $c_{\boldsymbol{X}}, c_{\boldsymbol{X},\boldsymbol{Y}} \in (0, \infty)$, for all $f_{\underline{\boldsymbol{X}}} \in \mathcal{E}$, $\|1/f_{\underline{\boldsymbol{X}}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq c_{\boldsymbol{X}}$, $\|f_{\underline{\boldsymbol{X}}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq C_{\boldsymbol{X}}$, and there exists a strict subset $\mathcal{S}$ of $\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$ such that, for all $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \in \mathcal{E}'$, $\|1/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\|_{L^\infty(\mathcal{S})} \leq c_{\boldsymbol{X},\boldsymbol{Y}}$; For $(v(G_0, \mathcal{E}))_{G_0 \in \mathbb{N}} \in (0,1)^{\mathbb{N}}$ and*

$(v(G_1,\mathcal{E}'))_{G_1\in\mathbb{N}} \in (0,1)^{\mathbb{N}}$ *which tend to 0, we have*

$$\frac{1}{v(G_0,\mathcal{E})}\sup_{f_{\underline{\boldsymbol{X}}}\in\mathcal{E}}\left\|\widehat{f}_{\underline{\boldsymbol{X}}}-f_{\underline{\boldsymbol{X}}}\right\|^2_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}=O_p\left(1\right), \tag{4.63}$$

$$\frac{1}{v(G_1,\mathcal{E}')}\sup_{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\in\mathcal{E}'}\left\|\widehat{f}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}-f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|^2_{L^\infty(\mathcal{S})}=O_p\left(1\right). \tag{4.64}$$

Giné and Nickl (2016); Tsybakov (2008) give examples of estimators for $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$, $\mathcal{E}$, $\mathcal{E}'$ rates (4.63) and (4.64). Define $\widehat{f}^{\delta}_{\underline{\boldsymbol{X}}} := \widehat{f}_{\underline{\boldsymbol{X}}} \vee \sqrt{\delta(G_0)}$ and $\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} := \widehat{f}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \vee \sqrt{\delta(G_1)}$, where δ is a trimming factor converging to zero. To deal with the statistical problem, I use

$$\widehat{\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}^{j_0}(\star|\cdot) := \sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(\star)2^j\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\cdot),$$

where I replace $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ by $\widehat{f}^{\delta}_{\underline{\boldsymbol{X}}}$ and $\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ in (4.14).

L^q **risk** In this context where $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ are estimated, I use the L^q risk on $\mathcal{S}$, which is defined in Assumption (Est.3), for $q\in\{2,\infty\}$, $r=1,\dots,d_R$

$$\mathcal{R}^q_{G_0,G_1}\left(\widehat{\boldsymbol{m}}_{r,1},\boldsymbol{m}_{r,1}\right) := \mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}_{r,1}-\boldsymbol{m}_{r,1}\right\|_{L^q(\mathcal{S})}\middle|\mathcal{P}_{G_0},\mathcal{P}_{G_1}\right]$$

and we use $G_e = G\wedge\lfloor(\delta(G_0)/v(G_0,\mathcal{E}))^{1/(1+\mathbb{1}\{q=\infty\})}\rfloor\wedge\lfloor(\delta(G_1)\delta(G_0)/v(G_1,\mathcal{E}'))^{1/(1+\mathbb{1}\{q=\infty\})}\rfloor$ for the sample size required for an ideal estimator where $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ are known to achieve the rate of the plug-in estimator. Instead of (4.10), the upper bounds of Theorem 2 in this context take the form, for $r=1,\dots,d_R-1$,

$$\frac{1}{r(G_e)}\sup_{\substack{f_{\boldsymbol{B}_{\cdot,1}}\in\mathcal{H}^{s+1}(l)\\ f_{\underline{\boldsymbol{X}}}\in\mathcal{E},\ f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\in\mathcal{E}'}}\mathcal{R}^q_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{j_0}_{r,1},\boldsymbol{m}_{r,1}\right)=O_p(1), \tag{4.65}$$

and in Theorem 2, G is replaced by G_e.

Different types of boundary corrections

I explain here how to obtain the orthonormal basis of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$,

$$\mathcal{B} = \left\{ \Phi_{J,\boldsymbol{k}}, \boldsymbol{k} \in \Lambda_J;\ \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}, j \in \mathbb{N} \setminus \{0, \dots, J-1\},\ \boldsymbol{k} \in \Lambda_j,\ \boldsymbol{w} \in \widetilde{\mathcal{W}} \right\}$$

based on the Daubechies family with $N \geq 1$ (see, *e.g.*, Härdle et al., 2012) and summarised its properties.

Case (RC.1), where $\mathbb{S}_{\underline{\boldsymbol{X}}}$ is a square. In this case, I use the boundary corrected wavelets introduced in Cohen et al. (1993) (see, *e.g.*, Section 4.3.5 in Giné and Nickl, 2016). Let $J, N \in \mathbb{N}$, $2^J \geq N$ and consider the standard $2^J - 2N$ Daubechies wavelets $\phi_{J,k} = 2^{J/2}\phi(2^J \cdot -k)$, $k \in \mathbb{Z}$ supported in the interior of $[0,1]$, the N left-edge basis functions $\phi^{\text{left}}_{J,k}$, and the right-edge basis functions $\phi^{\text{right}}_{J,k}$ introduced in Cohen et al. (1993) that are obtained from transformations (*e.g.* Gram-Schmidt orthonormalisation) of the standard wavelets. Together, they form an orthonormal system of $L^2([0,1])$ which we denote by $\left\{\phi^{bc}_{J,k}, k = 0, \dots, 2^J - 1\right\}$.
Cohen et al. (1993) then define the correspoding wavelets functions $\psi^{\text{left}}_{J,k}, \psi^{\text{right}}_{J,k}, \psi_{J,k}$ and introduced the dilated wavelets, for $k = 0, \dots, N-1$, $k' = N, \dots, 2N$ and $m = 2N+1, \dots, 2^j$,

$$\psi^{\text{left}}_{j,k} = 2^{(j-J)/2}\psi^{\text{left}}_{J,k}\left(2^{j-J}\cdot\right),\ \psi^{\text{right}}_{j,k'} = 2^{(j-J)/2}\psi^{\text{right}}_{J,k'}\left(2^{j-J}\cdot\right), \quad \psi_{j,m} = 2^{j/2}\psi\left(2^j \cdot -m + N\right),$$

which I denote using the common notation by $\psi^{bc}_{j,k}$, $k = 0, \dots, 2^j$. This yields

$$\left\{\psi^{bc}_{J,k}, \psi^{bc}_{j,l},\ k = 0, \dots, 2^{J-1},\ l = 0, \dots, 2^j - 1,\ j \geq J\right\}$$

forms a basis of $L^2([0,1])$. Theorem 4.4. in Cohen et al. (1993) and Theorem 4.2.10 (e) in Giné and Nickl (2016) shows that they belong to $C^\gamma([0,1])$ with $\gamma \geq 0.18(N-1)$.

Then using the construction of Section 4.3.6 in Giné and Nickl (2016), we introduce, for $\boldsymbol{k} \in \Lambda_j := \{\boldsymbol{k} : \ |\boldsymbol{k}|_\infty \le 2^J - 1\}$,

$$\widetilde{\Phi}_{J,\boldsymbol{k}} := \frac{1}{x_0^{(d_R-1)/2}} \prod_{r=1}^{d_R-1} \phi^{bc}_{J,\boldsymbol{k}_r}\left(\frac{\cdot_r - \widetilde{\boldsymbol{x}}_r}{x_0}\right), \quad \widetilde{\Psi}^{\boldsymbol{w}}_{j,\boldsymbol{k}_r} := \frac{1}{x_0^{(d_R-1)/2}} \prod_{r=1}^{d_R-1} (\psi^{bc}_{j,\boldsymbol{k}})^{\boldsymbol{w}_r}\left(\frac{\cdot_r - \widetilde{\boldsymbol{x}}_r}{x_0}\right),$$

where $\left(\psi^{bc}_{j,k}\right)^0 = 2^{(j-J)/2}\phi^{\text{left}}_{J,k}(2^{j-J}\cdot)$ for $k = 0, \dots, N-1$, $\left(\psi^{bc}_{j,k}\right)^0 = 2^{(j-J)/2}\phi^{\text{right}}_{J,k}(2^{j-J}\cdot)$ for $k = N, \dots, 2N-1$, and $\left(\psi^{bc}_{j,k}\right)^0 = 2^{j/2}\phi(2^j \cdot -k + N)$ for $k \ge 2N$. We thus have wavelet expansion (4.60).

The vaguelets $\left(\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}\right)$ form a Riesz basis (see Section 5 and condition (C) in Cai (2002)) which yields that there exist constants $A_l > a_l > 0$, which depend on $\mathbb{S}_{\underline{\boldsymbol{X}}}$, such that, for every sequence $(d_{j,\boldsymbol{k},\boldsymbol{w}})$,

$$a_l \left\|(d_{j,\boldsymbol{k},\boldsymbol{w}})\right\|_{l^2} \le \left\| \sum_{j \ge J} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \widetilde{\mathcal{W}}} d_{j,\boldsymbol{k},\boldsymbol{w}} \Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}} \right\|_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})} \le A_l \left\|(d_{j,\boldsymbol{k},\boldsymbol{w}})\right\|_{l^2}. \tag{4.66}$$

Important properties of $\mathcal{B}$. For simplicity of exposition, I summarize the following properties, where the constants can be adapted according to $\mathbb{S}_{\underline{\boldsymbol{X}}}$. Using that $(\psi^{bc}_{J,k})_{k \in \Lambda_J}$ belong to $C^{\lfloor \gamma \rfloor}([0,1])$ with $\gamma \ge 0.18(N-1)$ and the cardinal of Λ_J is finite, then for all J s.t. $2^J \ge 2N$ and all $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$,

$$\alpha_{1,J}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \Lambda_J} |\Phi_{J,\boldsymbol{k}}(\boldsymbol{x})| \le \kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) < \infty \tag{4.67}$$

and I denote by

$$\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) := \sup_{\boldsymbol{k} \in \Lambda_J} \|\partial_l \Phi_{J,\boldsymbol{k}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} < \infty. \tag{4.68}$$

Denote, for all $\boldsymbol{x} \in \boldsymbol{S}_{\underline{\boldsymbol{X}}}$ and $j \in N$, by $\alpha_{2,j}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}} |\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})|$. Then, using for the first inequality that

$$\sum_{\boldsymbol{k} \in \Lambda_j} \prod_{r=1}^{d_R-1} \left|\left(\psi^{bc}_{j,\boldsymbol{k}_r}\right)(\boldsymbol{x}_r)\right| \le \prod_{r=1}^{d_R-1} \sum_{w=0}^{1} \sum_{k=0}^{2^j-1} \left|\left(\psi^{bc}_{j,k}\right)(\boldsymbol{x}_r)\right|,$$

and Theorem 4.2.10 in Giné and Nickl (2016) for the last inequality, we obtain

$$\begin{aligned}
\alpha_{2,j}(\boldsymbol{x}) \leq & \prod_{r=1}^{d_R-1} \sum_{w=0}^{1} \left(\sum_{k=0}^{2N-1} \left| \left(\psi_{j,k}^{bc}\right)^{w} (\boldsymbol{x}_r) \right| + \sum_{k=2N}^{2^j-1} \left| \left(\psi_{j,k}^{bc}\right)^{w} (\boldsymbol{x}_r) \right| \right) \\
\leq & 2^{j(d_R-1)/2} \prod_{r=1}^{d_R-1} \sum_{w=0}^{1} \left(\sum_{k=0}^{N-1} \left| \left(\psi_{J,k}^{\text{left}}\right)^{w} \left(2^{j-J} \boldsymbol{x}_r\right) \right| + \sum_{k=N}^{2N-1} \left| \left(\psi_{J,k}^{\text{right}}\right)^{w} \left(2^{j-J} \boldsymbol{x}_r\right) \right| \right) \\
& + 2^{j(d_R-1)/2} \prod_{r=1}^{d_R-1} \sum_{w=0}^{1} \sum_{k=2N}^{2^j-1} \left| \psi^{w} \left(2^j \boldsymbol{x}_r - k + N\right) \right| \qquad (4.69) \\
\leq & 2^{j(d_R-1)/2} \left(\kappa_{2,1}^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) + \kappa_{2,2}^{N,d_R}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \right),
\end{aligned}$$

where $\kappa_{2,1}^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})$ and $\kappa_{2,2}^{N,d_R}(\mathbb{S}_{\underline{\boldsymbol{X}}})$ are finite, hence

$$\|\alpha_{2,j}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq 2^{j(d_R-1)/2} \kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) < \infty. \qquad (4.70)$$

Finally, using a similar decomposition as above and $\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}} = \partial_l \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}} / 2^j$, there exists $\nu_2^{J,N,d_R}(\mathbb{S}_{\underline{\boldsymbol{X}}}) < \infty$ such that for all $l = 1, \ldots, d_R - 1$, $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \mathcal{W}$, we have

$$\left\|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \leq 2^{j(d_R-1)/2} \nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}). \qquad (4.71)$$

When $N \geq 6$ we have $\phi, \psi \in C^1(\mathbb{R})$ by Theorem 4.2.10 in Giné and Nickl (2016) and by Proposition 4.2.5 in Giné and Nickl (2016), then using $\sum_{k \in \mathbb{Z}} \|\phi'(\cdot - k)\|_{L^\infty([0,1])} < \infty$ and $\sum_{k \in \mathbb{Z}} \|\psi'(\cdot - k)\|_{L^\infty([0,1])} < \infty$, we have, for all $l = 1, \ldots, d_R - 1$ and $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$,

$$\sum_{\boldsymbol{k} \in \Lambda_J} |\partial_l \Phi_{J,\boldsymbol{k}}(\boldsymbol{x})| \leq \widetilde{\kappa}_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) < \infty \qquad (4.72)$$

and, using a similar decomposition as (4.69),

$$\sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \widetilde{\mathcal{W}}} |\Omega_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})| \leq 2^{j(d_R-1)/2} \widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}). \qquad (4.73)$$

Proof of Section 4.2

The proofs use several times, for $j \geq 1$,

$$|\{\boldsymbol{k} \in \mathbb{N}_0^{d_R-1} : |\boldsymbol{k}|_\infty \leq 2^j - 1\}| \leq 2^{j(d_R-1)}, \tag{4.74}$$

$$\forall a, b > 0, \ \sup_{t \geq 1} \frac{\ln(t)^a}{t^b} = \left(\frac{a}{eb}\right)^a. \tag{4.75}$$

Proposition 9. *There exists a constant C_0 depending only on d_R such that for all $f \in L^2(\mathbb{R}^{d_R})$ compactly supported in $[-1,1]^{d_R}$ and with $s > (d_R - 1)/2$,*

$$\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \int_{\mathbb{R}} (1 \vee |t|)^{2s+(d_R-1)} \left|\mathcal{F}[f](t(\boldsymbol{x}, 1 - \boldsymbol{x}^\top \mathbf{1}))\right|^2 dt d\boldsymbol{x} \leq C_0 \int_{\mathbb{R}^{d_R}} (1 \vee |\boldsymbol{\xi}|_2)^{2s} |\mathcal{F}[f](\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.$$

Proof of Proposition 9. I borrow arguments from the proof of Theorem 4.6 in Hahn and Quinto (1985), without using the Radon transform. On the set $\mathbb{S}_{\underline{\boldsymbol{X}}} \times \mathbb{R} \setminus [-1,1]$, we use the bijective change of variable $F(t, \boldsymbol{x}) = (t\boldsymbol{x}_1, \ldots, t\boldsymbol{x}_{R-1}, t) = \boldsymbol{\xi} \in V$ with V a truncated cone in $\mathbb{R}^{d_R}$ and that for $|t| \geq 1$, $(1 \vee |t|)^{d_R-1} \leq 2^{(d_R-1)/2}|t|^{d_R-1}$ for the first equality

$$\begin{aligned}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \int_{\mathbb{R}} (1 \vee |t|)^{2s+(d_R-1)} |\mathcal{F}[f](t(\boldsymbol{x}, 1))|^2 dt d\boldsymbol{x} &\leq 2^{(d_R-1)/2} \int_V (1 \vee |\boldsymbol{\xi}_{d_R}|)^{2s} |\mathcal{F}[f](\boldsymbol{\xi})|^2 d\boldsymbol{\xi} \\ &\leq 2^{(d_R-1)/2} \int_{\mathbb{R}^{d_R}} (1 \vee |\boldsymbol{\xi}|)^{2s} |\mathcal{F}[f](\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.\end{aligned}$$

Then, for all $(\boldsymbol{x}, t) \in \mathbb{S}_{\underline{\boldsymbol{X}}} \times [-1,1]$, using the compact support of f,

$$\begin{aligned}&|\mathcal{F}[f](t(\boldsymbol{x}, 1))| \\ &= \left|\int_{\mathbb{R}^{d_R}} \mathbb{1}\{\boldsymbol{b} \in [-1,1]^{d_R}\} e^{i(t(\boldsymbol{x},1))\boldsymbol{b}} f(\boldsymbol{b}) d\boldsymbol{b}\right| \\ &\leq \left|\int_{\mathbb{R}^{d_R}} \mathcal{F}\left[\mathbb{1}\{\cdot \in [-1,1]^{d_R}\} e^{i(t(\boldsymbol{x},1))\cdot}\right](\boldsymbol{\xi}) \mathcal{F}[f](\boldsymbol{\xi}) d\boldsymbol{\xi}\right| \\ &\leq \int_{\mathbb{R}^{d_R}} \left|\mathcal{F}\left[\mathbb{1}\{\cdot \in [-1,1]^{d_R}\} e^{i(t(\boldsymbol{x},1))\cdot}\right](\boldsymbol{\xi})\right|^2 (1 \vee |\boldsymbol{\xi}|)^{-2s} d\boldsymbol{\xi} \int_{\mathbb{R}^{d_R}} (1 \vee |\boldsymbol{\xi}|)^{2s} |\mathcal{F}[f](\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.\end{aligned}$$

I conclude using that

$$\int_{\mathbb{R}^{d_R}} \left|\mathcal{F}\left[\mathbb{1}\{\cdot \in [-1,1]^{d_R}\} e^{i(t(\boldsymbol{x},1))\cdot}\right](\boldsymbol{\xi})\right|^2 (1 \vee |\boldsymbol{\xi}|)^{-2s} d\boldsymbol{\xi}$$
$$= \int_{\mathbb{R}^{d_R}} \frac{\prod_{i=1}^{d_R-1} |\mathrm{sinc}\,(\boldsymbol{\xi}_i + t\boldsymbol{x}_i)|^2 \left|\mathrm{sinc}\,\left(\boldsymbol{\xi}_{d_R} + t\right)\right|^2}{2^{-2d_R}(1 \vee |\boldsymbol{\xi}|)^{2s}} d\boldsymbol{\xi},$$

which is finite for $s > (d_R - 1)/2 \geq 1$ and that $\mathbb{S}_{\underline{\boldsymbol{X}}} \times [-1, 1]$ has finite measure. □

In the remaining, $\mathcal{E}$ and $\mathcal{E}'$ are classes of densities and conditional densities, $f_{\underline{\boldsymbol{X}}} \in \mathcal{E}$, $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \in \mathcal{E}'$, and $\eta, M > 0$. Denote also by $\Delta_{f,0} := 1/\widehat{f}^{\delta}_{\underline{\boldsymbol{X}}} - 1/f_{\underline{\boldsymbol{X}}}$, $\Delta_{f,1} := 1/\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - 1/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$,

$$Z_{G_0} := \sup_{f_{\underline{\boldsymbol{X}}} \in \mathcal{E}} \left\|\Delta_{f,0} f_{\underline{\boldsymbol{X}}}\right\|^2_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}, \quad Z_{G_1} := \sup_{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \in \mathcal{E}'} \left\|\Delta_{f,1} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|^2_{L^\infty(\mathbb{S}_{Y,\underline{\boldsymbol{X}}})}.$$

By Lemma A.3 in Gaillac and Gautier (2019c), there exists $M_{\mathcal{E},\eta,0}$ and $M_{\mathcal{E}',\eta,1}$ such that, for all $G_0, G_1 \in \mathbb{N}$, $\mathbb{P}\left(E\left(\mathcal{P}_{G_1}, \mathcal{E}', \eta\right)\right) \geq 1 - \eta/2$ and $\mathbb{P}\left(E\left(\mathcal{P}_{G_1}, \mathcal{E}, \eta\right)\right) \geq 1 - \eta/2$ where

$$E\left(\mathcal{P}_{G_0}, \mathcal{E}, \eta\right) := \left\{ Z_{G_0} \leq \frac{M_{\mathcal{E},\eta,0} v(G_0, \mathcal{E})}{\delta(G_0)} \right\}$$

and $E\left(\mathcal{P}_{G_1}, \mathcal{E}', \eta\right) := \{Z_{G_1} \leq M_{\mathcal{E}',\eta,1} v(G_1, \mathcal{E}')/\delta(G_1)\}$. I work on $E\left(\mathcal{P}_{G_0}, \mathcal{P}_{G_1}, \mathcal{E}, \mathcal{E}', \eta\right) := E\left(\mathcal{P}_{G_0}, \mathcal{E}, \eta\right) \cap E\left(\mathcal{P}_{G_1}, \mathcal{E}', \eta\right)$, where, using independence, $\mathbb{P}\left(E\left(\mathcal{P}_{G_0}, \mathcal{P}_{G_1}, \mathcal{E}, \mathcal{E}', \eta\right)\right) \geq 1 - \eta$, and use $M_{\mathcal{E},\mathcal{E}',\eta} := M_{\mathcal{E},\eta,0} \vee M_{\mathcal{E}',\eta,1}$.

All expectations are conditional on $\mathcal{P}_{G_0}$ and $\mathcal{P}_{G_1}$ when $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ are unknown and I rely on $\mathcal{P}_{G_0}$ and $\mathcal{P}_{G_1}$ to estimate it. I remove the conditioning in the notations for simplicity. Denote, for all $j \in \mathbb{N}$, $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \mathcal{W}_j$, by $\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}$ the quantities defined as in (4.14) replacing $\widehat{f}^{\delta}_{\underline{\boldsymbol{X}}}$ by $f_{\underline{\boldsymbol{X}}}$. Denote by $\widetilde{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ the estimator $\widehat{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ where $\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}$ is replaced by $\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}$. Denote also by $\widetilde{\boldsymbol{m}}^{j_0}$ the estimator $\widehat{\boldsymbol{m}}^{j_0}$ where $\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ is replaced by $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$.

Lemma 2. *For all* $J, j \in \mathbb{N}$, $j \geq J$, $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \mathcal{W}_j$, *and* $y \in [0, 1]$, *we have*

$\mathbb{E}\left[\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right] = d_{j,\boldsymbol{k},\boldsymbol{w}}(y)$, *and*

$$\mathbb{E}\left[\left|\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right] \le \frac{c_{\underline{\boldsymbol{X}}}}{G}.$$

Proof of Lemma 2. Let $J, j \in \mathbb{N}$, $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \mathcal{W}_j$, and $y \in [0,1]$. Using integration by part and that $\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}$ is compactly supported, yield

$$\mathbb{E}\left[\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right] = \mathbb{E}\left[\frac{\mathbb{1}\{\underline{\boldsymbol{Y}}_g \le y\}}{f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)}\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\left(\underline{\boldsymbol{X}}_g\right)\right] = \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \mathbb{E}\left[\mathbb{1}\{\underline{\boldsymbol{Y}} \le y\}|\underline{\boldsymbol{X}} = \boldsymbol{x}\right]\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})\,d\boldsymbol{x}$$

and, using that $\mathcal{B}$ is an orthonormal basis of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$, this yields

$$\mathbb{E}\left[\left|\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right] \le \frac{1}{G}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \frac{1}{f_{\underline{\boldsymbol{X}}}(\boldsymbol{x})}\left|\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})\right|^2 d\boldsymbol{x} \le \frac{c_{\underline{\boldsymbol{X}}}}{G}. \qquad \square \tag{4.76}$$

Proof of Theorem 2 when $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ are estimated as in Section 4.5.2. The proof of Theorem 2 has two parts: $q = \infty$ and $q = 2$. Let $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$ and use

$$R^{j_0}_{0,l} : (\boldsymbol{x}, y) \mapsto \left(\widehat{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \widetilde{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(\boldsymbol{x}, y) \tag{4.77}$$

$$R^{j_0}_{1,l} : (\boldsymbol{x}, y) \mapsto \left(\widetilde{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_{\boldsymbol{x}_l} F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(\boldsymbol{x}, y) \tag{4.78}$$

$$R^{j_0}_{2,l} : (\boldsymbol{x}, y) \mapsto \left(\partial_{\boldsymbol{x}_l} F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_{\boldsymbol{x}_l} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(\boldsymbol{x}, y). \tag{4.79}$$

Proof of Theorem (T2.1), L^∞ norm convergence rate. Let $N, J \in \mathbb{N}$, $2^J \ge 2N$. We obtain, using the triangle inequality for the first display and the notation (4.39) for the second one,

$$\begin{aligned}
&\left\|\widehat{\boldsymbol{m}}^{j_0}_r - \boldsymbol{m}_r\right\|_{L^\infty(\mathcal{S})} \\
&\le \left\|\widehat{\boldsymbol{m}}^{j_0}_r - \widetilde{\boldsymbol{g}}^{j_0}_r\right\|_{L^\infty(\mathcal{S})} + \left\|\widetilde{\boldsymbol{m}}^{j_0}_k - \boldsymbol{m}_k\right\|_{L^\infty(\mathcal{S})} \\
&\le \sum_{l=1}^{d_R-1} \|\boldsymbol{K}_{r,l}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}\left(Z^{1/2}_{G_1}\left\|\widehat{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} + \left\|\frac{\widetilde{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_{\boldsymbol{x}_l} F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}\right\|_{L^\infty(\mathcal{S})}\right).
\end{aligned} \tag{4.80}$$

Then, using (4.68) and (4.71) for the second inequality and using (4.67) and (4.70) for the thrid one, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left\|\widehat{\partial_{\boldsymbol{x}_l} F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right] \\
&\leq \sup_{y\in[0,1]} \sum_{\boldsymbol{k}\in\Lambda_J} \mathbb{E}\left[|\widehat{d}_{J,\boldsymbol{k},\underline{\boldsymbol{0}}}(y)|\right] \|\partial_l \Phi_{J,\boldsymbol{k}}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \\
&\quad + \sup_{y\in[0,1]} \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}} \mathbb{E}\left[\left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(\boldsymbol{t})\right|\right] 2^j \left\|\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \\
&\leq \frac{C_{\boldsymbol{X}}}{\sqrt{\delta(G_0)}} \sum_{\boldsymbol{k}\in\Lambda_J} \nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} |\Phi_{J,\boldsymbol{k}}(\boldsymbol{x})|\, d\boldsymbol{x} \\
&\quad + \frac{C_{\underline{\boldsymbol{X}}}}{\sqrt{\delta(G_0)}} \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}} 2^{j(d_R+1)/2} \nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})\right| d\boldsymbol{x} \\
&\leq \frac{|\mathbb{S}_{\underline{\boldsymbol{X}}}| C_{\underline{\boldsymbol{X}}}}{\sqrt{\delta(G_0)}} \left(\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) + \nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \sum_{j=J}^{j_0} 2^{j d_R} \right). \qquad (4.81)
\end{aligned}$$

Thus, using the triangle inequality and that $\sum_{j=J}^{j_0} 2^{(j-j_0)d_R} \leq 1/(1-2^{-d_R})$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}_r^{j_0} - \boldsymbol{m}_r\right\|_{L^\infty(\mathcal{S})}\right] \leq& \frac{C_0 Z_{G_1}^{1/2} 2^{j_0 d_R}}{\sqrt{\delta(G_0)}} \\
&+ \sum_{l=1}^{d_R-1} \|K_{r,l}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} c_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}} \sum_{j=0}^{2} \mathbb{E}\left[\left\|R^{j_0}_{j,l}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right], \qquad (4.82)
\end{aligned}$$

where

$$C_0 := C_{\underline{\boldsymbol{X}}} \left(\frac{\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{2^{J d_R}} + \frac{\nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{1-2^{-d_R}} \right) |\mathbb{S}_{\underline{\boldsymbol{X}}}| \sum_{l=1}^{d_R-1} \|K_{r,l}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}. \qquad (4.83)$$

Term $R_{0,l}$. We obtain, using (4.66) for the first display and using the same arguments as to obtain (4.81) for the last display as well as $\sum_{j=J}^{j_0} 2^{(j-j_0)d_R} \leq 1/(1-2^{-d_R})$, for all $l=1,\dots,d_R-1$,

$$\begin{aligned}
\mathbb{E}\left[\left\|R_{0,l}^{j_0}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right] &\leq \sup_{y\in[0,1]}\sum_{\boldsymbol{k}\in\Lambda_J}\mathbb{E}\left[\left|\widehat{c}_{J,\boldsymbol{k}}(y)-\widetilde{c}_{J,\boldsymbol{k}}(y)\right|\right]\left\|\Phi_{J,\boldsymbol{k}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \\
&\quad+\sup_{y\in[0,1]}\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\mathbb{E}\left[\left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)-\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|\right]\left\|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})} \\
&\leq Z_{G_0}^{1/2}\sum_{\boldsymbol{k}\in\Lambda_J}\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\mathbb{E}\left[\left|\Phi_{J,\boldsymbol{k}}\left(\underline{\boldsymbol{X}}_g\right)\right|\right] \\
&\quad+Z_{G_0}^{1/2}\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}2^{j(d_R+1)/2}\nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\mathbb{E}\left[\left|\Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\underline{\boldsymbol{X}}_g\right)\right|\right] \\
&\leq C_1 Z_{G_0}^{1/2}2^{j_0 d_R},
\end{aligned}\tag{4.84}$$

where

$$C_1:=2C_{\underline{\boldsymbol{X}}}|\mathbb{S}_{\underline{\boldsymbol{X}}}|\left(\frac{\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{2^{Jd_R}}+\frac{\nu_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{1-2^{-d_R}}\right).$$

Term $R_{1,l}$. We obtain, for all $l=1,\dots,d_R-1$,

$$\begin{aligned}
\mathbb{E}\left[\left\|R_{1,l}^{j_0}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right] \leq&\,\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_J}\left|\widetilde{c}_{J,\boldsymbol{k}}(y)-c_{J,\boldsymbol{k}}(y)\right|\right]\sum_{\boldsymbol{k}\in\Lambda_J}\left\|\partial_l\Phi_{J,\boldsymbol{k}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} \\
&+\sum_{j=J}^{j_0}\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\left|\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)-d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|\right]\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}2^j\left\|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}.
\end{aligned}$$

For all $y\in[0,1]$, we also have

$$\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_J}\left|\widetilde{c}_{J,\boldsymbol{k}}(y)-c_{J,\boldsymbol{k}}(y)\right|\right]=\frac{1}{G}\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_J}\left|\sum_{g=1}^{G}\left(f_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g)-\mathbb{E}\left[f_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g)\right]\right)\right|\right],$$

where $f_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g):=\mathbb{1}\{\underline{\boldsymbol{Y}}_g\leq y\}\Phi_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g)/f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)$. Using (4.70) we have, for all $\boldsymbol{k}\in\Lambda_J$,

$$\left\|f_{J,\boldsymbol{k}}-\mathbb{E}\left[f_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g)\right]\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\leq 2c_{\underline{\boldsymbol{X}}}\kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})$$

and using Lemma 2 that

$$\mathbb{E}\left[\left(f_{J,\boldsymbol{k}}(\underline{\boldsymbol{Y}}_g,\underline{\boldsymbol{X}}_g)\right)^2-\mathbb{E}\left[f_{J,\boldsymbol{k}}(\underline{\boldsymbol{Y}}_g,\underline{\boldsymbol{X}}_g)\right]^2\right]\leq\mathbb{E}\left[\left|\frac{\mathbb{1}\{\underline{\boldsymbol{Y}}_g\leq y\}\Phi_{J,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g)}{f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)}\right|^2\right]\leq c_{\underline{\boldsymbol{X}}}. \tag{4.85}$$

Then, Lemma 3.5.12 in Giné and Nickl (2016) yields

$$\begin{aligned}\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_J}|\widetilde{c}_{J,\boldsymbol{k}}(y)-c_{J,\boldsymbol{k}}(y)|\right]\leq&\frac{\sqrt{2Mc_{\underline{\boldsymbol{X}}}(J(d_R-1)+1)\ln(2)}}{G}\\&+\frac{2c_{\underline{\boldsymbol{X}}}\kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})(J(d_R-1)+1)}{3M}\ln(2).\end{aligned}$$

Similarly, using that $|\widetilde{\mathcal{W}}|=2^{d_R-1}$, we have, for all $j\geq J$

$$\begin{aligned}\mathbb{E}\left[\max_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\left|\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)-d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|\right]\leq&\frac{1}{G}\sqrt{2Mc_{\underline{\boldsymbol{X}}}((j+1)(d_R-1)+1)\ln(2)}\\&+\frac{c_{\underline{\boldsymbol{X}}}2^{j(d_R-1)/2+1}\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})((j+1)(d_R-1)+1)}{3M}\ln(2).\end{aligned}$$

Using (4.72), (4.73), and suppressing $\mathbb{S}_{\underline{\boldsymbol{X}}}$ in the notations, we obtain,

$$\begin{aligned}&\mathbb{E}\left[\left\|R_{1,l}^{j_0}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right]\\&\leq\sqrt{\frac{2c_{\underline{\boldsymbol{X}}}(J(d_R-1)+1)}{G}\ln(2)}\widetilde{\kappa}_1^{N,d_R,J}+2\frac{c_{\underline{\boldsymbol{X}}}(J(d_R-1)+1)\kappa_1^{N,d_R,J}\widetilde{\kappa}_1^{N,d_R,J}}{3M}\ln(2)\\&\quad+\sum_{j=J}^{j_0}\sqrt{\frac{2^{j(d_R+1)+1}c_{\underline{\boldsymbol{X}}}((j+1)(d_R-1)+1)}{G}\ln(2)}\widetilde{\kappa}_2^{N,d_R,J}\\&\quad+\sum_{j=J}^{j_0}\frac{c_{\underline{\boldsymbol{X}}}2^{jd_R+1}((j+1)(d_R-1)+1)\kappa_2^{N,d_R,J}\widetilde{\kappa}_2^{N,d_R,J}}{3M}\ln(2)\\&\leq(j_0-J)\sqrt{\frac{2^{j_0(d_R+1)+1}c_{\underline{\boldsymbol{X}}}((j_0+1)(d_R-1)+1)}{G}\ln(2)}\left(\widetilde{\kappa}_1^{N,d_R,J}+\widetilde{\kappa}_2^{N,d_R,J}\right)\\&\quad+(j_0-J)\frac{c_{\underline{\boldsymbol{X}}}2^{j_0d_R+1}((j_0+1)(d_R-1)+1)}{3M}\ln(2)\left(\kappa_1^{N,d_R,J}\widetilde{\kappa}_1^{N,d_R,J}+\kappa_2^{N,d_R,J}\widetilde{\kappa}_2^{N,d_R,J}\right).\end{aligned}\tag{4.86}$$

Term $R_{2,l}$. We have

$$\left\|R_{2,l}^{j_0}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} \le \sup_{y\in[0,1]} \sum_{j=j_0+1}^{\infty} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}} |d_{j,\boldsymbol{k},\boldsymbol{w}}(y)| 2^j \left\|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}.$$

Let $T_j > 0$, for all $j \ge J$. For all $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \widetilde{\mathcal{W}}$, denote by

$$H_1(j,\boldsymbol{k}) := \int_{|t|>T_j} \left| \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right|^2 dt$$

$$H_2(j,\boldsymbol{k}) := \int_{|t|\le T_j} \left| \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right|^2 dt.$$

Using for the second equality that under Assumption 3,

$$F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) = \int_0^y \mathcal{F}^{-1}\left[\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](\cdot(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\right](v)dv, \tag{4.87}$$

and using the Cauchy-Schwarz inequality for the third inequality, we obtain

$$\begin{aligned}
&|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)| \\
&= \left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right| \\
&= \frac{1}{2\pi}\left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\int_0^y\int_{\mathbb{R}} e^{-itv}\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}dvdt\right| \\
&\le \frac{y}{2\pi}\int_{\mathbb{R}}\left|\operatorname{sinc}\left(\frac{ty}{2}\right)\right|\left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right|dt \\
&\le \frac{y}{2\pi}\left(\int_{\mathbb{R}}\left|\operatorname{sinc}\left(\frac{ty}{2}\right)\right|^2 dt\right)^{1/2}\left(\int_{\mathbb{R}}\left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right|^2 dt\right)^{1/2} \\
&\le \frac{\sqrt{2y}}{\pi}\left(\int_{\mathbb{R}}\left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x}, 1-\boldsymbol{x}^\top \mathbf{1}))\overline{\Psi}_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})d\boldsymbol{x}\right|^2 dt\right)^{1/2} \\
&\le \frac{\sqrt{2y}}{\pi}\left(H_1(j,\boldsymbol{k}) + H_2(j,\boldsymbol{k})\right)^{1/2}
\end{aligned} \tag{4.88}$$

Using the Cauchy-Schwarz inequality and that $\left(\Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\right)_{j\ge J,\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}$ are orthonormal on $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$ for the first inequality and Assumption 6 and Proposition 9 for the second one, we have

$$H_1(j,\boldsymbol{k}) \leq \int_{|t|>T_j}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x},1-\boldsymbol{x}^\top\mathbf{1}))\right|^2 d\boldsymbol{x}dt \leq \frac{l^2C_0}{T_j^{2(s+1)+d_R-1}}. \tag{4.89}$$

Then, using for the fourth equality that $\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}$ has compact support in $\mathbb{S}_{\underline{\boldsymbol{X}}}$, we have

$$\begin{aligned}\left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \mathcal{F}\left[f_{\underline{\boldsymbol{B}}}\right](t(\boldsymbol{x},1-\boldsymbol{x}^\top\mathbf{1}))\overline{\Psi}^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x}\right| &= \left|\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\int_{[0,1]^{d_R}} e^{it(\boldsymbol{x},1-\boldsymbol{x}^\top\mathbf{1})^\top\boldsymbol{b}} f_{\underline{\boldsymbol{B}}}(\boldsymbol{b})\overline{\Psi}^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{b}d\boldsymbol{x}\right| \\ &= \left|\int_{[0,1]^{d_R}} f_{\underline{\boldsymbol{B}}}(\boldsymbol{b})\overline{e^{-it\boldsymbol{b}_{d_R}}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} e^{-it\boldsymbol{x}^\top(\underline{\boldsymbol{b}}-\boldsymbol{b}_{d_R})}\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x}}d\boldsymbol{b}\right| \\ &= 2\pi\left|\int_{[0,1]^{d_R}} f_{\underline{\boldsymbol{B}}}(\boldsymbol{b})\overline{e^{-it\boldsymbol{b}_{d_R}}\mathcal{F}^{-1}\left[\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\right](t(\underline{\boldsymbol{b}}-\boldsymbol{b}_{d_R}))}d\boldsymbol{b}\right| \\ &\leq 2\pi\sup_{\boldsymbol{b}\in[0,1]^{d_R}}\left|\mathcal{F}^{-1}\left[\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\right](t(\underline{\boldsymbol{b}}-\boldsymbol{b}_{d_R}))\right|.\end{aligned} \tag{4.90}$$

Using that, for all $\boldsymbol{k}\in\Lambda_j$ and $\boldsymbol{w}\in\widetilde{\mathcal{W}}$,

$$\left|\mathcal{F}^{-1}\left[\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\right](\cdot)\right| = \frac{1}{2^{j(d_R-1)/2}x_0^{(d_R-1)/2}}\prod_{r=1}^{d_R-1}\left|\mathcal{F}^{-1}\left[\psi^{\boldsymbol{w}_r}\right]\left(\frac{x_0\cdot}{2^j}\right)\right|,$$

where $x_0=1$ in the case of (RC.2), that $\mathcal{F}^{-1}[\psi](\cdot) = |m_0(\cdot+\pi)||\mathcal{F}^{-1}[\phi](\cdot)|$, that from (4.68) and (4.70) in Giné and Nickl (2016) there exists $C_2>0$ such that $|\mathcal{F}^{-1}[\psi](u)|\leq C_2|u|^N$ and $|\mathcal{F}^{-1}[\phi](u)|\leq C_2$ for all $u\in[-1,1]$, we obtain, for all $x_0|t|/2^j<1$ and $\boldsymbol{b}\in[0,1]^{d_R}$,

$$|\mathcal{F}^{-1}\left[\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\right](t(\boldsymbol{b}_{-d_R}-\boldsymbol{b}_{d_R}))| \leq \frac{C_2^{d_R-1}}{2^{j(d_R-1)/2}}\left(\frac{x_0|t|}{2^j}\right)^{N|\boldsymbol{w}|_1}. \tag{4.91}$$

We have, for all $\boldsymbol{w} \in \widetilde{\mathcal{W}}$,

$$H_2(j, \boldsymbol{k}) \leq 4\pi^2 \int_{|t|\leq T_j} \frac{C_2^{2(d_R-1)}}{2^{j(d_R-1)}} \left(\frac{x_0|t|}{2^j}\right)^{2N} dt$$
$$\leq \frac{8\pi^2 C_2^{2(d_R-1)}}{2^{j(d_R-1)}(2N+1)} \left(\frac{x_0 T_j}{2^j}\right)^{2N} T_j \tag{4.92}$$

Thus, using (4.88), (4.89), (4.92), and taking $T_j = 2^{j(2N+d_R-1)/(2N+2(s+1)+d_R)}$, we obtain, for all $y \in [0, 1]$,

$$|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)| \leq \frac{\sqrt{2}}{\pi} \left(\frac{8\pi^2 C_2^{2(d_R-1)}}{2^{j(d_R-1)}(2N+1)} \left(\frac{x_0}{2^j}\right)^{2N} T_j^{2N+1} + \frac{l^2 C_0}{T_j^{2(s+1)+d_R-1}} \right)^{1/2}$$
$$\leq \frac{\sqrt{2}}{\pi} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0 \right)^{1/2} \frac{1}{2^{j(d_R+1)/2} 2^{j s_N}}. \tag{4.93}$$

Then, using (4.73) for the first inequality, $N > 1 + (d_R+1)/s$ for the second and third inequalities, this yields

$$\left\|R_{2,l}^{j_0}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} \leq \frac{\sqrt{2}}{\pi} \sum_{j=j_0+1}^{\infty} \sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\widetilde{\mathcal{W}}} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0 \right)^{1/2} \frac{\left\|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}}{2^{j(d_R-1)/2} 2^{j s_N}}$$
$$\leq \frac{\sqrt{2}}{\pi} \widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0 \right)^{1/2} \sum_{j=j_0+1}^{\infty} \frac{1}{2^{j s_N}}$$
$$\leq \frac{\sqrt{2}}{\pi} \frac{\widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{\ln(2) s_N} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0 \right)^{1/2} \frac{1}{2^{(j_0+1)s_N}}. \tag{4.94}$$

Finally, using

$$C_{2,r} := c_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}} \sum_{l=1}^{d_R-1} \|K_{r,l}\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}, \qquad C_3 := \left(\kappa_1^{N,d_R,J} \widetilde{\kappa}_1^{N,d_R,J} + \kappa_2^{N,d_R,J} \widetilde{\kappa}_2^{N,d_R,J} \right) \sqrt{2 c_{\underline{\boldsymbol{X}}} \ln(2)},$$

$$C_4 := \frac{2 c_{\underline{\boldsymbol{X}}} \left(\widetilde{\kappa}_1^{N,d_R,J} + \widetilde{\kappa}_2^{N,d_R,J} \right) \ln(2)}{3}, \quad C_5 := \frac{\sqrt{2}}{\pi} \frac{\widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{\ln(2) s_N 2^{s_N}} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0 \right)^{1/2},$$

for the first inequality, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}_{r,1}^{j_0}-\boldsymbol{m}_{r,1}\right\|_{L^\infty(\mathcal{S})}\right] \leq & \frac{C_0 Z_{G_1}^{1/2} 2^{j_0 d_R}}{\sqrt{\delta(G_0)}} + C_{2,r} C_1 Z_{G_0}^{1/2} 2^{j_0 d_R} \\
& + C_{2,r}\left(C_3 (j_0 - J)\sqrt{\frac{2^{j_0(d_R+1)}((j_0+1)(d_R+1)+1)}{G}}\right) \\
& + C_{2,r}\left(C_4 \frac{2^{j_0 d_R}(j_0-J)((j_0+1)(d_R-1)+1)}{G} + \frac{C_5}{2^{j_0 s_N}}\right) \\
\leq & C_0 M_{\mathcal{E}',\eta,1}^{1/2} \frac{v(G_1,\mathcal{E}')^{1/2} 2^{j_0 d_R}}{\sqrt{\delta(G_1)\delta(G_0)}} + C_{2,r} C_1 M_{\mathcal{E},\eta,0}^{1/2} \frac{v(G_0,\mathcal{E})^{1/2} 2^{j_0 d_R}}{\sqrt{\delta(G_0)}} \\
& + C_{2,r}\left(C_3 j_0 \sqrt{\frac{2^{j_0(d_R+1)}(j_0+1)d_R}{G_e}} + C_4 R \frac{2^{j_0 d_R} j_0^2}{G_e} + \frac{C_5}{2^{j_0 s_N}}\right).
\end{aligned}
$$

Using $3\ln(\widetilde{j}) + (2s_N + d_R + 1)\widetilde{j}\ln(2) = \ln(G_e)$, we have

$$
\widetilde{j} \leq \frac{\ln(G_e)}{(2s_N + d_R + 1)\ln(2)},
$$

and using $j_0 \geq \widetilde{j} - 1$ yields

$$
\begin{aligned}
2^{-j_0 s_N} \leq \frac{2}{2^{\widetilde{j} s_N}} \leq 2\left(\frac{\widetilde{j}^3}{G_e}\right)^{s_N/(2s_N+d_R+1)} & \\
\leq \frac{2}{((2s_N + d_R + 1)\ln(2))^{3s_N/(2s_N+d_R+1)}} \left(\frac{\ln(G_e)^3}{G_e}\right)^{s_N/(2s_N+d_R+1)} &
\end{aligned}
$$

and

$$
\sqrt{\frac{2^{j_0(d_R+1)} j_0^3 d_R}{G_e}} \leq \sqrt{\frac{2^{\widetilde{j}(d_R+1)} \widetilde{j}^3 d_R}{G_e}} = \frac{\sqrt{d_R}}{2^{\widetilde{j} s_N}}.
$$

Finally, using $G_e v(G_1,\mathcal{E}')^{1/2}/\sqrt{\delta(G_1)\delta(G_0)} \leq 1$, we have

$$
\frac{v(G_1,\mathcal{E}')^{1/2} 2^{j_0 d_R} j_0}{\sqrt{\delta(G_1)\delta(G_0)}} \leq \frac{2^{\widetilde{j} d_R}\widetilde{j}}{G_e} \leq \frac{1}{\widetilde{j}^2 2^{\widetilde{j}(2s_N+1)}}
$$

and, using $G_e v(G_0,\mathcal{E})^{1/2}/\sqrt{\delta(G_0)} \le 1$,

$$\frac{v(G_0,\mathcal{E})^{1/2}2^{j_0 d_R} j_0}{\sqrt{\delta(G_0)}} \le \frac{2^{\widetilde{j}d_R}\widetilde{j}}{G_e} \le \frac{1}{\widetilde{j}^2 2^{\widetilde{j}(2s_N+1)}}$$

which yield the result.

Proof of Theorem (T2.2), L^2 norm convergence rates. Using (4.80) and the convexity of $x \mapsto x^2$, we obtain

$$\begin{aligned}
&\left\|\widehat{\boldsymbol{m}}_r^{j_0} - \boldsymbol{m}_r\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\\
&\le (d_R-1)\sum_{l=1}^{d_R-1}|\boldsymbol{K}_{r,l}|^2_\infty\left(Z_{G_1}\left\|\widehat{\partial_{\boldsymbol{x}_l}F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} + c^2_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\left\|\widehat{\partial_{\boldsymbol{x}_l}F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_{\boldsymbol{x}_l}F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right).
\end{aligned}$$

Then, using the convexity of $x \mapsto x^2$, the Cauchy-Swarz inequality, and (4.66) for the first display, that $(\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}})$ is an orthonormal system of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$, (4.67), (4.70), and (4.74) for the second inequality, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left\|\widehat{\partial_{\boldsymbol{x}_l}F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right]\\
&\le 2\mathbb{E}\left[\left\|\sum_{\boldsymbol{k}\in\Lambda_J}|\widehat{c}_{J,\boldsymbol{k}}(y)|\partial_l\Phi_{J,\boldsymbol{k}}\right\|_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}\right] + 2A_l\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\mathbb{E}\left[\left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right]2^{2j}\\
&\le 2\frac{\left(\kappa_1^{N,d_R,J}\widetilde{\kappa}_1^{N,d_R,J}\right)^2|\mathbb{S}_{\underline{\boldsymbol{X}}}|C_{\underline{\boldsymbol{X}}}}{\delta(G_0)} + 2\frac{A_l|\widetilde{\mathcal{W}}|C_{\underline{\boldsymbol{X}}}}{\delta(G_0)}\sum_{j=J}^{j_0}2^{j(d_R+1)}.
\end{aligned} \tag{4.95}$$

Using $C_{7,r,l} := 3(d_R-1)c^2_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\|K_{r,l}\|^2_{L^\infty}$, (4.95), that $\sum_{j=J}^{j_0}2^{(j-j_0)(d_R+1)} \le 1/(1-2^{-(d_R+1)})$, and the convexity of $x \mapsto x^2$ yield

$$\left\|\widehat{\boldsymbol{m}}_r^{j_0} - \boldsymbol{m}_r\right\|^2_{L^2(\mathcal{S})} \le \widetilde{C}_0\frac{Z_{G_1}2^{j_0(d_R+1)}}{\delta(G_0)} + \sum_{l=1}^{d_R-1}C_{7,r,l}\sum_{j=0}^{2}\int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}\left|R^{j_0}_{j,l}(\boldsymbol{x},y)\right|^2 dy d\boldsymbol{x}, \tag{4.96}$$

where

$$\widetilde{C}_0 := 2(d_R-1)C_{\underline{\boldsymbol{X}}}\left(\frac{(\kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\widetilde{\kappa}_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}))^2|\mathbb{S}_{\underline{\boldsymbol{X}}}|}{2^{J(d_R+1)}} + \frac{A_l|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}}\right)\sum_{l=1}^{d_R-1}\|K_{r,l}\|^2_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}.$$

Term $R_{0,l}$**.** Using the Cauchy-Schwarz inequality and (4.66) for the first display, that $(\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}})_{j\geq J,\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}$ is an orthonormal system of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$, (4.70), and (4.74) for the second, and using $\sum_{j=J}^{j_0} 2^{(j-j_0)(d_R+1)} \leq 1/(1-2^{-(d_R+1)})$ for the last display, we obtain, for all $l=1,\dots,d_R-1$,

$$
\begin{aligned}
&\mathbb{E}\left[\|R_{0,l}^{j_0}\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right] &(4.97)\\
&\leq 2\sup_{y\in[0,1]}\left(|\Lambda_J|\sup_{\boldsymbol{k}\in\Lambda_J}\mathbb{E}\left[|\widehat{c}_{J,\boldsymbol{k}}(y)-\widetilde{c}_{J,\boldsymbol{k}}(y)|\right]^2 dt + A_l\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\mathbb{E}\left[\left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)-\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right]2^{2j}\right)\\
&\leq 2Z_{G_0}\left(\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\right)^2|\Lambda_J| + 2A_lZ_{G_0}|\widetilde{\mathcal{W}}|C_{\underline{\boldsymbol{X}}}\sum_{j=J}^{j_0}2^{j(d_R+1)}\\
&\leq \widetilde{C}_1 Z_{G_0}2^{j_0(d_R+1)}, &(4.98)
\end{aligned}
$$

where $\widetilde{C}_1 := 2((\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}))^2/2^{J(d_R+1)} + \overline{A}|\widetilde{\mathcal{W}}|C_{\underline{\boldsymbol{X}}}/(1-2^{-(d_R+1)}))$.

Term $R_{1,l}$**.** We obtain, for all $l=1,\dots,d_R-1$, using the Cauchy-Schwarz inequality, the convexity of $x\mapsto x^2$, and (4.66) for the second display, Lemma 2, (4.74), and (4.68) for the third display,

$$
\begin{aligned}
&\int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}\mathbb{E}\left[\left|R_{1,l}^{j_0}(\boldsymbol{x},y)\right|^2\right]dyd\boldsymbol{x}\\
&\leq 2\int_{[0,1]}\mathbb{E}\left[\left\|\sum_{\boldsymbol{k}\in\Lambda_J}|\widetilde{c}_{J,\boldsymbol{k}}(y)-c_{J,\boldsymbol{k}}(y)|^2\,\partial_l\Phi_{J,\boldsymbol{k}}\right\|_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}\right]\\
&\quad + A_l\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}}\mathbb{E}\left[\left|\widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)-d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right]2^{2j}dy\\
&\leq 2\frac{c_{\underline{\boldsymbol{X}}}}{G}\left(|\Lambda_J|\nu_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}) + A_l|\widetilde{\mathcal{W}}|\sum_{j=J}^{j_0}2^{j(d_R+1)}\right)
\end{aligned}
$$

Thus, using $\sum_{j=J}^{j_0}2^{(j-j_0)(d_R+1)}\leq 1/(1-2^{-(d_R+1)})$, we obtain

$$
\int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}\mathbb{E}\left[\left|R_{1,l}^{j_0}(\boldsymbol{x},y)\right|^2\right]dyd\boldsymbol{x} \leq \frac{2c_{\underline{\boldsymbol{X}}}2^{j_0(d_R+1)}}{G}\left(\frac{|\Lambda_J|\nu_1^{N,d_R,J}}{2^{J(d_R+1)}}+\frac{A_l|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}}\right). \quad (4.99)
$$

Term $R_{2,l}$**.** We obtain, using (4.66) for the second display, (4.93) for the third, (4.74) and $|\widetilde{\mathcal{W}}| = 2^{d_R-1} - 1$ for the fifth,

$$\begin{aligned}
&\int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left|R_{2,l}^{j_0}(\boldsymbol{x}, y)\right|^2 dy d\boldsymbol{x} \\
&\leq \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left|\sum_{j=j_0+1}^{\infty} \sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\widetilde{\mathcal{W}}} d_{j,\boldsymbol{k},\boldsymbol{w}}(y) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})\right|^2 d\boldsymbol{x} dy \\
&\leq 2A_l \int_{[0,1]} \sum_{j=j_0+1}^{\infty} \sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\widetilde{\mathcal{W}}} \left|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 2^{2j} dy \\
&\leq A_l \int_{[0,1]} \frac{4}{\pi^2} \sum_{j=j_0+1}^{\infty} \sum_{\boldsymbol{w}\in\widetilde{\mathcal{W}}} \sum_{\boldsymbol{k}\in\Lambda_j} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0\right) \frac{1}{2^{j(d_R-1)} 2^{2js_N}} dy \\
&\leq \frac{4A_l(2^{d_R-1}-1)}{\pi^2} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0\right) \sum_{j=j_0+1}^{\infty} \frac{1}{2^{2js_N}} \\
&\leq \frac{4A_l(2^{d_R-1}-1)}{\pi^2 \ln(2) s_N} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0\right) \frac{1}{2^{2(j_0+1)s_N}}. \qquad (4.100)
\end{aligned}$$

Hence, using

$$\begin{aligned}
C_{8,r} &:= \sum_{l=1}^{d_R-1} C_{7,r,l}, \quad \widetilde{C}_3 := 2c_{\underline{\boldsymbol{X}}} \left(\frac{|\Lambda_J| \nu_1^{N,d_R,J}}{2^{J(d_R+1)}} + \frac{\overline{A}|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}}\right) \\
\widetilde{C}_4 &:= \frac{4\overline{A}(2^{d_R-1}-1)}{\pi^2 \ln(2) s_N 2^{2s_N}} \left(\frac{8\pi^2 C_2^{2(d_R-1)} x_0^{2N}}{2N+1} + l^2 C_0\right),
\end{aligned}$$

we obtain,

$$\begin{aligned}
\mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}_{r,1}^{j_0} - \boldsymbol{m}_{r,1}\right\|_{L^2(\mathcal{S})}^2\right] &\leq \widetilde{C}_0 \frac{Z_{G_1} 2^{j_0(d_R+1)}}{\delta(G_0)} + C_{8,r}\left(\widetilde{C}_1 Z_{G_0} 2^{j_0(d_R+1)} + \widetilde{C}_3 \frac{2^{j_0(d_R+1)}}{G} + \frac{\widetilde{C}_4}{2^{2s_N j_0}}\right) \\
&\leq \widetilde{C}_0 M_{\mathcal{E}',\eta,1} \frac{v(G_1, \mathcal{E}') 2^{j_0(d_R+1)}}{\delta(G_0)\delta(G_1)} \\
&\quad + C_{8,r}\left(\widetilde{C}_1 M_{\mathcal{E},\eta,0} \frac{v(G_0, \mathcal{E}) 2^{j_0(d_R+1)}}{\delta(G_0)} + \widetilde{C}_3 \frac{2^{j_0(d_R+1)}}{G} + \frac{\widetilde{C}_4}{2^{2s_N j_0}}\right).
\end{aligned}$$

Using $(2s_N + d_R + 1)\widetilde{j}\ln(2) = \ln(G_e)$ and $j_0 \geq \widetilde{j} - 1$ which yields

$$2^{-2j_0 s_N} \leq \frac{4}{2^{2\widetilde{j}s_N}} \leq 4\left(\frac{1}{G_e}\right)^{2s_N/(2s_N+d_R+1)}$$

and $2^{j_0(d_R+1)}/G_e \le 2^{\widetilde{j}(d_R+1)}\widetilde{j}^2/G_e = 1/2^{2\widetilde{j}s_N}$ which yields the result. $\square$

Proofs of Proposition 4

Denoting by $\mathbb{P}_{\boldsymbol{B},j}$ the law of $\mathbb{P}_{\boldsymbol{B}}$, $\boldsymbol{m}_{r,1,j}(\boldsymbol{x},y) = \mathbb{E}\left[\boldsymbol{B}_{r,1}|\underline{\boldsymbol{X}} = \boldsymbol{x}, \underline{\boldsymbol{Y}} = y\right]$, $r = 1,\dots,d_R$ the associated functions of interest, and by $\mathbb{P}_{j,G}$ the law of an i.i.d $(\underline{\boldsymbol{X}}_g, \underline{\boldsymbol{Y}}_g)_{g=1}^G$ sample of size G, for $j = 0,\dots,K$, $K \ge 1$, and use

$$\inf_{\widehat{\boldsymbol{m}}_{r,1}} \sup_{\mathbb{P}_{\boldsymbol{B}} \in \mathcal{H}^{s+1}(l)} \mathbb{E}\left[\|\widehat{\boldsymbol{m}}_{r,1} - \boldsymbol{m}_{r,1}\|_{L^q(\mathcal{S})}\right] \ge \inf_{\widehat{\boldsymbol{m}}_{r,1}} \sup_{\mathbb{P}_{\boldsymbol{B},j} \in \mathcal{H}^{s}(l), j=0,\dots,K} \mathbb{E}\left[\|\widehat{\boldsymbol{m}}_{r,1} - \boldsymbol{m}_{r,1,j}\|_{L^q(\mathcal{S})}\right]$$

and Theorem 2.6 , (2.5), and (2.9) in Tsybakov (2000) that we now recall.

Proposition 10 (Theorem 2.6 in Tsybakov (2000))**.** *Assume that* $\mathcal{H}^s(l)$ *contains* $\{\mathbb{P}_{\boldsymbol{B},j}, j = 0,\dots,K\}$, $K \ge 1$, *which satisfy:*

1. $\|\boldsymbol{m}_{r,1,1} - \boldsymbol{m}_{r,1,0}\|_{L^q(\mathcal{S})} \ge 2r(G)$, *for* $r = 1,\dots,d_R$;

2. *for all* $j = 1,\dots,K$,
$$\frac{1}{K}\sum_{j=1}^{K} \chi^2\left(\mathbb{P}_{\boldsymbol{B},j}, \mathbb{P}_{\boldsymbol{B},0}\right) \le \xi K; \tag{4.101}$$

Then, we have

$$\frac{1}{r(G)} \inf_{\widehat{\boldsymbol{m}}_{r,1}} \sup_{\mathbb{P}_{\boldsymbol{B},j} \in \mathcal{H}^{s+1}(l), j=0,\dots,K} \mathbb{E}\left[\|\widehat{\boldsymbol{m}}_{r,1} - \boldsymbol{m}_{r,1,j}\|_{L^q(\mathcal{S})}\right] \ge \frac{1}{2}\left(1 - \xi - \frac{1}{K}\right).$$

Proof when $q = \infty$. I consider the following distributions:

- $\mathbb{P}_{\boldsymbol{B},0} = \bigotimes_{l=1}^{d_R-1} \mathbb{P}_{\boldsymbol{B}_l,0}$ and $\mathbb{P}_{\boldsymbol{B}_1,0} = \dots = \mathbb{P}_{\boldsymbol{B}_{R-1},0} = \mathbb{P}_0$. This yields for all $\mathbb{P}_Y = \mathbb{P}_0$ hence $g_{r,1,0}(\boldsymbol{x},y) = y$;

- $\mathbb{P}_{\boldsymbol{B},\boldsymbol{k}}$, $\boldsymbol{k} \in \overline{\Lambda}_{j_0}$ are the compactly supported function in $[0,1]^{d_R}$ such that, for all $t \in \mathbb{R}$, $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}}$,

$$\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}},1}\right](t\boldsymbol{x}, t(1 - \boldsymbol{x}^\top \mathbf{1})) = \gamma(t)\Phi_{j_0,\boldsymbol{k}}(\boldsymbol{x}) + \mathcal{F}\left[\mathbb{P}_0\right](t), \tag{4.102}$$

with $\gamma(0)=0$, $\overline{\Lambda}_{j_0}\subset\Lambda_{j_0}$ such that the support of all functions $(\Phi_{j_0,\boldsymbol{k}})_{\boldsymbol{k}\in\overline{\Lambda}_{j_0}}$ is a strict subset of $\mathbb{S}_{\underline{\boldsymbol{X}}}$. There exists a constant c_0 such that $|\overline{\Lambda}_{j_0}|\ge c_0 2^{j_0(d_R-1)}$. This yields

$$\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}},\boldsymbol{k}}\right](\mathbf{0})=\gamma(0)\Phi_{j_0,\boldsymbol{k}}\left(\boldsymbol{x}\right)+\mathcal{F}\left[\mathbb{P}_0\right](0)=1. \tag{4.103}$$

We have, using (4.7), on $\mathcal{S}$,

$$\boldsymbol{m}_{r,1,\boldsymbol{k}}(\boldsymbol{x},y)=y+\sum_{l=1}^{d_R-1}\frac{\boldsymbol{x}_l-\mathbb{1}\{l=r\}}{f^{\boldsymbol{k}}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\int_0^y\mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv2^{j_0}\Omega_{l,j_0,\boldsymbol{k}}\left(\boldsymbol{x}\right). \tag{4.104}$$

From the end of page 724 in Rullgård and Quinto (2010) and arguments from Proposition 9, there exists a constant $\widetilde{C}_0$ depending only on d_R such that for all $f\in L^2(\mathbb{R}^{d_R})$ compactly supported in $[-1,1]^{d_R}$ and with $s>(d_R-1)/2$,

$$\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\int_{\mathbb{R}}(1\vee|t|)^{2s+(d_R-1)}\left|\mathcal{F}[f](t(\boldsymbol{x},1-\boldsymbol{x}^\top\mathbf{1}))\right|^2dtd\boldsymbol{x}\ge\frac{1}{\widetilde{C}_0}\int_{\mathbb{R}^{d_R}}(1\vee|\boldsymbol{\xi}|_2)^{2s}\left|\mathcal{F}[f](\boldsymbol{\xi})\right|^2d\boldsymbol{\xi}.$$

Thus, using (4.102), $\mathcal{H}^{s+1}(l)$ contains $\{\mathbb{P}_{\boldsymbol{B},j},j=0,1\}$, if

$$\int_{\mathbb{R}}(1\vee|t|)^{2(s+1)+(d_R-1)}\gamma(t)^2dt+\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\int_{\mathbb{R}}(1\vee|t|)^{2(s+1)+(d_R-1)}\left|\mathcal{F}\left[\mathbb{P}_0\right](t)\right|^2dt\le\frac{l^2}{\widetilde{C}_0}. \tag{4.105}$$

Then, using (4.104), that for ϵ small enough

$$\sup_{(\boldsymbol{x},y)\in\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}f^{\boldsymbol{k}}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\le2\sup_{(\boldsymbol{x},y)\in\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}f^{0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})=:K_0$$

and that, for all $\boldsymbol{k}\in\overline{\Lambda}_{j_0}$, $\Omega_{j_0,\boldsymbol{k}}$ have disjoint support,

$$\begin{aligned}\left\|\boldsymbol{m}_{r,1,\boldsymbol{k}}-\boldsymbol{m}_{r,1,\boldsymbol{k}'}\right\|_{L^\infty(\mathcal{S})}&\ge\left\|\sum_{l=1}^{d_R-1}\frac{\boldsymbol{x}_l-\mathbb{1}\{l=r\}}{f^{\boldsymbol{k}}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\int_0^y\mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv2^{j_0}\Omega_{l,j_0,\boldsymbol{k}}\left(\boldsymbol{x}\right)\right\|_{L^\infty(\mathcal{S})}\\&\ge\frac{2^{j_0(d_R+1)/2}K_1}{K_0}\sup_{y\in\mathbb{S}_{\underline{\boldsymbol{Y}}}}\left|\int_0^y\mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv\right|\end{aligned} \tag{4.106}$$

where $K_1:=\sup_{\boldsymbol{x}\in\mathbb{S}_{\underline{\boldsymbol{X}}}}\left|\sum_{l=1}^{d_R-1}\left(\boldsymbol{x}_l-\mathbb{1}\{l=r\}\right)\phi'\left(2^{j_0}\boldsymbol{x}_l-\boldsymbol{k}_l\right)\prod_{t=1,t\neq l}^{d_R-1}\phi\left(2^{j_0}\boldsymbol{x}_t-\boldsymbol{k}_t\right)\right|$ can be lower bounded independently of j_0. We also have, using Step 3. in Gaillac and Gautier (2019c), $\chi_2(\mathbb{P}_{\boldsymbol{k},n}, \mathbb{P}_{0,n}) \le en\chi_2(\mathbb{P}_{\boldsymbol{k}}, \mathbb{P}_0)$ and

$$\chi_2(\mathbb{P}_{\boldsymbol{k}}, \mathbb{P}_0) = \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \frac{f_{\underline{\boldsymbol{X}}}(\boldsymbol{x})\left(f^0_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) - f^{\boldsymbol{k}}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right)^2}{f^0_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})} d\boldsymbol{x}dy.$$

Using that $f^0_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) = f^0_{\underline{\boldsymbol{Y}}}(y) \ge \inf_{y\in\mathbb{S}_{\underline{\boldsymbol{Y}}}} f^0_{\underline{\boldsymbol{Y}}}(y) =: 1/c_{\underline{\boldsymbol{Y}}} > 0$ on $\mathbb{S}_{\underline{\boldsymbol{Y}}}$, we have

$$\begin{aligned}
\chi_2(\mathbb{P}_1, \mathbb{P}_0) &\le C_{\underline{\boldsymbol{X}}} c_{\underline{\boldsymbol{Y}}} \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left(f^0_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) - f^{\boldsymbol{k}}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right)^2 d\boldsymbol{x}dy \\
&\le C_{\underline{\boldsymbol{X}}} c_{\underline{\boldsymbol{Y}}} \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \int_{\mathbb{R}} \left|\mathcal{F}\left[\mathbb{P}_{\boldsymbol{B},\boldsymbol{k}}\right](t\boldsymbol{x}, t(1 - \boldsymbol{x}^\top \mathbf{1}))\right|^2 d\boldsymbol{x}dt \\
&\le C_{\underline{\boldsymbol{X}}} c_{\underline{\boldsymbol{Y}}} \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} |\boldsymbol{\Phi}_{j_0,\boldsymbol{k}}(\boldsymbol{x})|^2 d\boldsymbol{x} \int_{\mathbb{R}} \gamma(t)^2 dt \\
&= C_{\underline{\boldsymbol{X}}} c_{\underline{\boldsymbol{Y}}} \int_{\mathbb{R}} \gamma(t)^2 dt.
\end{aligned}$$

Hence, (4.101) is satisfied if

$$G\int_{\mathbb{R}} \gamma(t)^2 dt \le \frac{\xi|\overline{\Lambda}_{j_0}|}{C_{\underline{\boldsymbol{X}}} c_{\underline{\boldsymbol{Y}}} e}. \tag{4.107}$$

Take, for all $t \in \mathbb{R}$,

$$\gamma(t) = \frac{\epsilon(1 \wedge |t/\tau|^\nu)}{(1 + (t/\tau)^{s+(d_R+1)/2})(\tau^{s+(d_R+1)/2+1/2}(e \vee |t|)^{1/2} \ln(e \vee t/\tau))^{1/2}},$$

with $\nu \ge 1/2$,

- $\tau = 2^{j_0}$ and j_0 such that $2^{j_0} \sim (G/\ln(G))^{1/(2s+d_R+1)}$;

- ϵ such that

$$\epsilon \int_{\mathbb{R}} \frac{(1 \wedge |t|^{2\nu})}{(e \vee |t|)\ln(e \vee (t/\tau))} dt + \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \int_{\mathbb{R}} (1 \vee |t|)^{2(s+1)+(d_R-1)} \left|\mathcal{F}\left[\mathbb{P}_0\right](t)\right|^2 dt \le \frac{l^2}{\widetilde{C}_0};$$

and $\epsilon^2 \leq \xi/(C_{\underline{\boldsymbol{X}}} c_Y e\,(1+1/(2s+d_R+1)))$ which ensures that

$$\begin{aligned} G\int_{\mathbb{R}} \gamma(t)^2 dt &\leq \int_{\mathbb{R}} \frac{G\epsilon^2}{\tau^{2s+d_R+2}(1+(t/\tau)^{s+(d_R+1)/2})^2} dt \\ &\leq \left(1+\frac{1}{2s+d_R+1}\right) G 2^{-j_0(2s+d_R+1)}\epsilon^2 \\ &\leq \frac{\xi \ln(G)}{C_{\underline{\boldsymbol{X}}} c_Y e} \end{aligned}$$

hence with $\ln(G) \leq K = |\overline{\Lambda}_{j_0}|$ that (4.107) is satisfied.

Finally, we have, for j_0 sufficiently large for the second inequality to have $\tau\pi/2 \geq e$,

$$\begin{aligned} \left|\int_0^y \mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv\right| &= \frac{\epsilon\tau^{1/2}}{\tau^{s+(d_R+1)/2}} \int_{\mathbb{R}} \frac{\operatorname{sinc}\left(\tau t y/2\right)\left(1\wedge |t|^{\nu}\right)}{(1+|t|^{s+(d_R+1)/2})(e\vee|\tau t|)^{1/2}\ln(e\vee|t|)^{1/2}} dt \\ &\geq \frac{2^{-j_0(s+(d_R+1)/2)}\epsilon 2\pi}{y(\pi/2)^{1/2}\ln(e\vee(\pi/2))^{1/2}} \frac{\operatorname{sinc}(\pi/2)(1\wedge|\pi/2|^{\nu})}{(1+(\pi/2)^{s+(d_R+1)/2})}. \end{aligned}$$

Thus, using (4.106), we obtain

$$\|\boldsymbol{m}_{r,1,1}-\boldsymbol{m}_{r,1,0}\|_{L^\infty(\mathcal{S})} \geq \frac{2^{-j_0 s}K_1}{K_0} \sup_{y\in\mathbb{S}_Y} \frac{1}{y} \frac{\epsilon 2\pi}{(\pi/2)^{1/2}\ln(e\vee(\pi/2))^{1/2}} \frac{\operatorname{sinc}(\pi/2)(1\wedge|\pi/2|^{\nu})}{(1+(\pi/2)^{s+(d_R+1)/2})},$$

which yields the result using Proposition 10.

Proof when $q=2$. I consider here the following distributions:

- $\mathbb{P}_{\boldsymbol{B},0}$ is the same as in $q=\infty$;
- $K=2$ and $\mathbb{P}_{\boldsymbol{B},1}$ is the compactly supported function in $[0,1]^{d_R}$ such that, for all $t\in\mathbb{R}$, $\boldsymbol{x}\in\mathbb{S}_{\underline{\boldsymbol{X}}}$,

$$\mathcal{F}\left[\mathbb{P}_{\underline{\boldsymbol{B}},1}\right](t\boldsymbol{x}, t(1-\boldsymbol{x}^\top\mathbf{1})) = \gamma(t)\sum_{\boldsymbol{k}\in\overline{\Lambda}_{j_0}} \Phi_{j_0,\boldsymbol{k}}(\boldsymbol{x}) + \mathcal{F}\left[\mathbb{P}_0\right](t), \tag{4.108}$$

  where $\gamma(0)=0$, $\overline{\Lambda}_{j_0}\subset\Lambda_{j_0}$ such that the support of all functions $(\Phi_{j_0,\boldsymbol{k}})_{\boldsymbol{k}\in\overline{\Lambda}_{j_0}}$ is a strict subset of $\mathbb{S}_{\underline{\boldsymbol{X}}}$.

Consider $d_R=2$. The first condition (4.105) remains the same using that $(\Phi_{j,\boldsymbol{k}})_{j\geq J,\boldsymbol{k}\in\Lambda_j}$ is an orthonormal system of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$. Then, using (4.66) and $|\overline{\Lambda}_{j_0}|\geq c_0 2^{j_0(d_R-1)}$, for

$r = 1, 2,$

$$\|\boldsymbol{m}_{r,1,1} - \boldsymbol{m}_{r,1,0}\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})} = \left\| \frac{\boldsymbol{x}_r - \mathbb{1}\{r=1\}}{f^1_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})} \int_0^y \mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv \sum_{\boldsymbol{k}\in\overline{\Lambda}_{j_0}} 2^{j_0}\Omega_{l,j_0,\boldsymbol{k}}(\boldsymbol{x}) \right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}$$
$$\geq \frac{b_1^2 c_0 2^{j_0(d_R+1)}}{K_0^2} \inf_{\boldsymbol{x}_1\in\mathbb{S}_{\underline{\boldsymbol{X}}}} |\boldsymbol{x}_r - \mathbb{1}\{r=1\}|^2 \int_{\mathbb{S}_Y} \left| \int_0^y \mathcal{F}^{-1}\left[\gamma(\cdot)\right](v)dv \right|^2 dy.$$

The end of the proof is similar to the case $q = \infty$ taking here $\tau = 2^{j_0}$ and j_0 such that $2^{j_0} \sim n^{1/(2s+d_R+1)}$. □

Proofs of Section 4.2.3

We consider the more general version of Proposition 5 below, where $f_{\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ are estimated under assumptions of Section 4.5.2.

Proposition 11 (Data-driven convergence rates for the L^2 risk)**.** *Let $d_C = 2$, $l > 0$, $N \in \mathbb{N}$. Make assumptions 1-3, 8 and 7, then we have that, for $r = 1, \dots, d_R - 1$,*

$$\frac{1}{r(G_e)} \sup_{\substack{\mathbb{P}_{\boldsymbol{B}}\in\mathcal{H}^{s+1}(l) \\ f_{\underline{\boldsymbol{X}}}\in\mathcal{E},\ f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\in\mathcal{E}'}} \mathcal{R}^2_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j_0}}_{r,1}, \boldsymbol{m}_{r,1}\right) = \underset{\mathcal{O}_{G_0,G_1,G}}{O_p}(1), \tag{4.109}$$

where $\mathcal{O}_{G_0,G_1,G} = \{v(G_0,\mathcal{E})/\delta(G_0) \leq G^{-2}\ln(G)^{-1}, v(G_1,\mathcal{E}')/(\delta(G_0)\delta(G_1)) \leq G^{-2}\ln(G)^{-1}\}$, and $r(G_e) = (G_e/\ln(G_e))^{-s_N/(2s_N+d_R+1)}$, where s_N is defined in Proposition 4.

Let $\mathcal{J}_G$ be the set of functions $j \in \mathbb{N}_0^{\mathbb{R}}$ such that for all $y \in \mathbb{S}_{\underline{\boldsymbol{Y}}}$, $j(y) \in \{0, \dots, j_{\max}\}$. I use, for all $j \geq J$, $\boldsymbol{k} \in \Lambda_j$, $\boldsymbol{w} \in \widetilde{\mathcal{W}}$, $\Delta_{j,\boldsymbol{k},\boldsymbol{w}} := \widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - \widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)$, $\widetilde{\Delta}_{j,\boldsymbol{k},\boldsymbol{w}} := \widetilde{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - d_{j,\boldsymbol{k},\boldsymbol{w}}(y)$,

$$L_l^{j_0} : (y, \boldsymbol{x}) \mapsto \left(\widehat{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(y, \boldsymbol{x}).$$

I also use

$$\Xi_l\left(y,j_0\right) := \sum_{j>j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\widetilde{\mathcal{W}}} 2^{2j}\left|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2, \quad S_1\left(y,j_0\right) := \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} 2^{2j}\left|\Delta_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2,$$

$$S_{2,l}\left(y,j_0\right) := \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\left(\widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} - \partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(y,\boldsymbol{x})\right|^2 d\boldsymbol{x}, \; L := \frac{\sqrt{2|\widetilde{\mathcal{W}}|}}{42\kappa_2^{N,d_R,J}},$$

$$\Psi_{0,G} := \exp\left(-\frac{p_G}{6}\right) + \frac{2942^{j_{\max}(d_R+1)}\kappa_2^{N,d_R,J}}{G|\widetilde{\mathcal{W}}|}\exp\left(-L\sqrt{Gp_G}\right).$$

Lemma 3. *For all $q=2$, $y\in\mathbb{S}_{\underline{\boldsymbol{Y}}}$, $l=1,\ldots,d_R-1$, and $j_0\in\{0,\ldots,j_{\max}\}$, we have*

$$\mathbb{E}\left[S_1\left(y,\widehat{j}_0(y)\right)\right] \le \frac{Z_{G_0}C_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|2^{j_{\max}(d_R+1)}}{1-2^{-(d_R+1)}}, \tag{4.110}$$

$$\mathbb{E}\left[\left(S_{2,l}(y,j_0) - \frac{\Sigma(j_0)}{6\widetilde{A}}\right)_+\right] \le 48\frac{2^{j_0(d_R+1)}c_{\underline{\boldsymbol{X}}}\overline{A}|\widetilde{\mathcal{W}}|}{G(1-2^{-(d_R+1)})}\Psi_{0,G}. \tag{4.111}$$

Proof. Let the parameters in the *for all* statement be given and $l\in 1,\ldots,d_R-1$.
Proof of (4.110). Using

$$\mathbb{E}\left[\left|\Delta_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right] \le \mathbb{E}\left[\frac{Z_{G_0}}{G^2}\left|\sum_{g=1}^{G}\left|\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\underline{\boldsymbol{X}}_g)\right|\right|^2\right] \le Z_{G_0}C_{\underline{\boldsymbol{X}}}$$

(4.74), and $\sum_{j=J}^{j_{\max}} 2^{(j-j_{\max})(d_R+1)} \le 1/(1-2^{-(d_R+1)})$, we obtain (4.110) from the following inequalities

$$\mathbb{E}\left[S_1\left(y,\widehat{j}_0(y)\right)\right] \le \sum_{j=J}^{j_{\max}} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} 2^{2j}\mathbb{E}\left[\left|\Delta_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2\right] \le \frac{Z_{G_0}C_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|2^{j_{\max}(d_R+1)}}{1-2^{-(d_R+1)}}.$$

Proof of (4.111)**.** We use

$$S_{2,l}(y,j_0)=\sup_{v\in\mathcal{U}}\left|\nu_G^y(v)\right|^2,$$

$$\nu_G^y(v):=\left\langle\left(\partial_l\widetilde{F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}-\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(y,\cdot),v(\cdot)\right\rangle_{\mathbb{S}_{\underline{\boldsymbol{X}}}}=\frac{1}{G}\sum_{g=1}^{G}\left(f_v^y(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g)-\mathbb{E}\left[f_v^y(\underline{\boldsymbol{X}}_g,\underline{\boldsymbol{Y}}_g)\right]\right),$$

$$f_v^y(\cdot,\star):=\frac{\mathbb{1}\{\star\le y\}}{f_{\underline{\boldsymbol{X}}}(\cdot)}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}}\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\cdot)2^j\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\boldsymbol{x})\overline{v}(\boldsymbol{x})d\boldsymbol{x},$$

where $\mathcal{U}$ is a countable dense set of measurable functions of $\left\{v:\ \|v\|_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}=1\right\}$ and check the conditions of the Talagrand inequality given in Lemma B.15 in Gaillac and Gautier (2019c) with $\eta=p_G$ and $\Lambda(p_G)=1$. For all $u\in\mathcal{U}$, using the Cauchy-Schwarz inequality for the first display, (4.66) for the second inequality, and (4.70) for the third one, we obtain

$$\begin{aligned}\|f_v^y\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}}\times[0,1])}&\le\sqrt{c_{\underline{\boldsymbol{X}}}}\left\|\left(\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}\left|\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\cdot)\right|^2 2^{2j}\int_{\mathbb{S}_{\underline{\boldsymbol{X}}}}\left|\Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}(\boldsymbol{x})\right|^2d\boldsymbol{x}\right)^{1/2}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}\\&\le\sqrt{c_{\underline{\boldsymbol{X}}}A_l}\left\|\left(\sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}2^{2j}\left|\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}(\cdot)\right|^2\right)^{1/2}\right\|_{L^\infty(\mathbb{S}_{\underline{\boldsymbol{X}}})}\\&\le\sqrt{c_{\underline{\boldsymbol{X}}}A_l\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}\left(\sum_{j=J}^{j_0}2^{j(d_R+1)}\right)^{1/2}\\&\le 2^{j_0(d_R+1)/2}\frac{\sqrt{c_{\underline{\boldsymbol{X}}}A_l\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}}{(1-2^{-(d_R+1)})^{1/2}}.\end{aligned}\tag{4.112}$$

By the Cauchy-Schwarz inequality, (4.66), Lemma 2, and (4.74), we have

$$\begin{aligned}\mathbb{E}\left[\sup_{v\in\mathcal{U}}\left|\nu_G^y(v)\right|\right]^2\le\mathbb{E}\left[\sup_{v\in\mathcal{U}}\left|\nu_G^y(v)\right|^2\right]&\le\mathbb{E}\left[\left\|\left(\widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}-\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(y,\cdot)\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}\right]\\&\le A_l\frac{c_{\underline{\boldsymbol{X}}}}{G}|\widetilde{\mathcal{W}}|\sum_{j=J}^{j_0}2^{j(d_R+1)}\\&\le 2^{j_0(d_R+1)}\frac{A_l c_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|}{G(1-2^{-(d_R+1)})}=\frac{\Sigma(j_0)}{24(1+2p_G)\widetilde{A}}.\end{aligned}$$

Finally, by the Cauchy-Schwarz inequality and (4.112), we have

$$\mathrm{Var}\left(\Re(f_v^y(\underline{\boldsymbol{Y}}_g,\underline{\boldsymbol{X}}_g))\right)\vee\mathrm{Var}\left(\Im(f_v^y(\underline{\boldsymbol{Y}}_g,\underline{\boldsymbol{X}}_g))\right)\le\int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}}\left|f_v^y(y',\boldsymbol{x})\right|^2 f_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}(y',\boldsymbol{x})dy'd\boldsymbol{x}$$
$$\le 2^{j_0(d_R+1)}\frac{c_{\underline{\boldsymbol{X}}}A_l|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}}.$$

□

Denote by

$$\mathcal{R}^{2,2}_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j}_0}_{r,1},\boldsymbol{m}_{r,1}\right):=\mathbb{E}\left[\left\|\widehat{\boldsymbol{m}}^{\widehat{j}_0}_{r,1}-\boldsymbol{m}_{r,1}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}})}\right].$$

Lemma 4. *For all $r=1,\dots,d_R-1$, and $j_0\in\mathcal{J}_G$,*

$$\mathcal{R}^{2,2}_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j}_0}_{r,1},\boldsymbol{m}_{r,1}\right)\le C_{8,r}\sum_{l=1}^{d_R-1}\mathbb{E}\left[\left\|L_l^{j_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right]+12\overline{A}C_{7,r}\int_{y\in\mathbb{S}_{\underline{\boldsymbol{Y}}}}\Sigma(y,j_0(y))dy$$
$$+\widetilde{C}_0\frac{Z_{G_1}2^{j_{\max}(d_R+1)}}{\delta(G_0)}+j_{\max}2^{j_{\max}(d_R+1)}C_{7,r}\Pi(G,Z_{G_0},j_0),$$

where $C_{7,r}:=\max_{l=1,\dots,d_R-1}C_{7,r,l}$, $C_{8,r}:=(18\overline{A}+3)\widetilde{A}C_{7,r}/(3(d_R-1))$, and

$$\Pi(G,Z_{G_0},j_0):=1152\frac{c_{\underline{\boldsymbol{X}}}\overline{A}^2\widetilde{A}|\widetilde{\mathcal{W}}|}{G(1-2^{-(d_R+1)})}\Psi_{0,G}+Z_{G_0}\frac{24\overline{A}^2C_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}},$$

Proof of Lemma 4. Let $j_0\in\{0,\dots,j_{\max}\}$. We have, using (4.96)

$$\mathcal{R}^{2,2}_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j}_0}_{r,1},\boldsymbol{m}_{r,1}\right)\le\widetilde{C}_0\frac{Z_{G_1}2^{j_{\max}(d_R+1)}}{\delta(G_0)}+\frac{C_{7,r}}{3(d_R-1)}\sum_{l=1}^{d_R-1}\mathbb{E}\left[\left\|L_l^{\widehat{j}_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right].\tag{4.113}$$

Using, for all $j_1,j_2\in\mathbb{N}$ and $y\in[0,1]$,

$$\widetilde{R}^{j_2}_{j_1,l}(y,\cdot):=\left(\widehat{\partial_lF}^{j_2\vee j_1}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}-\widehat{\partial_lF}^{j_1}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)(y,\cdot),$$

we have $L_l^{\widehat{j}_0} = \widetilde{R}^{j_0}_{\widehat{j}_0(y),l} - \widetilde{R}^{\widehat{j}_0(y)}_{j_0,l} + L_l^{j_0}$. We obtain, using the convexity of $x \mapsto x^2$,

$$\begin{aligned}\mathbb{E}\left[\left\|L_l^{\widehat{j}_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right] \leq & 3\mathbb{E}\left[\|\widetilde{R}^{j_0}_{\widehat{j}_0,l}\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right] + 3\mathbb{E}\left[\|\widetilde{R}^{\widehat{j}_0}_{j_0,l}\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right] \\ & + 3\mathbb{E}\left[\left\|L_l^{j_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right].\end{aligned}$$

Because

$$\beta\left(y, j_0\right) = \max_{j' \in \mathbb{N}_0:\ J \leq j' \leq j_{\max}} \left(\sum_{j=j_0}^{j' \vee j_0} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} 2^{2j} \left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 - \Sigma\left(j'\right)\right)_+ ,$$

we have

$$\sum_{l=1}^{d_R-1} \mathbb{E}\left[\|\widetilde{R}^{j_2}_{j_1,l}\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right] \leq \overline{A}(d_R-1)\int_{\mathbb{S}_{\underline{\boldsymbol{Y}}}} \left(\mathbb{E}\left[\beta\left(y, j_1\right)\right] + \mathbb{E}\left[\Sigma\left(j_2\right)\right]\right) dy$$

for possibly random j_1 and j_2. Using (4.18) yields

$$\begin{aligned}\sum_{l=1}^{d_R-1} \mathbb{E}\left[\left\|L_l^{\widehat{j}_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right] \leq & 6\overline{A}(d_R-1)\int_{\mathbb{S}_{\underline{\boldsymbol{Y}}}} \left(\mathbb{E}\left[\beta\left(y, j_0\right)\right] + \Sigma\left(j_0\right)\right) dy \\ & + 3\sum_{l=1}^{d_R-1} \mathbb{E}\left[\left\|L_l^{j_0}\right\|^2_{L^2\left(\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\right)}\right].\end{aligned}$$

Using the convexity of $x \mapsto x^2$ and, for all $j' \in \mathcal{J}_G$,

$$\begin{aligned}\widetilde{K}^{j'}_{j_0,a}(y) &:= \sum_{j=j_0}^{j_0 \vee j'} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} \left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 2^{2j} \\ \widetilde{K}^{j'}_{j_0,b}(y) &:= \sum_{j=0}^{j_0} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} \left|\widehat{d}_{j,\boldsymbol{k},\boldsymbol{w}}(y) - d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 2^{2j} \\ \widetilde{K}^{j'}_{j_0,c}(y) &:= \sum_{j=j_0}^{j_0 \vee j'} \sum_{\boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j} \left|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 2^{2j},\end{aligned}$$

we have

$$\beta(y,j_0) \leq \max_{\substack{J\leq j'\leq j_{\max}\\ j'\in\mathbb{N}}} \left(3\sum_{m\in\{a,b,c\}} \widetilde{K}^{j'}_{j_0,m}(y) - \Sigma(j')\right)_+ .$$

We obtain, for all $y\in[0,1]$ and $l\in\{1,\dots,d_R-1\}$,

$$\widetilde{K}^{j'}_{j_0,c}(y) \leq \sum_{j=j_0}^{\infty}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} 2^{2j}|d_{j,\boldsymbol{k},\boldsymbol{w}}(y)|^2 \leq \frac{1}{a_l}\left\|L_l^{j_0}(y,\cdot_1)\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}$$

hence

$$\begin{aligned}(d_R-1)\beta(y,j_0) \leq &(d_R-1)\max_{\substack{0\leq j'\leq j_{\max}\\ j'\in\mathbb{N}}}\left(6\sum_{j=j_0}^{j'}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} 2^{2j}\left|\Delta_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 - \Sigma(j')\right)_+\\ &+3\sum_{l=1}^{d_R-1}\frac{1}{a_l}\left\|L_l^{j_0}(y,\cdot_1)\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}.\end{aligned}$$

Finally, denoting by

$$\widetilde{\beta}(y,j_0) := \max_{\substack{J\leq j'\leq j_{\max}\\ j'\in\mathbb{N}}}\left(\sum_{j=j_0}^{j'}\sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} 2^{2j}\left|\Delta_{j,\boldsymbol{k},\boldsymbol{w}}(y)\right|^2 - \frac{\Sigma(j')}{6}\right)_+$$

we have

$$\begin{aligned}\sum_{l=1}^{d_R-1}\mathbb{E}\left[\left\|L_l^{\widehat{j}_0}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},Y})}\right] \leq &36\overline{A}(d_R-1)\int_{\mathbb{S}_Y}\left(\widetilde{\beta}(y,j_0)+\Sigma\left(j_0\right)\right)dy\\ &+(18\overline{A}+3)\sum_{l=1}^{d_R-1}\frac{1}{a_l}\mathbb{E}\left[\left\|L_l^{j_0}\right\|^2_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}},Y})}\right].\end{aligned}$$

Using Lemma 3 for the third, we obtain

$$\begin{aligned}\widetilde{\beta}(y,j_0) &\leq 2\widetilde{A}\sum_{l=1}^{d_R-1}\mathbb{E}\left[\max_{J\leq j'\leq j_{\max}}\left(\underline{S}_{2,l}(y,j')-\frac{\Sigma(j')}{6\widetilde{A}}\right)_+\right] + 2\overline{A}\mathbb{E}\left[\max_{J\leq j'\leq j_{\max}}\underline{S}_1(y,j')\right]\\ &\leq 96 j_{\max}\frac{2^{j_{\max}(d_R+1)}c_{\underline{\boldsymbol{X}}}\overline{A}\widetilde{A}|\widetilde{\mathcal{W}}|}{G(1-2^{-(d_R+1)})}\Psi_{0,M} + Z_{G_0}j_{\max}\frac{2\overline{A}C_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|2^{j_{\max}(d_R+1)}}{1-2^{-(d_R+1)}},\end{aligned}$$

Hence the result. $\square$

Proof of propositions 5 and 11.

Let G_0, G_1, G such that $v(G_0,\mathcal{E})/\delta(G_0) \leq G^{-2}\ln(G)^{-1}$, $v(G_1,\mathcal{E}')/(\delta(G_0)\delta(G_1)) \leq G^{-2}\ln(G)^{-1}$, and $G_e \geq 3$. Let $r = 1,\dots,d_R-1$ and $j_0 \in \mathcal{J}_G$. Start from Lemma 4 and use (4.96), (4.98), (4.99), and (4.100), which yield

$$\begin{aligned}
\mathcal{R}^{2,2}_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j}_0}_{r,1}, \boldsymbol{m}_{r,1}\right) \leq & C_{8,r}\left(\widetilde{C}_1 Z_{G_0} 2^{j_{\max}(d_R+1)} + \frac{\widetilde{C}_4}{2^{2j_0 s_N}}\right) \\
& + \left(C_{8,r}\widetilde{C}_3 + \frac{576 C_{7,r}}{1-2^{-(d_R+1)}}\left(\frac{1}{\theta\ln(2)} + 2\right)\widetilde{A}\overline{A}c_{\underline{\boldsymbol{X}}}|\widetilde{\mathcal{W}}|\right)\frac{2^{j_0(d_R+1)}p_G}{G} \\
& + \widetilde{C}_2\frac{Z_{G_1}2^{j_{\max}(d_R+1)}}{\delta(G_0)} + j_{\max}2^{j_{\max}(d_R+1)}C_{7,r}\Pi(G, Z_{G_0}, j_0),
\end{aligned}$$

where

$$\widetilde{C}_2 := \widetilde{C}_0 + C_{8,r}2(d_R-1)C_{\underline{\boldsymbol{X}}}\left(\frac{(\kappa_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})\widetilde{\kappa}_1^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}}))^2|\mathbb{S}_{\underline{\boldsymbol{X}}}|}{2^{J(d_R+1)}} + \frac{A_l|\widetilde{\mathcal{W}}|}{1-2^{-(d_R+1)}}\right).$$

Then, we have

$$\begin{aligned}
j_{\max}\exp\left(-p_G\right)2^{j_{\max}(d_R+1)} &\leq \frac{\ln(G)}{d_R+1} \\
Z_{G_0}j_{\max}2^{j_{\max}(d_R+1)} &\leq \frac{M_{\mathcal{E},\eta,0}v(G_0,\mathcal{E})\ln(G)G}{\delta(G_0)} \leq \frac{M_{\mathcal{E},\eta,0}}{G}, \\
\frac{Z_{G_1}2^{j_{\max}(d_R+1)}}{\delta(G_0)} &\leq \frac{M_{\mathcal{E}',\eta,1}Gv(G_1,\mathcal{E}')}{\delta(G_1)\delta(G_0)} \leq \frac{M_{\mathcal{E}',\eta,1}}{G}.
\end{aligned}$$

Using (4.75), we obtain

$$\begin{aligned}
\frac{j_{\max}2^{2j_{\max}(d_R+1)}\exp(-L\sqrt{Gp_G})}{G} &\leq \frac{\ln(G)G}{d_R+1}\exp(-L\sqrt{\theta e}\ln(G)^2) \\
&\leq G^{1-L\sqrt{\theta e}}\frac{\ln(G)}{d_R+1} \\
&\leq \frac{\ln(G)}{d_R+1} \quad (\text{using} \quad \theta > 1/(eL^2)).
\end{aligned}$$

We conclude using that the solution j^* of $\ln(j^*) + (2s_N + d_R + 1)j^*\ln(2) = \ln(G_e)$

satisfy

$$j^* \leq \frac{\ln(G_e)}{(2s_N + d_R + 1)\ln(2)},$$

hence belongs to $\mathcal{J}_G$, which yields the result as

$$\begin{aligned}
&\mathcal{R}^{2,2}_{G_0,G_1}\left(\widehat{\boldsymbol{g}}^{\widehat{j}_0}_{r,1}, \boldsymbol{g}_{r,1}\right) \\
&\leq \frac{\ln(G)}{G}\left(C_{8,r}\widetilde{C}_1 + \left(\widetilde{C}_2 + \frac{24\overline{A}^2 C_{7,r} C_{\underline{\mathbf{X}}}|\widetilde{\mathcal{W}}|}{1 - 2^{-(d_R+1)}}\right) M_{\mathcal{E}',\eta,1}\right) \\
&\quad + 2^{-2j^* s_N}\left(\left(C_{8,r}\widetilde{C}_3 + \frac{576 C_{7,r}}{1 - 2^{-(d_R+1)}}\left(\frac{1}{\theta\ln(2)} + 2\right)\widetilde{A}\overline{A} c_{\underline{\mathbf{X}}}|\widetilde{\mathcal{W}}|\right)\frac{2^{j^*(2s_N+d_R+1)}\ln(G)}{G}\right) \\
&\quad + 2^{-2j^* s_N} C_{8,r}\widetilde{C}_4 + \frac{1152 c_{\underline{\mathbf{X}}}\widetilde{A}\overline{A}|\widetilde{\mathcal{W}}|\ln(G)}{G(d_R+1)(1 - 2^{-(d_R+1)})}
\end{aligned}$$

and, using $\ln(j_*) \geq \ln(G_e)$,

$$2^{-2j^* s_N} \leq \left(\frac{G_e}{\ln(j_*)}\right)^{-2s_N/(2s_N+d_R+1)} \leq \left(\frac{G_e}{\ln(G_e)}\right)^{-2s_N/(2s_N+d_R+1)}. \qquad \square$$

Proofs of Section 4.2.4

We consider the context of Section 4.5.2, where $f_{\underline{\mathbf{X}}}$ and $f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}$ are estimated. We add the following assumption.

Assumption 13. *(Asn.4)* $\sqrt{G}v(G_1, \mathcal{E}')2^{j_0 d_R}/\delta(G_1) \underset{G,G_1\to\infty}{\longrightarrow} 0;$

(Asn.5) $\sqrt{G}v(G_0, \mathcal{E})2^{j_0 d_R}/\delta(G_0).$

Under this assumptions 13 and 12, Proposition 6 holds with $f_{\underline{\mathbf{X}}}$ and $f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}$ replaced by their respective trimmed estimators. We consider this context for the proof hereafter.

We use the notation

$$\mathcal{K}_{l,j_0}\left(\underline{\boldsymbol{X}}_g, \boldsymbol{x}\right) := \sum_{j=J}^{j_0}\sum_{\boldsymbol{k}\in\Lambda_j, \boldsymbol{w}\in\mathcal{W}_j} \Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\left(\underline{\boldsymbol{X}}_g\right) 2^j \Omega^{\boldsymbol{w}}_{l,j,\boldsymbol{k}}\left(\boldsymbol{x}\right).$$

Proof of Proposition 6. We have, using the notation (4.39), for all $r = 1, \dots, d_R - 1$,

$$\sqrt{G}\left(\widehat{\boldsymbol{m}}^{j_0}_{r,1}(\boldsymbol{x}, y) - \boldsymbol{m}_{r,1}(\boldsymbol{x}, y)\right) = \sqrt{G}\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\left(\frac{\widehat{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}{\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})} - \frac{\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\right)$$
$$= \sqrt{G}\sum_{j=1}^{4} R_j(\boldsymbol{x}, y),$$

where

$$R_1(\boldsymbol{x}, y) := \sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\left(\frac{1}{\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})} - \frac{1}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\right)\widehat{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})$$
$$R_2(\boldsymbol{x}, y) := \frac{1}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\left(\widehat{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) - \widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right)$$
$$R_3(\boldsymbol{x}, y) := \frac{1}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\left(\widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) - \partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right)$$
$$R_4(\boldsymbol{x}, y) := \frac{1}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\left(\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) - \partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right).$$

We have

$$\frac{\sqrt{G}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})}\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) = G^{-1/2}\sum_{g=1}^{G} \boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y),$$

and $\mathbb{E}\left[\widetilde{\partial_l F}^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})\right] = \partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})$. Using (4.19), we show below that $\boldsymbol{\zeta}^{j_0}_{r}(\boldsymbol{x}, y)$ satisfies the Lyapounov condition holds, for $\nu > 0$,

$$\frac{\mathbb{E}\left[\left|\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y) - \mathbb{E}\left[\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y)\right]\right|^{2+\nu}\right]}{G^{\nu/2}\mathrm{Var}(\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y))^{1+\nu/2}} \longrightarrow 0.$$

Lower bound on $\mathbf{Var}(\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y))^{1+\nu/2}$**.** Because $\mathbb{E}\left[\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y)\right]$ converges to $\boldsymbol{m}_{r,1}(\boldsymbol{x}, y)$, it is sufficient to get a lower bound on $\mathbb{E}\left[\left|\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y)\right|^2\right]$. We have, using that $\left(\Psi^{\boldsymbol{w}}_{j,\boldsymbol{k}}\right)_{j \geq J, \boldsymbol{k} \in \Lambda_j, \boldsymbol{w} \in \mathcal{W}_j}$

are orthonormal on $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$ for the last display,

$$\begin{aligned}
\mathbb{E}\left[\left|\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x},y)\right|^2\right] &= \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left|\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\mathcal{K}_l\left(\boldsymbol{v},\boldsymbol{x}\right)\right|^2 \frac{\mathbb{1}\{z\leq y\} f_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}(\boldsymbol{v},z)}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^2 f_{\underline{\boldsymbol{X}}}(\boldsymbol{v})^2} dz d\boldsymbol{v} \\
&= \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x})\mathcal{K}_l\left(\boldsymbol{v},\boldsymbol{x}\right)\right|^2 \frac{F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{v})}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^2 f_{\underline{\boldsymbol{X}}}(\boldsymbol{v})} d\boldsymbol{v} \\
&\geq \frac{\widetilde{c}_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}(y)}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^2} \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left(\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{v}\right)\right|^2 d\boldsymbol{v} \\
&\geq \frac{\widetilde{c}_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}(y)}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^2} \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left|\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right|^2, \qquad (4.114)
\end{aligned}$$

where $\widetilde{c}_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}(y) := \inf_{\boldsymbol{v}\in\mathbb{S}_{\underline{\boldsymbol{X}}}} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{v})/f_{\underline{\boldsymbol{X}}}(\boldsymbol{v})$.

Upper bound on the Lyapounov condition. We have,

$$\begin{aligned}
&\mathbb{E}\left[\left|\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x},y)\right|^{2+\nu}\right] \\
&= \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left|\sum_{l=1}^{d_R-1} \frac{\boldsymbol{K}_{r,l}(\boldsymbol{x})\mathcal{K}_l\left(\boldsymbol{v},\boldsymbol{x}\right)}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) f_{\underline{\boldsymbol{X}}}(\boldsymbol{v})}\right|^{2+\nu} f_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}(\boldsymbol{v},z) dz d\boldsymbol{v} \\
&\leq \frac{c_{\underline{\boldsymbol{X}}}^{1+\nu}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^{2+\nu}} \int_{\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}} \left|\sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}} \left(\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{v}\right)\right|^{2+\nu} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(z|\boldsymbol{v}) dz d\boldsymbol{v}, \\
&\leq \frac{c_{\underline{\boldsymbol{X}}}^{1+\nu}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x})^{2+\nu}} \sup_{\boldsymbol{v}\in\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left(\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{v}\right)\right|^{\nu} B(\boldsymbol{x}),
\end{aligned}$$

where, using that $\left(\Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\right)_{j\geq J,\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j}$ are orthonormal on $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$,

$$\begin{aligned}
M(\boldsymbol{x}) &:= \int_{\mathbb{S}_{\underline{\boldsymbol{X}}}} \left|\sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left(\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{v}\right)\right|^2 d\boldsymbol{v} \\
&= \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left|\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right|^2.
\end{aligned}$$

This yields, using (4.70) for the second inequality and (4.73) for the third one,

$$\begin{aligned}
&\frac{\mathbb{E}\left[\left|\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x},y)-\mathbb{E}\left[\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x},y)\right]\right|^{2+\nu}\right]}{G^{\nu/2}\mathrm{Var}(\boldsymbol{\zeta}_{r,g}^{j_0}(\boldsymbol{x},y))^{1+\nu/2}} \\
&\leq \frac{c_{\underline{\mathbf{X}}}^{1+\nu}}{\widetilde{c}_{\underline{\mathbf{Y}},\underline{\mathbf{X}}}(y)^{1+\nu/2}G^{\nu/2}} \sup_{\boldsymbol{v}\in\mathbb{S}_{\underline{\mathbf{X}}}} \left| \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left(\sum_{l=1}^{d_R-1} \boldsymbol{K}_{r,l}(\boldsymbol{x}) 2^j \Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x}) \right) \Psi_{j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{v}) \right|^{\nu} \\
&\leq \frac{2^{j_0\nu(d_R+1)/2}}{G^{\nu/2}} \frac{c_{\underline{\mathbf{X}}}^{1+\nu}\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\mathbf{X}}})^{\nu}}{\widetilde{c}_{\underline{\mathbf{Y}},\underline{\mathbf{X}}}(y)^{1+\nu/2}} \left(\sum_{l=1}^{d_R-1} |\boldsymbol{K}_{r,l}(\boldsymbol{x})| \sum_{j=J}^{j_0} \sum_{\boldsymbol{k}\in\Lambda_j,\boldsymbol{w}\in\mathcal{W}_j} \left|\Omega_{l,j,\boldsymbol{k}}^{\boldsymbol{w}}(\boldsymbol{x})\right| \right)^{\nu} \\
&\leq \frac{j_0^{\nu}2^{j_0\nu R}}{G^{\nu/2}} \frac{c_{\underline{\mathbf{X}}}^{1+\nu}\kappa_2^{N,d_R,J}(\mathbb{S}_{\underline{\mathbf{X}}})^{\nu}\widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\mathbf{X}}})^{\nu}}{\widetilde{c}_{\underline{\mathbf{Y}},\underline{\mathbf{X}}}(y)^{1+\nu/2}} \Phi(\boldsymbol{x})^{\nu},
\end{aligned}$$

where

$$\Phi(\boldsymbol{x}) := \widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\mathbf{X}}}) \sum_{l=1}^{d_R-1} |\boldsymbol{K}_{r,l}(\boldsymbol{x})| .$$

Thus, under condition (Asn.1), the Lyapounov condition is satisfied and we have $\sqrt{G/v^{j_0}(\boldsymbol{x},y)}R_3(\boldsymbol{x},y) \xrightarrow{d} \mathcal{N}(0,1)$.

We now need to prove that the remaining terms $\sqrt{G/v^{j_0}(\boldsymbol{x},y)}R_j(\boldsymbol{x},y)$, $j=1,2,4$ are $o_p(1)$. Using the lower bound (4.114), it suffices to show that $\sqrt{G/2^{j_0(d_R+1)}}R_j(\boldsymbol{x},y) = o_p(1)$ for $j\in\{1,2,4\}$.

Term $\sqrt{G/v^{j_0}(\boldsymbol{x},y)}R_1(\boldsymbol{x},y)$. We have, using that $Z_{G_1} = O_p(v(G_1,\mathcal{E}')/\delta(G_1))$, (4.81), and (4.83),

$$|R_1(\boldsymbol{x},y)| \leq \sum_{l=1}^{d_R-1} |\boldsymbol{K}_{r,l}(\boldsymbol{x})| \left| \frac{1}{\widehat{f}_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(y|\boldsymbol{x})} - \frac{1}{f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(y|\boldsymbol{x})} \right| \left| \widehat{\partial_l F}^{j_0}_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(y|\boldsymbol{x}) \right| \leq \frac{C_0 Z_{G_1} 2^{j_0 d_R}}{\sqrt{\delta_0(G_0)}}.$$

Thus, under condition (Asn.4) we have $\sqrt{G/2^{j_0(d_R+1)}}R_1(\boldsymbol{x},y) = o_p(1)$.

Term $\sqrt{G/v^{j_0}(\boldsymbol{x},y)}R_2(\boldsymbol{x},y)$. Using (4.84) we have,

$$|R_2(\boldsymbol{x},y)| \leq c_{\underline{\mathbf{Y}},\underline{\mathbf{X}}} C_1 Z_{G_0} 2^{j_0 d_R} \sum_{l=1}^{d_R-1} |\boldsymbol{K}_{r,l}(\boldsymbol{x})| .$$

Thus, under condition (Asn.5) we have $\sqrt{G/2^{j_0(d_R+1)}}R_2(\boldsymbol{x},y)=o_p(1)$.

Term $\sqrt{G/v^{j_0}(\boldsymbol{x},y)}R_4(\boldsymbol{x},y)$. Using (4.94), we have

$$|R_4(\boldsymbol{x},y)|\le \frac{c_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}\sqrt{2}}{\pi}\sum_{l=1}^{d_R-1}|\boldsymbol{K}_{r,l}(\boldsymbol{x})|\frac{\widetilde{\kappa}_2^{N,d_R,J}(\mathbb{S}_{\underline{\boldsymbol{X}}})}{\ln(2)s_N}\left(\frac{8\pi^2C_2^{2(d_R-1)}x_0^{2N}}{2N+1}+l^2C_0\right)^{1/2}\frac{1}{2^{(j_0+1)s_N}}$$

hence $\sqrt{G/2^{j_0(d_R+1)}}R_4(\boldsymbol{x},y)=o_p(1)$ using condition (Asn.2). This yields the result.

□

4.5.3 Handling inference with contextual effects

Proposition 12. *Let the distribution of $(\boldsymbol{B},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ satisfy (4.1) and assumptions 2 and 4. The identified set for $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\in\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}},\boldsymbol{Z}}\mapsto\mathbb{E}\left[\boldsymbol{B}|\underline{\boldsymbol{X}}=\boldsymbol{x},\underline{\boldsymbol{Y}}=\boldsymbol{y},\boldsymbol{Z}=\boldsymbol{z}\right]$ is included into the set of functions taking the form $\boldsymbol{m}=\boldsymbol{M}/f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}},\boldsymbol{Z}}$, where $\boldsymbol{B}_{r,c}:\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}},\boldsymbol{Z}}\mapsto[0,1]$ for $r=1,\dots,d_R$ and $c=1,\dots,d_C$ are continuous functions which admit a continuous derivative with respect to $\boldsymbol{y}_c$, for $c=1,\dots,d_C-1$, $\boldsymbol{M}_{r,C}=1-\sum_{c=1}^{d_C-1}\boldsymbol{M}_{r,c}$, and, for all $r=1,\dots,d_R$, $c=1,\dots,d_C-1$, and $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\in\mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}},\boldsymbol{Z}}$,*

$$\sum_{r=1}^{d_R-1}\boldsymbol{x}_r\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})+(1-\boldsymbol{x}^\top\boldsymbol{1})\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}), \tag{4.115}$$

$$\sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\sum_{c=1}^{d_C-1}\partial_{\boldsymbol{y}_c}\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})+\sum_{l=1}^{d_R-1}(\boldsymbol{x}_l-\mathbb{1}\{l=r\})\partial_{\boldsymbol{x}_l}f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}},\boldsymbol{Z}}(\boldsymbol{x},y,\boldsymbol{z}), \tag{4.116}$$

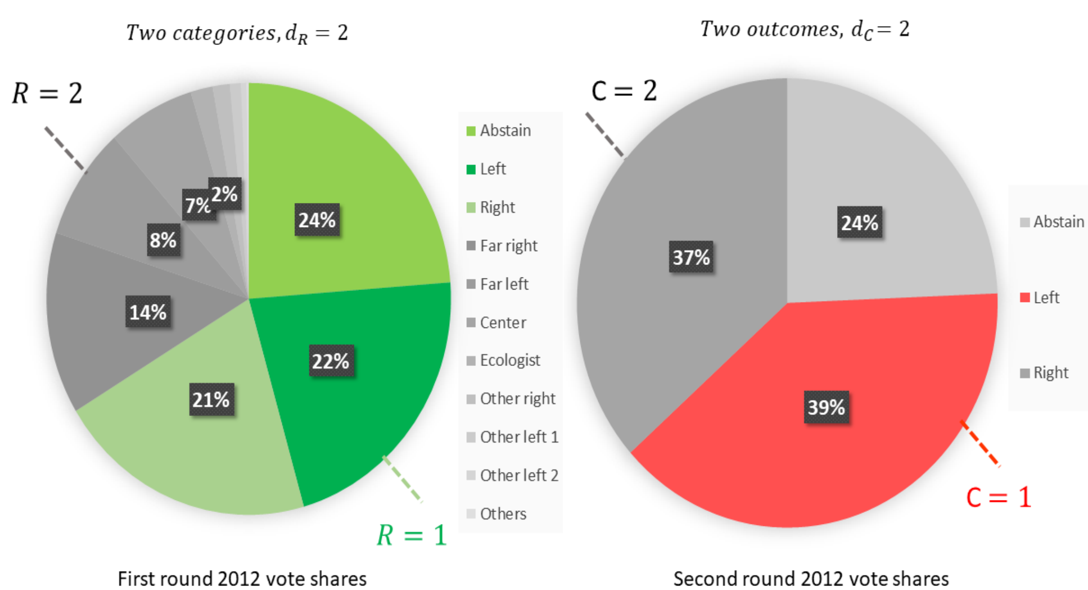
where and $\boldsymbol{\rho}_c(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}):=f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}},\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})\boldsymbol{y}_c$. Moreover, for all $c=1,\dots,d_C-1$ and $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{S}_{\underline{\boldsymbol{Y}},\underline{\boldsymbol{X}}}$, $\boldsymbol{M}_{r,c}(\boldsymbol{x},\boldsymbol{y}_1,\dots,\boldsymbol{y}_c=0,\dots,\boldsymbol{y}_{d_C-1},\boldsymbol{z})=0$.

When $d_C=2$, this set of functions is reduced to one element, for all $r=1,\dots,d_R$, (4.9) holds and $\boldsymbol{m}_{r,2}=1-\boldsymbol{m}_{r,1}$.

Proof of Proposition 12. This is a direct adaptation of the proof of Propostition 7. □

Proof of Proposition 3. We use Proposition 1 in Masten and Torgovitsky (2013), which ensures that $\boldsymbol{X}\perp\boldsymbol{B}|\boldsymbol{Z}$, where $\boldsymbol{Z}$ are gerenated covariates. Then, the result follows directly from Proposition 12. □

Complements on Section 4.3



Notes: Shares among registered voters, null votes are included in abstention.

Figure 4-2: Electoral results of the first and second rounds of 2012 Presidential elections, and the two categories and two outcome variable possibilities that correspond to the ones in the first decomposition of Section 4.3.

Table 4.7: Impact on the left-wing candidate's 2nd round vote shares among different categories of voters for Decomposition 2

Treatment	**Category 1**	**Category 2**
ITT estimation	0.0129	-0.0229
	[-0.0144,0.0310]	[-0.0444,0.0137]
Instrumental variable estimation	0.0235	-0.0416
	[-0.0260,0.0562]	[-0.0804,0.0249]

Notes: ITT estimation shows the effect of a precinct being assigned to the treatment group (ITT results from (1)) on the two different types of individuals: Category 1 are individual who voted for party at the first round of the 2007 election that qualify at the second round of the 2012 election (*i.e.* left (Royal), right (Sarkozy)) while category 2 are the others, undecided active voters plus those who abstain. Instrumental variable estimation shows the effect of a precinct being allocated to canvassers using the assignment dummy T_g as instrument. All the results use $W_g = \widetilde{PO}_g$ as control, which enters nonparametrically. The unit of observation is the unit of randomization (precinct or municipality) and each regression is based on 2,665 observations. 95% bootstrap confidence intervals are in parentheses, computed via 300 stratum-clustered bootstrap. The outcome variables are estimated using our main adaptive estimator based on Legendre functions.

Impact on vote shares according to past votes

Impact on vote shares according to level of education Part of the literature considers the role of voters' prior knowledge in the persuasion effect of political campaign communications. Prior (2006) merged local TV coverage measure with the National Election Studies surveys in the 1960s. He shows a correlation between television availability and less-educated voters support for the incumbents, who are more likely to receive media coverage. I complement this literature by looking at the difference in beliefs or preference updating after the visits according to education using actual electoral results in Pons (2018) randomized experiment.[18]

Thus, I consider another decomposition than in Section 4.3 in two categories, $d_R = 2, d_C = 2$, based on the level of education. I distinguish voters according to

[18]See Section C in Pons (2018) for the question of whether visits act through beliefs versus preferences updating, where he gives arguments favoring the first mechanism.

whether they graduated from high school or not (*i.e*, more specifically, have their *baccalauréat*). I focus, as in Section 4.3, on the impact on the left-wing candidate vote shares at the second round. I keep the same vector of controls $\boldsymbol{W}$. Results are robust to the use of the first or second round 2007 left-wing vote shares as additional controls.

Results are displayed in Table 4.8. They show that there is a positive impact of the visits on less-educated voters. There is negative but not significant effect on the more educated ones. It suggests that beliefs might be more affected by the information brought by personal visits among those with less education. This underlines the importance of prior knowledge in the persuasion mechanism, which is coherent with Bayesian models of beliefs formation.

Table 4.8: Impact on the left-wing candidate's 2nd round vote shares among different categories of voters, based on education

Category	$R = 1$, "Less educated"	$R = 2$, "More educated"
ITT estimation	0.0101	-0.0120
	[0.0010,0.0146]	[-0.0195,0.0030]
Instrumental variable estimation	0.0183	-0.0217
	[0.0016,0.0265]	[-0.0352,0.0056]

Notes: ITT estimation shows the effect of a precinct being assigned to the treatment group (ITT results from equation (1)) on the two different types of individuals: Category $R = 1$ are individuals who have not graduated high school, while Category $R = 2$ are the others. Instrumental variable estimation shows the effect of a precinct being allocated to canvassers using the assignment dummy T_g as instrument. All the results use $W_g = \widetilde{PO}_g$ as control, which enters nonparametrically. The unit of observation is the unit of randomization (precinct or municipality) and each regression is based on 2,665 observations. 95% bootstrap confidence intervals are in parentheses, computed via 300 stratum-clustered bootstrap. The outcome variables are estimated using our main adaptive estimator based on Legendre functions.

4.5.4 Main estimator based on Legendre polynomials instead of vaguelet-wavelets

Assumption 14. *Assume that* $\mathbb{S}_{\underline{\boldsymbol{X}}} = \prod_{l=1}^{d_R-1}[\widetilde{\boldsymbol{x}}_l, \widetilde{\boldsymbol{x}}_l + x_0]$*, where* $\widetilde{\boldsymbol{x}} \in [0,1]^{d_R-1}$ *and* $x_0 > 0$.

I make Assumption 14 for simplicity but one can handle the case (RC.2) which is a triangle using wavelets adapted to the triangle. The proof of Theorem 1 is constructive and my estimator is based on a plug-in approach of an estimator of $\left(\partial_{\boldsymbol{x}_l} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}\right)_{j=1}^{d_R-1}$ for $d_C = 2$. The strategy implies having first-step estimators of $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{X}}}$, similarly to Section 4.2.

Estimator and convergence rates of the L^q risk when $d_C = 2$

My approximation is based on an approximation of the derivatives, for all $y \in [0,1]$ and $l = 1, \ldots, d_R - 1$, $\boldsymbol{x} \in \mathbb{S}_{\underline{\boldsymbol{X}}} \mapsto \partial_{\boldsymbol{x}_l} F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot)$ using a truncation of its decomposition on normalized Legendre polynomials $(L_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}_0^{d_R-1}}$ in $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$ that I describe now. Let $y \in [0,1]$. Assuming that $F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) \in L^2(\mathbb{S}_{\boldsymbol{X}})$, we have the expansion

$$F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^{d_R-1}} d_{\boldsymbol{k}}(y) L_{\boldsymbol{k}}(\cdot), \tag{4.117}$$

where $d_{\boldsymbol{k}}(y) := \langle \mathbb{E}\left[\mathbb{1}\{\underline{\boldsymbol{Y}} \le y\}|\boldsymbol{X} = \cdot\right], L_{\boldsymbol{k}}\rangle_{L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})}$. Assume that $F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) \in L^2(\mathbb{S}_{\boldsymbol{X}})$ admits a square integrable derivative with respect to the $l \in \{1, \ldots, d_R - 1\}$ variable such that $\mu(\cdot)\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) \in L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$, where $\mu(\cdot) = (1 - (2(\cdot_l - \widetilde{x})/x_0 - 1)^2)^{1/2}$. Then, a valid decomposition of $\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot)$ in the space $L^2_\mu(\mathbb{S}_{\boldsymbol{X}})$ is simply

$$\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^{d_R-1}} d_{\boldsymbol{k}}(y) \partial_l L_{\boldsymbol{k}}(\cdot). \tag{4.118}$$

because the functions $\Omega_{\boldsymbol{k},l}(\cdot) = \partial_l L_{\boldsymbol{k}}(\cdot)\mu(\cdot)/\sqrt{\boldsymbol{k}_l(\boldsymbol{k}_l + 1)}$ are tensor products of associated Legendre functions and Legendre polynomials. $(\Omega_{\boldsymbol{k},l})_{\boldsymbol{k}\in\mathbb{N}_0^{d_R-1}}$ constitute also an orthonormal basis of $L^2(\mathbb{S}_{\underline{\boldsymbol{X}}})$ using, *e.g.*, 14.17.6 in Olver et al. (2010) and as they are solutions of the Sturm-Liouville equation 14.2.2 in Olver et al. (2010).[19] The use of weight μ means that I do not weight the boundaries of $\mathbb{S}_{\underline{\boldsymbol{X}}}$ in our asymptotic analysis.

Let $j_0 \geq 0$ a parameter chosen a posteriori as a function of the sample size G. For all $l = 1, \ldots, d_R - 1$, and $y \in [0, 1]$, I thus consider the approximation

$$\partial_l F^{j_0}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) := \sum_{|\boldsymbol{k}|_\infty \leq j_0} d_{\boldsymbol{k}}(y) \partial_l L_{\boldsymbol{k}}(\cdot),$$

which yields our approximation of $\boldsymbol{m}_{r,1}$.

To deal with the statistical problem, I use

$$\widehat{\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}}^{j_0}(\star|\cdot) := \sum_{|\boldsymbol{k}|_\infty \leq j_0} \widehat{d}_{\boldsymbol{k}}(\star) \partial_l L_{\boldsymbol{k}}(\cdot),$$

where, for all $y \in [0, 1]$,

$$\widehat{d}_{\boldsymbol{k}}(y) := \frac{1}{G} \sum_{g=1}^{G} \frac{\mathbb{1}\{\underline{\boldsymbol{Y}}_g \leq y\}}{\widehat{f}^{\delta}_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)} L_{\boldsymbol{k}} \left(\underline{\boldsymbol{X}}_g\right) \tag{4.119}$$

and replace $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ by $\widehat{f}^{\delta}_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ in (4.62).

L^2 risk and smoothness assumptions. I study here the $L^2_\mu(\mathcal{S})$ risk, where $\mu(\cdot) = (1-(2(\cdot_l - \widetilde{x})/x_0 - 1)^2)^{1/2}$ and $\mathcal{S}$ is defined in Assumption (Est.3), for all $r = 1, \ldots, d_R$

$$\mathcal{R}^2_{G_0,G_1}\left(\widehat{\boldsymbol{m}_{r,1}}, \boldsymbol{m}_{r,1}\right) := \mathbb{E}\left[\|\widehat{\boldsymbol{m}_{r,1}} - \boldsymbol{m}_{r,1}\|_{L^2_\mu(\mathcal{S})} \middle| \mathcal{P}_{G_0}, \mathcal{P}_{G_1}\right],$$

and use $G_e = G \wedge \lfloor (\delta(G_0)/v(G_0, \mathcal{E}))^{1/(1+\mathbb{1}\{q=\infty\})} \rfloor \wedge \lfloor (\delta(G_1)\delta(G_0)/v(G_1, \mathcal{E}'))^{1/(1+\mathbb{1}\{q=\infty\})} \rfloor$ for the sample size required for an ideal estimator where $f_{\boldsymbol{X}}$ and $f_{\boldsymbol{Y}|\boldsymbol{X}}$ are known to achieve the rate of the plug-in estimator.

[19]Note that we have

$$\mu(\cdot)\partial_l F_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\cdot) = \sum_{\boldsymbol{k} \in \mathbb{N}_0^{d_R - 1}} d_{\boldsymbol{k}}(y) \sqrt{\boldsymbol{k}_l(\boldsymbol{k}_l + 1)} \Omega_{\boldsymbol{k},l}(\cdot),$$

hence the link with the wavelet-vaguelet formulation of this inverse problem in Cai (2002) and that I use in (4.61). The wavelet-vaguelet formulation is more complex but allows to handle more general geometry of $\mathbb{S}_{\boldsymbol{X}}$ and without the weight μ.

Upper bounds on L^2 risk. The upper bounds below take the form, for $r = 1, \dots, d_R - 1$,

$$\frac{1}{r(G_e)} \sup_{\substack{f_{\boldsymbol{B}_{\cdot,1}} \in \mathcal{H}^{s+1}(l) \\ f_{\boldsymbol{X}} \in \mathcal{E},\ f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \in \mathcal{E}'}} \mathcal{R}^q_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{j_0}_{r,1}, \boldsymbol{m}_{r,1}\right) = O_p(1). \tag{4.120}$$

Proposition 13 (L^2_μ convergence rates). *Let $d_C = 2$, $l > 0$, $N \in \mathbb{N}$ and $s > 0$ such that $s_N = s\nu_{N+1} - 1/2 > 0$ where $\nu_{N+1} = 1/(1 + (d_R + 2(s+1))/N + 1) \to_{N\to\infty} 1$, $j_0 = \lfloor \widetilde{j} \rfloor$, $\widetilde{j}$ is solution of $\widetilde{j} = G_e^{1/(2s_N+d_R+1)}$. Make assumptions 1-3, 8 and 7, then (4.120) holds with $q = 2$ and $r(G_e) = G_e^{-s_N/(2s_N+d_R+1)}$.*

Proposition 13 shows that our main estimator based on Legendre polynomials admits a polynomial weighted L^2 convergence rate. In the latter, note that s_N converges as $N \to \infty$ to $s - 1/2$, hence this estimator is simpler yet non-optimal contrary to the one in Section 4.2 based on wavelets.

Data-driven estimation

Similarly to Section 4.2.3, I use the Goldenshluger-Lepski method (see, *e.g.*, Goldenshluger and Lepski, 2014; Lacour and Massart, 2016) for the data-driven choice of j_0. I focus on the adaptation with the weigthed L^2 risk. Let $p_G := \theta \ln(G)$, $\theta > 6$ and, for all $j_0 \in \mathbb{N}^{\mathbb{R}}$, $j \in \mathbb{N}$, $j_{\max} = \lfloor \widetilde{j} \rfloor$, where $\widetilde{j}$ is solution of $2^{\widetilde{j}} = G^{1/(d_R+1)}$,

$$\beta(y, j_0) := \max_{j_0+1 \le j' \le j_{\max}} \left(\sum_{|\boldsymbol{k}|_\infty \le j'} \left| \widehat{d}_{\boldsymbol{k}}(y) \right|^2 - \Sigma(j') \right)_+ ,$$
$$\Sigma(j_0) := \frac{24(1+2p_G) j_0^{d_R+1} c_{\boldsymbol{X}}}{G},$$

and $\widehat{j}_0$ is defined as

$$\forall y \in \mathbb{S}_{\boldsymbol{Y}}, \quad \widehat{j}_0(y) \in \operatorname*{argmin}_{J \le j \le j_{\max}} \left(\beta(y,j) + \Sigma(j)\right). \tag{4.121}$$

Proposition 14 (Data-driven convergence rates for the L^2 risk). *Let $d_C = 2$, $l > 0$,*

$N \in \mathbb{N}$. *Make assumptions 1-3, 14 and 7, then we have that, for* $r = 1, \ldots, d_R - 1$,

$$\frac{1}{r(G_e)} \sup_{\substack{f_{\boldsymbol{B}_{\cdot,1}} \in \mathcal{H}^{s+1}(l) \\ f_{\boldsymbol{X}} \in \mathcal{E},\ f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}} \in \mathcal{E}'}} \mathcal{R}^2_{G_0,G_1}\left(\widehat{\boldsymbol{m}}^{\widehat{j_0}}_{r,1}, \boldsymbol{m}_{r,1}\right) = \underset{\mathcal{O}_{G_0,G_1,G}}{O_p}(1), \tag{4.122}$$

$\mathcal{O}_{G_0,G_1,G} = \{v(G_0,\mathcal{E})/\delta(G_0) \le G^{-2}\ln(G)^{-1}, v(G_1,\mathcal{E}')/(\delta(G_0)\delta(G_1)) \le G^{-2}\ln(G)^{-1}\}$, *and* $r(G_e) = (G_e/\ln(G_e))^{-s_N/(2s_N+d_R+1)}$, *where* s_N *is defined in Proposition 13.*

Proof is available upon request, but can steadily be adapted from the wavelets' case.

Asymptotic normality

Assumption 15. *Assume (R.1)* $(j_0+1)^{d_R+1}/\sqrt{G} \longrightarrow 0$; *(R.2)* $\sqrt{G}v(G_1,\mathcal{E}')(j_0+1)^{d_R+1}/\delta(G_1) \underset{G,G_1\to\infty}{\longrightarrow} 0$; *(R.3)* $\inf_{\boldsymbol{v}\in\mathbb{S}_Y} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\boldsymbol{v}|\boldsymbol{x}) > 0$; *(R.4)* $\sqrt{G}v(G_0,\mathcal{E})(j_0+1)^{d_R+1}/\delta(G_0)$; *(R.5)* $G/(j_0+1)^{2s+d_R+1} \underset{G\to\infty}{\longrightarrow} 0$.

Let $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$. We have, when $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ and $f_{\underline{\boldsymbol{X}}}$ are known,

$$\widehat{\boldsymbol{m}}^{j_0}_{r,1}(\boldsymbol{x}, y) - y = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y), \tag{4.123}$$

where

$$\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y) := \sum_{l=1}^{d_R-1} \frac{(\boldsymbol{x}_l - \mathbb{1}\{l=r\})\mathbb{1}\{\underline{\boldsymbol{Y}}_g \le y\}}{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(y|\boldsymbol{x}) f_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{X}}_g)} \sum_{|\boldsymbol{k}|_\infty \le j_0} L_{\boldsymbol{k}}\left(\underline{\boldsymbol{X}}_g\right)\partial_l L_{\boldsymbol{k}}(\boldsymbol{x}).$$

I show that, the impact of estimating $f_{Y|\boldsymbol{X}}$ and $f_{\boldsymbol{X}}$ under Assumption 15 is negligible.

Proposition 15. *(Asymptotic normality) Let* $(\boldsymbol{x}, y) \in \mathbb{S}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$, $d_C = 2$, $s \ge (d_R-3)/2$. *Let* $r = 1, \ldots, d_R - 1$, $j_0 \in \mathbb{N}$, *and* $\boldsymbol{v}^{j_0}_r(\boldsymbol{x}, y) := Var\left(\boldsymbol{\zeta}^{j_0}_{r,g}(\boldsymbol{x}, y)\right)$. *Make assumptions 1-3, 7, 6, and 15, then we have,*

$$\sqrt{\frac{G}{\boldsymbol{v}^{j_0}_r(\boldsymbol{x}, y)}}\left(\widehat{\boldsymbol{m}}^{j_0}_{r,1}(\boldsymbol{x}, y) - \boldsymbol{m}_{r,1}(\boldsymbol{x}, y)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Proof is available upon request, but can steadily be adapted from the wavelets' case.

Chapter 5

Rationalizing Rational Expectations: Characterization and Tests

Joint with Xavier D'Haultfoeuille (CREST) and Arnaud Maurel (Duke University, NBER and IZA), *Forthcoming, Quantitative Economics.*

Preprint version available at here.

Associated R Package (RationalExp) and vignette available here.

Abstract

In this paper, we build a new test of rational expectations based on the marginal distributions of realizations and subjective beliefs. This test is widely applicable, including in the common situation where realizations and beliefs are observed in two different datasets that cannot be matched. We show that whether one can rationalize rational expectations is equivalent to the distribution of realizations being a mean-preserving spread of the distribution of beliefs. The null hypothesis can then be rewritten as a system of many moment inequality and equality constraints, for which tests have been recently developed in the literature. The test is robust to measurement errors under some restrictions and can be extended to account for aggregate shocks. Finally, we apply our methodology to test for rational expectations about

future earnings. While individuals tend to be right on average about their future earnings, our test strongly rejects rational expectations.

Keywords: Rational expectations, Test, Subjective expectations, Data combination.

5.1 Introduction

How individuals form their beliefs about uncertain future outcomes is critical to understanding decision making. Despite longstanding critiques (see, among many others, Pesaran, 1987; Manski, 2004), rational expectations remain by far the most popular framework to describe belief formation (Muth, 1961). This theory states that agents have expectations that do not systematically differ from the realized outcomes, and efficiently process all private information to form these expectations. Rational expectations (RE) are a key building block in many macro- and micro-economic models, and in particular in most of the dynamic microeconomic models that have been estimated over the last two decades (see, e.g., Aguirregabiria and Mira, 2010; Blundell, 2017, for recent surveys).

In this paper, we build a new test of RE. Our test only requires having access to the marginal distributions of subjective beliefs and realizations, and, as such, can be applied quite broadly. In particular, this test can be used in a data combination context, where individual realizations and subjective beliefs are observed in two different datasets that cannot be matched. Such situations are common in practice (see, e.g., Delavande, 2008; Arcidiacono, Hotz and Kang, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; Stinebrickner and Stinebrickner, 2014*a*; Gennaioli, Ma and Shleifer, 2016; Kuchler and Zafar, 2019; Boneva and Rauh, 2018; Biroli, Boneva, Raja and Rauh, 2020). Besides, even in surveys for which an explicit aim is to measure subjective expectations, such as the Michigan Survey of Consumers or the Survey of Consumer Expectations of the New York Fed, expectations and realizations can typically only be matched for a subset of the respondents. And of course, regardless of attrition, whenever one seeks to measure long or medium-term outcomes, matching beliefs with realizations does require waiting for a long period of time before the data can be made available to researchers.[1]

The tests of RE implemented so far in this context only use specific implications of the RE hypothesis. In contrast, we develop a test that exploits all possible implications of RE. Using the key insight that we can rationalize RE if and only if the distribution of realizations is a mean-preserving spread of the distribution of beliefs, we show that rationalizing RE is equivalent to satisfying one moment equality and (infinitely) many moment inequalities.[2] As a consequence, if these moment conditions hold, RE cannot be refuted, given the data at our disposal. By exhausting all relevant implications of RE, our test is able to detect much more violations of rational expectations than existing tests.

To develop a statistical test of RE rationalization, we build on the recent literature on inference based on moment inequalities, and more specifically, on Andrews and Shi (2017). By applying their results to our context, we show that our test controls size asymptotically and is consistent over fixed alternatives. We also provide conditions under which the test is not conservative.

We then consider several extensions to our baseline test. First, we show that by using a set of covariates that are common to both datasets, we can increase our ability to detect violations of RE. Another important issue is that of unanticipated aggregate shocks. Even if individuals have rational expectations, the mean of observed outcomes may differ from the mean of individual beliefs simply because of aggregate shocks. We show that our test can be easily adapted to account for such shocks.

Finally, we prove that our test is robust to measurement errors in the following sense. If individuals have rational expectations but both beliefs and outcomes are measured with (classical) errors, then we can still rationalize RE with such data provided that the amount of measurement errors on beliefs does not exceed the amount of

[1] Situations where realizations can be perfectly predicted beforehand, such as in school choice settings where assignments are a known function of observed inputs, are notable exceptions.

[2] Interestingly, the equivalence on which we rely, which is based on Strassen's theorem (Strassen, 1965), is also used in the microeconomic risk theory literature, see in particular Rothschild and Stiglitz (1970).

intervening transitory shocks plus the measurement errors on the realized outcomes. In that specific sense, imperfect data quality does not jeopardize the validity of our test. In particular, this allows for elicited beliefs to be noisier than realized outcomes. This provides a rationale for our test even in cases where realizations and beliefs are observed in the same dataset, since a direct test based on a regression of the outcome on the beliefs (see, e.g., Lovell, 1986) is, at least at the population level, not robust to any amount of measurement errors on the subjective beliefs.

We apply our framework to test for rational expectations about future earnings. To do so, we combine elicited beliefs about future earnings with realized earnings, using data from the Labor Market module of the Survey of Consumer Expectations (SCE, New York Fed), and test whether household heads form rational expectations on their annual labor earnings. While a naive test of equality of means between earnings beliefs and realizations shows that earnings expectations are realistic in the sense of not being significantly biased, thus not rejecting the rational expectations hypothesis, our test does reject rational expectations at the 1% level. Taken together, our findings illustrate the practical importance of incorporating the additional restrictions of rational expectations that are embedded in our test. The results of our test also indicate that the RE hypothesis is more credible for certain subpopulations than others. For instance, we reject RE for individuals without a college degree, who exhibit substantial deviations from RE. On the other hand, we fail to reject the hypothesis that college-educated workers have rational expectations on their future earnings.

By developing a test of rational expectations in a setting where realizations and subjective beliefs are observed in two different datasets, we bring together the literature on data combination (see, e.g., Cross and Manski, 2002*b*, Molinari and Peski, 2006*b*, Fan, Sherman and Shum, 2014, Buchinsky, Li and Liao, 2019, and Ridder and Moffitt, 2007 for a survey), and the literature on testing for rational expectations in a micro environment (see, e.g., Lovell, 1986; Gourieroux and Pradel, 1986; Ivaldi, 1992, for seminal contributions).

On the empirical side, we contribute to a rapidly growing literature on the use of subjective expectations data in economics (see, e.g., Manski, 2004; Delavande, 2008; Van der Klaauw and Wolpin, 2008; Van der Klaauw, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; de Paula, Shapira and Todd, 2014; Stinebrickner and Stinebrickner, 2014*b*; Wiswall and Zafar, 2015). In this paper, we show how to incorporate all of the relevant information from subjective beliefs combined with realized data to test for rational expectations.

The remainder of the paper is organized as follows. In Section 5.2, we present the general set-up and the main theoretical equivalences underlying our RE test. In Section 5.3, we introduce the corresponding statistical tests and study their asymptotic properties. Section 5.4 illustrates the finite sample properties of our tests through Monte Carlo simulations. Section 5.5 applies our framework to expectations about future earnings. Finally, Section 5.6 concludes. The appendix gathers the proofs of the equivalence results. We consider in the Web Appendix various theoretical extensions, additional simulation results, additional material on the application, and all the remaining proofs. Finally, the companion R package RationalExp, described in the user guide (D'Haultfœuille, Gaillac and Maurel, 2018*a*), performs the test of RE.

5.2 Set-up and characterizations

5.2.1 Set-up

We assume that the researcher has access to a first dataset containing the individual outcome variable of interest, which we denote by Y. She also observes, through a second dataset drawn from the same population, the elicited individual expectation on Y, denoted by ψ. The two datasets, however, cannot be matched. We focus on situations where the researcher has access to elicited beliefs about mean outcomes, as opposed to probabilistic expectations about the full distribution of outcomes. The type of subjective expectations data we consider in the paper has been collected in various contexts, and used in a number of prior studies (see, among others, Delavande, 2008; Zafar, 2011*b*; Arcidiacono, Hotz and Kang, 2012; Arcidiacono, Hotz, Maurel and

Romano, 2014; Hoffman and Burks, 2020).

Formally, $\psi = \mathcal{E}[Y|\mathcal{I}]$, where $\mathcal{I}$ denotes the σ-algebra corresponding to the agent's information set and $\mathcal{E}[\cdot|\mathcal{I}]$ is the subjective expectation operator (i.e. for any U, $\mathcal{E}[U|\mathcal{I}]$ is a $\mathcal{I}$-measurable random variable). We are interested in testing the rational expectations (RE) hypothesis $\psi = \mathbb{E}[Y|\mathcal{I}]$, where $\mathbb{E}[\cdot|\mathcal{I}]$ is the conditional expectation operator generated by the true data generating process. Importantly, we remain agnostic throughout most of our analysis on the information set $\mathcal{I}$. Our setting is also compatible with heterogeneity in the information different agents use to form their expectations. To see this, let $(U_1, ..., U_m)$ denote m variables that agents may or may not observe when they form their expectations, and let $D_k = 1$ if U_k is observed, 0 otherwise. Then, if $\mathcal{I}$ is the information set generated by $(D_1U_1, ..., D_mU_m)$, agents will use different subsets of the $(U_k)_{k=1...m}$ (i.e., different pieces of information) depending on the values of the $(D_k)_{k=1...m}$. Our setup encompasses a wide variety of situations, where individuals have private information and form their beliefs based on their information set. This includes various contexts where individuals form their expectations about future outcomes, including education, labor market as well as health outcomes. By remaining agnostic on the information set, our analysis complements several studies which primarily focus on testing for different information sets, while maintaining the rational expectations assumption (see Cunha and Heckman, 2007, for a survey).

It is easy to see that the RE hypothesis imposes restrictions on the joint distribution of realizations Y and beliefs ψ. In this data combination context, the relevant question of interest is then whether one can rationalize RE, in the sense that there exists a triplet $(Y', \psi', \mathcal{I}')$ such that (i) the pair of random variables (Y', ψ') are compatible with the marginal distributions of Y and ψ; and (ii) ψ' correspond to the rational expectations of Y', given the information set $\mathcal{I}'$, i.e., $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Hence, we consider the test of the following hypothesis:

H_0 : there exists a pair of random variables (Y', ψ') and a sigma-algebra $\mathcal{I}'$ such that $\sigma(\psi') \subset \mathcal{I}', Y' \sim Y,\ \psi' \sim \psi$ and $\mathbb{E}[Y'|\mathcal{I}'] = \psi'$,

where $\sim$ denotes equality in distribution. Rationalizing RE does not mean that the true realizations Y, beliefs ψ and information set $\mathcal{I}$ are such that $\mathbb{E}[Y|\mathcal{I}] = \psi$. Instead, it means that there exists a triplet $(Y', \psi', \mathcal{I}')$ consistent with the data and such that $\mathbb{E}[Y'|\mathcal{I}'] = \psi'$. In other words, a violation of H_0 implies that RE does not hold, in the sense that the true realizations, beliefs, and information set do not satisfy RE ($\mathbb{E}[Y|\mathcal{I}] \neq \psi$). The converse, however, is not true.

5.2.2 Equivalences

Main equivalence

Let F_ψ and F_Y denote the cumulative distribution functions (cdf) of ψ and Y, $x^+ = \max(0, x)$, and define

$$\Delta(y) = \int_{-\infty}^{y} F_Y(t) - F_\psi(t)dt.$$

Throughout most of our analysis, we impose the following regularity conditions on the distributions of realized outcomes (Y) and subjective beliefs (ψ):

Assumption 1. $\mathbb{E}(|Y|) < \infty$ *and* $\mathbb{E}(|\psi|) < \infty$.

The following preliminary result will be useful subsequently.

Lemma 1. *Suppose that Assumption 1 holds. Then H_0 holds if and only if there exists a pair of random variables (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$.*

Lemma 1 states that in order to test for H_0, we can focus on the constraints on the joint distribution of Y and ψ, and ignore those related to the information set. This is intuitive given that we impose no restrictions on this set. Our main result is Theorem 1 below. It states that rationalizing RE (i.e., H_0) is equivalent to a continuum of moment inequalities, and one moment equality.

Theorem 1. *Suppose that Assumption 1 holds. The following statements are equivalent:*

(i) H_0 *holds;*

(ii) *(*F_Y *mean-preserving spread of* F_ψ*)* $\Delta(y) \geq 0$ *for all* $y \in \mathbb{R}$ *and* $\mathbb{E}[Y] = \mathbb{E}[\psi]$*;*

(iii) $\mathbb{E}\left[(y-Y)^+ - (y-\psi)^+\right] \geq 0$ *for all* $y \in \mathbb{R}$ *and* $\mathbb{E}[Y] = \mathbb{E}[\psi]$*.*

The implication (i) $\Rightarrow$ (iii) and the equivalence between (ii) and (iii) are simple to establish. The key part of the result is to prove that (iii) implies (i). To show this, we first use Lemma 1, which states that H_0 is equivalent to the existence of (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$. Then the result essentially follows from Strassen's theorem (Strassen, 1965, Theorem 8).

It is interesting to note that Theorem 1 is related to the theory of risk in microeconomic theory. In particular, using the terminology of Rothschild and Stiglitz (1970), (ii) states that realizations (Y) are more risky than beliefs (ψ). The main value of Theorem 1, from a statistical point of view, is to transform H_0 into the set of moment inequality (and equality) restrictions given by (iii). We show in Section 5.3 how to build a statistical test of these conditions.

Comparison with alternative approaches We now compare our approach with alternative ones that have been proposed in the literature. In the following discussion, as in this whole section, we reason at the population level and thus ignore statistical uncertainty. Accordingly, the "tests" we consider here are formally deterministic, and we compare them in terms of data generating processes violating the null hypothesis associated with each of them.

Our approach can clearly detect many more violations of rational expectations than the "naive" approach based solely on the equality $\mathbb{E}(Y) = \mathbb{E}(\psi)$. It also detects more violations than the approach based on the restrictions $\mathbb{E}(Y) = \mathbb{E}(\psi)$ and $\mathbb{V}(Y) \geq \mathbb{V}(\psi)$ (approach based on the variance), which has been considered in particular in the macroeconomic literature on the accuracy and rationality of forecasts (see, e.g.

Patton and Timmermann, 2012). On the other hand, and as expected since it relies on the joint distribution of (Y, ψ), the "direct" approach for testing RE, based on $\mathbb{E}(Y|\psi) = \psi$, can detect more violations of rational expectations than ours.

To better understand the differences between these four different approaches ("naive", variance, "direct", and ours), it is helpful to consider important particular cases. Of course, if $\psi = \mathbb{E}[Y|\mathcal{I}]$, individuals are rational and none of the four approaches leads to reject RE. Next, consider departures from rational expectations of the form $\psi = \mathbb{E}[Y|\mathcal{I}] + \eta$, with η independent of $\mathbb{E}[Y|\mathcal{I}]$. If $\mathbb{E}(\eta) \neq 0$, subjective beliefs are biased, and individuals are on average either over-pessimistic or over-optimistic. It follows that $\mathbb{E}(Y) \neq \mathbb{E}(\psi)$, implying that all four approaches lead to reject RE.

More interestingly, if $\mathbb{E}(\eta) = 0$, individuals' expectations are right on average, and the naive approach does not lead to reject RE. However, it is easy to show that, as long as deviations from RE are heterogeneous in the population ($\mathbb{V}(\eta) > 0$), the direct approach always leads to a rejection. In this setting, our approach constitutes a middle ground, in which rejection of RE depends on the degree of dispersion of the deviations from RE (η) relative to the uncertainty shocks ($\varepsilon = Y - \mathbb{E}(Y|\mathcal{I})$). In other words and intuitively, we reject RE whenever departures from RE dominate the uncertainty shocks affecting the outcome. Formally, and using similar arguments as in Proposition 4 in Subsection 5.2.2, one can show that if ε is independent of $\mathbb{E}[Y|\mathcal{I}]$, we reject H_0 as long as the distribution of the uncertainty shocks stochastically dominates at the second-order the distribution of the deviations from RE.

Specifically, if $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, we reject RE if and only if $\sigma_\eta^2 > \sigma_\varepsilon^2$. In such a case, our approach boils down to the variance approach mentioned above: we reject whenever $\mathbb{V}(\psi) > \mathbb{V}(Y)$. But interestingly, if the discrepancy (η) between beliefs and RE is not normally distributed, we can reject H_0 even if $\mathbb{V}(\psi) \leq \mathbb{V}(Y)$. Suppose for instance that $\varepsilon \sim \mathcal{N}(0, 1)$ and

$$\eta = a\left(-\mathbb{1}\{U \leq 0.1\} + \mathbb{1}\{U \geq 0.9\}\right), \ U \sim \mathcal{U}[0,1] \text{ and } a > 0.$$

In other words, 80% of individuals are rational, 10% are over-pessimistic and form expectations equal to $\mathbb{E}[Y|\mathcal{I}]-a$, whereas 10% are over-optimistic and expect $\mathbb{E}[Y|\mathcal{I}]+a$. Then one can show that our approach leads to reject RE when $a \geq 1.755$, while for $a = 1.755$, $\mathbb{V}(\eta) \simeq 0.616 < \mathbb{V}(\varepsilon) = 1$.

Binary outcome Our equivalence result does not require the outcome Y to be continuously distributed. In the particular case where Y is binary, our test reduces to the naive test of $\mathbb{E}(Y) = \mathbb{E}(\psi)$. Indeed, when Y is a binary outcome and $\psi \in [0,1]$, one can easily show that as long as $\mathbb{E}(Y) = \mathbb{E}(\psi)$, the inequalities $\mathbb{E}\left[(y-Y)^+ - (y-\psi)^+\right] \geq 0$ automatically hold for all $y \in \mathbb{R}$. This applies to expectations about binary events, such as, e.g., being employed or not at a given date.

Interpretation of the boundary condition To shed further light on our test and on the interpretation of H_0, it is instructive to derive the distributions of $Y|\psi$ that correspond to the boundary condition ($\Delta(y) = 0$). The proposition below shows that, in the presence of rational expectations, agents whose beliefs ψ lies at the boundary of H_0 have perfect foresight, i.e. $\psi = \mathbb{E}[Y|\mathcal{I}] = Y$.[3]

Proposition 1. *Suppose that (Y,ψ) satisfies RE, $u \mapsto F^{-1}_{Y|\psi}(\tau|u)$ is continuous for all $\tau \in (0,1)$, and $\Delta(y_0) = 0$ for some y_0 in the interior of the support of ψ. Then the distribution of Y conditional on $\psi = y_0$ is degenerate: $P(Y = y_0|\psi = y_0) = 1$.*

Equivalence with covariates

In practice we may observe additional variables $X \in \mathbb{R}^{d_X}$ in both datasets. Assuming that X is in the agent's information set, we modify H_0 as follows:[4]

H_{0X} : there exists a pair of random variables (Y',ψ') and a sigma-algebra $\mathcal{I}'$ such that

$$\sigma(\psi',X) \subset \mathcal{I}',\ Y'|X \sim Y|X,\ \psi'|X \sim \psi|X \text{ and } \mathbb{E}[Y'|\mathcal{I}'] = \psi'.$$

[3] For any cdf F, we let F^{-1} denote its quantile function, namely $F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$.

[4] See complementary work by Gutknecht et al. (2018), who use subjective expectations data to relax the rational expectations assumption, and propose a method allowing to test whether specific covariates are included in the agents' information sets.

Adding covariates increases the number of restrictions that are implied by the rational expectation hypothesis, thus improving our ability to detect violations of rational expectations. Proposition 2 below formalizes this idea and shows that H_{0X} can be expressed as a continuum of conditional moment inequalities, and one conditional moment equality.

Proposition 2. *Suppose that Assumption 1 holds. The following two statements are equivalent:*

(i) H_{0X} *holds;*

(ii) Almost surely, $\mathbb{E}\left[(y-Y)^+ - (y-\psi)^+ \middle| X\right] \geq 0$ *for all* $y \in \mathbb{R}$ *and* $\mathbb{E}\left[Y - \psi | X\right] = 0$.

Moreover, if H_{0X} *holds,* H_0 *holds as well.*

Equivalence with unpredictable aggregate shocks

Oftentimes, the outcome variable is affected not only by individual-specific shocks, but also by aggregate shocks. We denote by C the random variable corresponding to the aggregate shocks. The issue, in this case, is that we observe a single realization of C (c, say), along with the outcome variable conditional on that realization $C = c$. In other words, we only identify $F_{Y|C=c}$ rather than F_Y, as the latter would require to integrate over the distribution of all possible aggregate shocks. Moreover, the restriction $\mathbb{E}\left[Y|C=c,\psi\right] = \psi$ is generally violated, even though the rational expectations hypothesis holds. It follows that one cannot directly apply our previous results by simply replacing F_Y by $F_{Y|C=c}$. In such a case, one has to make additional assumptions on how the aggregate shocks affect the outcome.

To illustrate our approach, let us consider the example of individual income. Suppose that the logarithm of income of individual i at period t, denoted by Y_{it}, satisfies a Restricted Income Profile model:

$$Y_{it} = \alpha_i + \beta_t + \varepsilon_{it},$$

where β_t capture aggregate (macroeconomic) shocks, ε_{it} follows a zero-mean random walk, and α_i, $(\beta_t)_t$ and $(\varepsilon_{it})_t$ are assumed to be mutually independent. Let $\mathcal{I}_{it-1}$ denote individual i's information set at time $t-1$, and suppose that $\mathcal{I}_{it-1} = \sigma\left(\alpha_i, (\beta_{t-k})_{k\geq 1}, (\varepsilon_{it-k})_{k\geq 1}\right)$. If individuals form rational expectations on their future outcomes, their beliefs in period $t-1$ about their future log-income in period t are given by

$$\psi_{it} = \mathbb{E}\left[Y_{it}|\mathcal{I}_{it-1}\right] = \alpha_i + \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq 1}\right] + \varepsilon_{it-1}.$$

Thus, $Y_{it} = \psi_{it} + C_t + \varepsilon_{it} - \varepsilon_{it-1}$, with $C_t = \beta_t - \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq 1}\right]$. The corresponding conditional expectation is given by:

$$\mathbb{E}\left[Y_{it}|\mathcal{I}_{it-1}, C_t = c_t\right] = \psi_{it} + c_t \neq \psi_{it}.$$

To get closer to our initial set-up, we now drop indexes i and t and maintain the conditioning on the aggregate shocks $C = c$ implicit. Under these conventions, rationalizing RE does not correspond to $\mathbb{E}\left[Y|\mathcal{I}\right] = \psi$, but instead to $\mathbb{E}\left[Y|\mathcal{I}\right] = c_0 + \psi$ for some $c_0 \in \mathbb{R}$. A similar reasoning applies to multiplicative instead of additive aggregate shocks. In such a case, the null takes the form $\mathbb{E}\left[Y|\mathcal{I}\right] = c_0\psi$, for some $c_0 > 0$. In these two examples, c_0 is identifiable: by $c_0 = \mathbb{E}(Y) - \mathbb{E}(\psi)$ in the additive case, by $c_0 = \mathbb{E}(Y)/\mathbb{E}(\psi)$ in the multiplicative case. Moreover, there exists in both cases a known function $q(y, c)$ such that $\mathbb{E}(q(Y, c_0)) = \mathbb{E}(\psi)$, namely $q(y, c) = y - c$ and $q(y, c) = y/c$ for additive and multiplicative shocks, respectively.

More generally, we consider the following null hypothesis for testing RE in the presence of aggregate shocks:

H_{0S} : there exist random variables (Y', ψ'), a sigma-algebra $\mathcal{I}'$ and $c_0 \in \mathbb{R}$ such that $\sigma(\psi') \subset \mathcal{I}'$, $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}\left[q\left(Y', c_0\right)|\mathcal{I}'\right] = \psi'$.

where $q(.,.)$ is a known function supposed to satisfy the following restrictions.

Assumption 2. $\mathbb{E}\left(|\psi|\right) < \infty$ *and for all* c, $\mathbb{E}\left(\left|q\left(Y, c\right)\right|\right) < \infty$. *Moreover,* $\mathbb{E}\left[q(Y, c)\right] =$

$\mathbb{E}[\psi]$ *admits a unique solution,* c_0*.*

By applying our main equivalence result (Theorem 1) to $q(Y, c_0)$ and ψ, we obtain the following result.

Proposition 3. *Suppose that Assumption 2 holds. Then the following statements are equivalent:*

(i) H_{0S} *holds;*

(ii) $\mathbb{E}\left[(y - q(Y, c_0))^+ - (y - \psi)^+\right] \geq 0$ *for all* $y \in \mathbb{R}$*.*

A few remarks on this proposition are in order. First, this result can be extended in a straightforward way to a setting with covariates. This is important not only to increase the ability of our test to detect violations of RE, but also because this allows for aggregate shocks that differ across observable groups. We discuss further this extension, and the corresponding statistical test, in Appendix 5.8.1. Second, in the presence of aggregate shocks, the null hypothesis does not involve a moment equality restriction anymore; the corresponding moment is used instead to identify c_0. Related, a clear limitation of the naive test ($\mathbb{E}(Y) = \mathbb{E}(\psi)$) is that, unlike our test, it is not robust to aggregate shocks. In this case, rejecting the null could either stem from violations of the rational expectation hypothesis, or simply from the presence of aggregate shocks. Third, in Appendix 5.8.1, we examine whether one can extend the results above to test for RE when aggregate shocks affect the outcomes in a more general way. Proposition 6 establishes a negative result in this respect: as long as one allows for a sufficiently flexible dependence between the outcome and the aggregate shocks, any given distribution of subjective expectations is arbitrarily close to a distribution for which RE can be rationalized. This implies that, within this more general class of outcome models, there does not exist any almost-surely continuous RE test that has non-trivial power.

Robustness to measurement errors

We have assumed so far that Y and ψ were perfectly observed; yet measurement errors in survey data are pervasive (see, e.g. Bound, Brown and Mathiowetz, 2001).

We explore in the following the extent to which our test is robust to measurement errors. By robust, we mean that we still rationalize RE, when they in fact hold. Specifically, assume that the true variables (ψ and Y) are unobserved. Instead, we only observe $\widehat{\psi}$ and $\widehat{Y}$, which are affected by classical measurement errors.[5] Namely:

$$\begin{aligned} \widehat{\psi} &= \psi + \xi_\psi \quad \text{with} \quad \xi_\psi \perp\!\!\!\perp \psi,\ \mathbb{E}[\xi_\psi] = 0 \\ \widehat{Y} &= Y + \xi_Y \quad \text{with} \quad \xi_Y \perp\!\!\!\perp Y,\ \mathbb{E}[\xi_Y] = 0. \end{aligned} \tag{5.1}$$

The following proposition shows that our test is robust to a certain degree of measurement errors on the beliefs.

Proposition 4. *Suppose that Y and ψ satisfy H_0, and let $\varepsilon = Y - \psi$ and $\left(\widehat{\psi}, \widehat{Y}\right)$ be defined as in (5.1). Suppose also that $\varepsilon + \xi_Y \perp\!\!\!\perp \psi$ and F_{ξ_ψ} dominates at the second order $F_{\xi_Y+\varepsilon}$. Then $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0.*

The key condition is that F_{ξ_ψ} dominates at the second order $F_{\xi_Y+\varepsilon}$, or, equivalently here, that $F_{\xi_Y+\varepsilon}$ is a mean-preserving spread of F_{ξ_ψ}. Recall that in the case of normal variables, $\xi_\psi \sim \mathcal{N}(0, \sigma_1^2)$ and $\xi_Y + \varepsilon \sim \mathcal{N}(0, \sigma_2^2)$, this is in turn equivalent to imposing $\sigma_1^2 \leq \sigma_2^2$. Thus, even if there is no measurement error on Y, so that $\xi_Y = 0$, this condition may hold provided that the variance of measurement errors on ψ is smaller than the variance of the uncertainty shocks on Y. More generally, this allows elicited beliefs to be - potentially much - noisier than realized outcomes, a setting which is likely to be relevant in practice. One should not infer, however, that measurement errors are innocuous in our set-up. Indeed, the converse of Proposition 4 does not hold: $\widehat{Y}$ and $\widehat{\psi}$ may satisfy H_0, though Y and ψ do not. As a simple example, suppose that $Y \sim \mathcal{N}(0, \sigma_Y^2)$, $\psi \sim \mathcal{N}(0, \sigma_\psi^2)$, $\xi_Y \sim \mathcal{N}(0, \sigma_3^2)$, $\xi_\psi = 0$ and $\sigma_\psi^2 \in (\sigma_Y^2, \sigma_Y^2 + \sigma_3^2]$. Then, $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0, since $\sigma_\psi^2 \leq \sigma_Y^2 + \sigma_3^2$, whereas Y and ψ do not, since $\sigma_\psi^2 > \sigma_Y^2$. Importantly though, Proposition 4 does show that our test is conservative in the sense that measurement errors cannot result in incorrectly concluding that the

[5]See Zafar (2011*a*) who does not find evidence of non-classical measurement errors on subjective beliefs elicited from a sample of Northwestern undergraduate students. We conjecture that our test is robust to some forms of non-classical measurement errors. However, it seems difficult in this case to obtain a general result similar to the one in Proposition 4.

RE hypothesis does not hold.

In situations where $(\widehat{Y}, \widehat{\psi})$ are jointly observed, one could in principle alternatively implement the direct test. However, in contrast to our test, the direct test is not robust to any measurement errors on the subjective beliefs ψ. Indeed, if RE holds, so that $\mathbb{E}[Y|\psi] = \psi$, it is nevertheless the case that $\mathbb{E}\left[\widehat{Y}\middle|\widehat{\psi}\right] \neq \widehat{\psi}$, as long as $\text{Cov}(\xi_Y, \widehat{\psi}) = \text{Cov}(\xi_\psi, Y) = 0$ and $\mathbb{V}(\xi_\psi) > 0$. In other words, even if individuals have rational expectations, the direct test will reject the null hypothesis in the presence of even an arbitrarily small degree of measurement errors on the elicited beliefs.

Also, it is unclear whether, in the presence of measurement errors on the elicited beliefs and beyond the restrictions on the marginal distributions, there are restrictions on the copula of $(\widehat{Y}, \widehat{\psi})$ that are implied by RE. For instance, we show in Proposition 7 in Appendix 5.8.2 that under RE, and without imposing restrictions on the dependence between $\xi_Y + \varepsilon$ and ξ_ψ, the coefficient of the (theoretical) linear regression of $\widehat{Y}$ on $\widehat{\psi}$ remains unrestricted.[6] On the other hand, if one assumes that $\text{Cov}(\xi_Y + \varepsilon, \xi_\psi) \geq 0$ and $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi) \geq \underline{\lambda}$ for some $\underline{\lambda} \geq 0$, Proposition 7 also shows that the coefficient of the linear regression of $\widehat{Y}$ on $\widehat{\psi}$ is bounded from below under RE. Such a restriction, which does require to take a stand on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$, can be easily added to the moment inequalities of our test if $(\widehat{Y}, \widehat{\psi})$ is observed.

Other extensions

We now briefly discuss other relevant directions in which Theorem 1 can be extended. First, another potential source of uncertainty on ψ is rounding. Rounding practices by interviewees are common in the case of subjective beliefs. Under additional restrictions, it is possible in such a case to construct bounds on the true beliefs ψ (see, e.g., Manski and Molinari, 2010). We show in Appendix 5.8.3 that our test can be generalized to accommodate this rounding practice.

Second, we have implicitly maintained the assumption so far that subjective beliefs

[6] There might of course possibly be additional relevant information in the higher-order moments, although we have not been able to find any.

and realized outcomes are drawn from the same population. In Appendix 5.8.4, we relax this assumption and show that our test can be easily extended to allow for sample selection under unconfoundedness, through an appropriate reweighting of the observations.

Third, our equivalence result and our test can be extended to accommodate situations with multiple outcomes $(Y_k)_{k=1,..,K}$ and multiple subjective beliefs $(\psi_k)_{k=1,..,K}$ associated with each of these outcomes. Specifically, whether one can rationalize rational expectations in this environment can be written as:

$$\mathbb{E}(Y_k|\psi_1, ..., \psi_K) = \psi_k, \text{ for all } k \in \{1, ..., K\}$$

which, in turn, is equivalent to the distribution of the outcomes Y_k being a mean-preserving spread of the distribution of the beliefs ψ_k. This situation arises in various contexts, including cases where respondents declare their subjective probabilities of making particular choices among $K + 1$ possible alternatives. This also arises in situations where expectations about the distribution of a continuous outcome Y are elicited through questions of the form "what do you think is the percent chance that [Y] will be greater than [y]?", for different values $(y_k)_{k=1,..,K}$. In such cases, it is natural to build a RE test based on the multiple outcomes $(\mathbb{1}\{Y > y_k\})_{k=1,..,K}$ and subjective beliefs $(\psi_k)_{k=1,..,K}$, where ψ_k is the subjective survival function of Y evaluated at y_k.

5.3 Statistical tests

We now propose a testing procedure for H_{0X}, which can be easily adapted to the case where no covariate common to both datasets is available to the analyst. To simplify notation, we use a potential outcome framework to describe our data combination problem. Specifically, instead of observing (Y, ψ), we suppose to observe only, in addition to the covariates X, $\widetilde{Y} = DY + (1 - D)\psi$ and D, where $D = 1$ (resp. $D = 0$) if the unit belongs to the dataset of Y (resp. ψ). As in Subsection 5.2.1, we assume that the two samples are drawn from the same population, which amounts to supposing that $D \perp\!\!\!\perp (X, Y, \psi)$ (see Assumption 3-(i) below). In order to build our test, we use the characterization (ii) of Proposition 2:

$$\mathbb{E}\left[(y-Y)^{+}-(y-\psi)^{+}\middle|X\right] \geq 0 \quad \forall y \in \mathbb{R} \;\text{ and }\; \mathbb{E}\left[Y-\psi|X\right]=0.$$

Equivalently but written more compactly with $\widetilde{Y}$ only,

$$\mathbb{E}\left[W\left(y-\widetilde{Y}\right)^{+}\middle|X\right] \geq 0 \quad \forall y \in \mathbb{R} \;\text{ and }\; \mathbb{E}\left[W\widetilde{Y}\middle|X\right]=0,$$

where $W = D/\mathbb{E}(D)-(1-D)/\mathbb{E}(1-D)$. This formulation of the null hypothesis allows us to apply the instrumental functions approach of Andrews and Shi (2017, AS), who consider the issue of testing many conditional moment inequalities and equalities. We then build on their results to establish that our test controls size asymptotically and is consistent over fixed alternatives.[7] The initial step is to transform the conditional moments into the following unconditional moments conditions:

$$\mathbb{E}\left[W\left(y-\widetilde{Y}\right)^{+}g(X)\right] \geq 0, \quad \mathbb{E}\left[\left(Y-\psi\right)g(X)\right]=0,$$

for all $y \in \mathbb{R}$ and g belonging to a suitable class of non-negative functions.

We suppose to observe a sample $(D_i, X_i, \widetilde{Y}_i)_{i=1\ldots n}$ of n i.i.d. copies of $(D, X, \widetilde{Y})$. We consider instrumental functions g that are indicators of belonging to specific hypercubes within $[0,1]^{d_X}$, hence we tranform the variables X_i to lie in $[0,1]^{d_X}$. For notational convenience, we let $\widetilde{X}_i$ denote the nontransformed vector of covariates, and redefine X_i as:

$$X_i = \Phi_0\left(\widehat{\Sigma}_{\widetilde{X},n}^{-1/2}\left(\widetilde{X}_i-\overline{\widetilde{X}}_i\right)\right),$$

where, for any $x=(x_1,\ldots,x_{d_X})$, we let $\Phi_0(x)=\left(\Phi(x_1),\ldots,\Phi\left(x_{d_X}\right)\right)^{\top}$. Here Φ denotes the standard normal cdf, $\widehat{\Sigma}_{\widetilde{X},n}$ is the sample covariance matrix of $\left(\widetilde{X}_i\right)_{i=1\ldots n}$ and $\overline{\widetilde{X}}_n$ its sample mean.

[7] Other testing procedures could be used to implement our test, such as that proposed by Linton et al. (2010).

Specifically, we consider instrumental functions g belonging to the class of functions $\mathcal{G}_r = \{g_{a,r},\ a \in A_r\}$, with $A_r = \{1, 2, \dots, 2r\}^{d_X}$ $(r \geq 1)$, $g_{a,r}(x) = \mathbb{1}\{x \in C_{a,r}\}$ and, for any $a = (a_1, ..., a_{d_X})^\top \in A_r$,

$$C_{a,r} = \prod_{u=1}^{d_X} \left(\frac{a_u - 1}{2r}, \frac{a_u}{2r} \right].$$

Finally, to define the test statistic T, we need to introduce additional notations. First, let $w_i = nD_i / \sum_{j=1}^n D_j - n(1-D_i)/\sum_{j=1}^n (1-D_j)$ and define, for any $y \in \mathbb{R}$,

$$m\left(D_i, \widetilde{Y}_i, X_i, g, y\right) = \begin{pmatrix} m_1\left(D_i, \widetilde{Y}_i, X_i, g, y\right) \\ m_2\left(D_i, \widetilde{Y}_i, X_i, g, y\right) \end{pmatrix} = \begin{pmatrix} w_i\left(y - \widetilde{Y}_i\right)^+ g\left(X_i\right) \\ w_i \widetilde{Y}_i g\left(X_i\right) \end{pmatrix}. \quad (5.2)$$

Let $\overline{m}_n(g, y) = \sum_{i=1}^n m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)/n$ and define similarly $\overline{m}_{n,j}$ for $j = 1, 2$. For any function g and any $y \in \mathbb{R}$, we also define, for some $\epsilon > 0$,

$$\overline{\Sigma}_n(g, y) = \widehat{\Sigma}_n(g, y) + \epsilon \text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}\right), \widehat{\mathbb{V}}\left(\widetilde{Y}\right)\right),$$

where $\widehat{\Sigma}_n(g, y)$ is the sample covariance matrix of $\sqrt{n}\overline{m}_n(g, y)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}\right)$ is the empirical variance of $\widetilde{Y}$. We then denote by $\overline{\Sigma}_{n,jj}(g, y)$ $(j = 1, 2)$ the j-th diagonal term of $\overline{\Sigma}_n(g, y)$.

Then the (Cramér-von-Mises) test statistic T is defined by

$$T = \sup_{y \in \widehat{\mathcal{Y}}} \sum_{r=1}^{r_n} \frac{(2r)^{-d_X}}{(r^2 + 100)} \sum_{a \in A_r} \left[(1-p) \left(-\frac{\sqrt{n}\overline{m}_{n,1}(g_{a,r}, y)}{\overline{\Sigma}_{n,11}(g_{a,r}, y)^{1/2}} \right)^{+2} + p \left(\frac{\sqrt{n}\overline{m}_{n,2}(g_{a,r}, y)}{\overline{\Sigma}_{n,22}(g_{a,r}, y)^{1/2}} \right)^2 \right],$$

where $\widehat{\mathcal{Y}} = \left[\min_{i=1,\dots,n} \widetilde{Y}_i, \max_{i=1,\dots,n} \widetilde{Y}_i \right]$, $p \in (0, 1)$ is a parameter weighting the moments inequalities versus equalities and $(r_n)_{n \in \mathbb{N}}$ is a deterministic sequence tending to infinity.

To test for rational expectations in the absence of covariates, we set the instrumental function equal to the constant function $g(X) = 1$, and the test statistic is simply written as:

$$T=\sup_{y\in\widehat{\mathcal{Y}}}\left[(1-p)\left(-\frac{\sqrt{n}\overline{m}_{n,1}(y)}{\overline{\Sigma}_{n,11}(y)^{1/2}}\right)^{+2}+p\left(\frac{\sqrt{n}\overline{m}_{n,2}(y)}{\overline{\Sigma}_{n,22}(y)^{1/2}}\right)^{2}\right],$$

where, using the notations introduced above, $\overline{m}_{n,j}(y)=\overline{m}_{n,j}(1,y)$ and $\overline{\Sigma}_{n,jj}(y)=\overline{\Sigma}_{n,jj}(1,y)$ $(j=1,2)$.

Whether or not covariates are included, the resulting test is of the form $\varphi_{n,\alpha}=\mathbb{1}\left\{T>c^*_{n,\alpha}\right\}$ where the estimated critical value $c^*_{n,\alpha}$ is obtained by bootstrap using as in AS the Generalized Moment Selection method. Specifically, we follow three steps:

1. Compute the function $\overline{\varphi}_n(y,g)=\left(\overline{\varphi}_{n,1}(y,g),0\right)^\top$ for (y,g) in $\widehat{\mathcal{Y}}\times\cup_{r=1}^{r_n}\mathcal{G}_r$, with

$$\overline{\varphi}_{n,1}(y,g)=\overline{\Sigma}_{n,11}^{1/2}B_n\mathbb{1}\left\{\frac{n^{1/2}}{\kappa_n}\overline{\Sigma}_{n,11}^{-1/2}\overline{m}_{n,1}(y,g)>1\right\},$$

   and where $B_n=(b_0\ln(n)/\ln(\ln(n)))^{1/2}$, $b_0>0$, $\kappa_n=(\kappa\ln(n))^{1/2}$, and $\kappa>0$. To compute $\overline{\Sigma}_{n,11}$, we fix ϵ to 0.05, as in AS.

2. Let $\left(D_i^*,\widetilde{Y}_i^*,X_i^*\right)_{i=1,\ldots,n}$ denote a bootstrap sample, i.e., an i.i.d. sample from the empirical cdf of $\left(D,\widetilde{Y},X\right)$, and compute from this sample the bootstrap counterparts of $\overline{m}_n$ and $\overline{\Sigma}_n$, $\overline{m}_n^*$ and $\overline{\Sigma}_n^*$. Then compute the bootstrap counterpart of T, T^*, replacing $\overline{\Sigma}_n(y,g_{a,r})$ and $\sqrt{n}\overline{m}_n(y,g_{a,r})$ by $\overline{\Sigma}_n^*(y,g_{a,r})$ and $\sqrt{n}\left(\overline{m}_n^*-\overline{m}_n\right)(y,g_{a,r})+\overline{\varphi}_n(y,g_{a,r})$, respectively.

3. The threshold $c^*_{n,\alpha}$ is the quantile (conditional on the data) of order $1-\alpha+\eta$ of $T^*+\eta$ for some $\eta>0$. Following AS, we set η to 10^{-6}.

Note that, despite the multiple steps involved, the testing procedure remains computationally easily tractable. In particular, for the baseline sample we use in our application (see Section 5.5.1), the RE test only takes 2 minutes.[8]

We now turn to the asymptotic properties of the test. For that purpose, it is convenient to introduce additional notations. Let $\mathcal{Y}$ and $\mathcal{X}$ denote the support of Y

[8]This CPU time is obtained using our companion R package, on an Intel Xeon CPU E5-2643, 3.30GHz with 256Gb of RAM.

and X respectively, and

$$\mathcal{L}_F = \left\{ (y, g_{a,r}) : y \in \mathcal{Y},\ (a,r) \in A_r \times \mathbb{N} : \ \mathbb{E}_F \left[W \left(y - \widetilde{Y} \right)^+ g_{a,r}(X) \right] = 0 \right\},$$

where, to make the dependence on the underlying probability measure explicit, $\mathbb{E}_F$ denotes the expectation with respect to the distribution F of $\left(D, \widetilde{Y}, X\right)$. Finally, let $\mathcal{F}$ denote a subset of all possible cumulative distribution functions of $\left(D, \widetilde{Y}, X\right)$ and $\mathcal{F}_0$ be the subset of $\mathcal{F}$ such that H_{0X} holds. We impose the following conditions on $\mathcal{F}$ and $\mathcal{F}_0$.

Assumption 3.

(i) *For all* $F \in \mathcal{F}$, $D \perp\!\!\!\perp (X, Y, \psi)$*;*

(ii) *There exists* $M > 0$ *such that* $\widetilde{Y} \in [-M, M]$ *for all* $F \in \mathcal{F}$*. Also,* $\inf_{F \in \mathcal{F}} \mathbb{V}_F \left(\widetilde{Y} \right) > 0$ *and* $0 < \inf_{F \in \mathcal{F}} \mathbb{E}_F [D] \leq \sup_{F \in \mathcal{F}} \mathbb{E}_F [D] < 1$*;*

(iii) *For all* $F \in \mathcal{F}_0$, K_F*, the asymptotic covariance kernel of* $n^{-1/2} \mathrm{Diag} \left(\mathbb{V}_F \left(\widetilde{Y} \right) \right)^{-1/2} \overline{m}_n$ *is in a compact set* $\mathcal{K}_2$ *of the set of all* 2×2 *matrix valued covariance kernels on* $\mathcal{Y} \times \cup_{r \geq 1} \mathcal{G}_r$ *with uniform metric* d *defined by*

$$d(K, K') = \sup_{(y,g,y',g') \in \left(\mathcal{Y} \times \cup_{r \geq 1} H_r \right)^2} \left\| K(y,g,y',g') - K'(y,g,y',g') \right\|.$$

The main result of this section is Theorem 2. It shows that, under Assumption 3, the test $\varphi_{n,\alpha}$ controls the asymptotic size and is consistent over fixed alternatives.

Theorem 2. *Suppose that* $r_n \to \infty$ *and Assumption 3 holds. Then:*

(i) $\limsup_{n \to \infty} \sup_{F \in \mathcal{F}_0} \mathbb{E}_F [\varphi_{n,\alpha}] \leq \alpha$*;*

(ii) *If there exists* $F_0 \in \mathcal{F}_0$ *such that* $\mathcal{L}_{F_0}$ *is nonempty and there exists* (j, y_0, g_0) *in* $\{1, 2\} \times \mathcal{L}_{F_0}$ *such that* $K_{F_0, jj}(y_0, g_0, y_0, g_0) > 0$*, then, for any* $\alpha \in [0, 1/2)$*,*

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \sup_{F \in \mathcal{F}_0} \mathbb{E}_F [\varphi_{n,\alpha}] = \alpha.$$

(iii) If $F \in \mathcal{F} \backslash \mathcal{F}_0$, then $\lim_{n\to\infty} \mathbb{E}_F(\varphi_{n,\alpha}) = 1$.

Theorem 2 (i) is closely related to Theorem 5.1 and Lemma 2 in AS. It shows that the test $\varphi_{n,\alpha}$ controls the asymptotic size, in the sense that the supremum over $\mathcal{F}_0$ of its level is asymptotically lower or equal to α. To prove this result, the key is to establish that, under Assumption 3, the class of transformed unconditional moment restrictions that characterize the null hypothesis satisfies a manageability condition (see Pollard, 1990). Using arguments from Hsu (2016), we then exhibit cases of equality in Theorem 2 (ii), showing that, under mild additional regularity conditions, the test has asymptotically exact size (when letting η tend to zero). Finally, Theorem 2 (iii), which is based on Theorem 6.1 in AS, shows that the test is consistent over fixed alternatives.

Extension to account for aggregate shocks This testing procedure can be easily modified to accommodate unanticipated aggregate shocks. Specifically, using the notation defined in Section 5.2.2, we consider the same test as above after replacing $\widetilde{Y}$ by $\widetilde{Y}_{\widehat{c}} = Dq(Y, \widehat{c}) + (1 - D)\psi$, where $\widehat{c}$ denotes a consistent estimator of c_0. The resulting test is given by $\varphi_{n,\alpha,\widehat{c}} = \mathbb{1}\left\{T(\widehat{c}) > c^*_{n,\alpha}\right\}$ (where $T(\widehat{c})$ is obtained by replacing $\widetilde{Y}$ by $\widetilde{Y}_{\widehat{c}}$ in the original test statistic). Such tests have the same properties as those above under some mild regularity conditions on $q(\cdot, \cdot)$, which hold in particular for the leading examples of additive and multiplicative shocks ($q(y, c) = y - c$ and $q(y, c) = y/c$). We refer the reader to Appendix 5.8.1 for a detailed discussion of this extension.

5.4 Monte Carlo simulations

In the following we study the finite sample performances of the test without covariates through Monte Carlo simulations. The finite sample performances of the version of our test that accounts for covariates are reported and discussed in Appendix 5.8.5.

We suppose that the outcome Y is given by

$$Y = \rho\psi + \varepsilon,$$

with $\rho \in [0,1]$, $\psi \sim \mathcal{N}(0,1)$ and

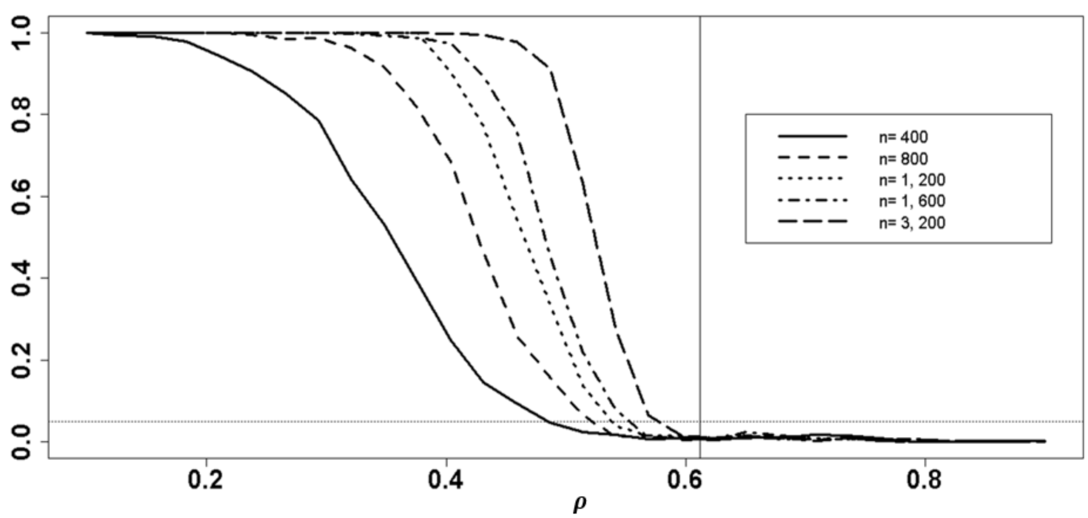
$$\varepsilon = \zeta \left(-\mathbb{1}\{U \leq 0.1\} + \mathbb{1}\{U \geq 0.9\}\right),$$

where ζ, U and ψ are mutually independent, $\zeta \sim \mathcal{N}(2, 0.1)$ and $U \sim \mathcal{U}[0,1]$. In this setup, $\mathbb{E}(Y|\psi) = \rho\psi$ and expectations are rational if and only if $\rho = 1$. But since we observe Y and ψ in two different datasets, there are values of $\rho \neq 1$ for which our test is not consistent. More precisely, we can show that the test is consistent if and only if $\rho \leq \rho^* \simeq 0.616$. Besides, given this data generating process, the naive test $\mathbb{E}(Y) = \mathbb{E}(\psi)$ is not consistent for any ρ, while the RE test based on variances is only able to detect a subset of violations of RE that correspond to $\rho < 0.445$.

To compute our test, we need to choose the tuning parameters b_0, κ, ϵ and η (see Section 5.3 for definitions). As mentioned in Section 5.3, we set $\epsilon = 0.05$ and $\eta = 10^{-6}$, following Andrews and Shi (2017). Andrews and Shi (2013) show that there exists in practice a large range of admissible values for the other tuning parameters parameters. Regarding b_0 and κ, we follow Beare and Shi (2019, Section 4.2) and compute, for a grid of candidate parameters, the rejection rate under the null and under one alternative (namely, $\rho = 0.5$), through Monte Carlo simulations. Then, we set (b_0, κ) so as to maximize the power subject to the constraint that the rejection rate under the null is below the nominal size 0.05. That way, we obtain $b_0 = 0.3$ and $\kappa = 0.001$. The parameter p has a distinct effect, in that its choice does not affect size, at least asymptotically. Rather, this parameter selects to what extent the test aims power at the equality constraint $E(Y - \psi) = 0$ versus the inequalities $E[(y - Y)^+ - (y - \psi)^+] \geq 0$ ($y \in \mathbb{R}$). Setting p to 0.05 leads to slightly higher power in our DGP, but values of p in $[0, 0.31]$ provide similar finite sample performances, with power always greater than 90% of the maximal power.

Results reported in Figure 5-1 show the power curves of the test φ_α for five different sample sizes ($n_Y = n_\psi = n \in \{400; 800; 1,200; 1,600; 3,200\}$) as a function of the parameter ρ, using 800 simulations for each value of ρ. We use 500 bootstrap simulations to compute the critical values of the test.

Several remarks are in order. First, as expected, under the alternative (i.e. for values of $\rho \leq \rho^* = 0.616$), rejection frequencies increase with the sample size n. In particular, for the largest sample size $n = 3,200$, our test always results in rejection of the RE hypothesis for values of ρ as large as .45. Second, in this setting, our test is conservative in the sense that rejection frequencies under the null are smaller than $\alpha = 0.05$, for all sample sizes. This should not necessarily come as a surprise since the test proposed by AS has been shown to be conservative in alternative finite-sample settings (see, *e.g.* Table 1 p.22 in AS for the case of first-order stochastic dominance tests). However, for the version of our test that accounts for covariates and for the data generating process considered in Section 5.8.5 of the Web Appendix, rejection frequencies under the null are very close to the nominal level.



Notes: The vertical line at $\rho \simeq 0.616$ corresponds to the theoretical limit for the rejection of the null hypothesis using our test. The dotted horizontal line corresponds to the 5% level.

Figure 5-1: Power curves.

5.5 Application to earnings expectations

5.5.1 Data

Using the tests developed in Section 5.3, we now investigate whether household heads form rational expectations on their future earnings. We use for this purpose data from the Survey of Consumer Expectations (SCE), a monthly household survey that has been conducted by the Federal Reserve Bank of New York since 2012 (see Armantier, Topa, Van der Klaauw and Zafar, 2017, for a detailed description of the survey, and Kuchler and Zafar, 2019; Conlon, Philossoph, Wiswall and Zafar, 2018; Fuster, Kaplan and Zafar, 2020 for recent articles using the SCE). The SCE is conducted with the primary goal of eliciting consumer expectations about inflation, household finance, labor market, as well as housing market. It is a rotating internet-based panel of about 1,200 household heads, in which respondents participate for up to twelve months.[9] Each month, the panel consists of about 180 entrants, and 1,100 repeated respondents. While entrants are overall fairly similar to the repeated respondents, they are slightly older and also have slightly lower incomes (see Table 1 in Armantier et al., 2017).

Of particular interest for this paper is the supplementary module on labor market expectations. This module is repeated every four months since March 2014. Since March 2015, respondents are asked the following question about labor market earnings expectations (ψ) over the next four months: "What do you believe your annual earnings will be in four months?". Implicit throughout the rest of our analysis is the assumption that these elicited beliefs correspond to the mean of the subjective beliefs distribution.[10] In this module, respondents are also asked about current job outcomes, including their current annual earnings (Y), through the following question: "How much do you make before taxes and other deductions at your [main/current]

[9]Each survey takes on average about fifteen minutes to complete, and respondents are paid $15 per survey completed.

[10]This assumption, while often made in the subjective expectations literature, is *a priori* restrictive. In this application, for the vast majority of the sub-groups of the population, the mean of ψ cannot be statistically distinguished from the one of Y (see Table 5.2 below). This provides empirical support for this assumption.

job, on an annual basis?".

Specifically, we use for our baseline test the elicited earnings expectations (ψ), which are available for two cross-sectional samples of household heads who were working either full-time or part-time at the time of the survey, and responded to the labor market module in March 2015 and July 2015 respectively. We combine this data with current earnings (Y) declared in July 2015 and November 2015 by the respondents who are working full-time or part-time at the time of the survey.[11] This leaves us with a final sample of 2,993 observations, which is composed of 1,565 earnings expectation observations, and 1,428 realized earnings observations. 51% (1,536) of these observations correspond to the sub-sample of respondents who are reinterviewed at least once. We refer to Table 5.1 for additional details on our sample.

Table 5.1: Descriptive statistics of the SCE sample

	Mean	Std. dev.
Male	0.53	0.50
White	0.74	0.43
College degree	0.49	0.46
Low numeracy	0.33	0.47
Tenure $\leq$ 6 months	0.17	0.38
Age	45.8	13.0
ψ (Earnings beliefs)	\$50,592	\$40,889
Y (Realized earnings)	\$52,354	\$38,634

5.5.2 Implementation of the test

We summarize how we implemented the test in practice, either on the overall sample or on each subsample corresponding to the binary covariates in Table 5.1. For each case, we start by winsorizing the distribution of realized earnings (Y) and earnings

[11] Throughout our analysis (with the exception of the number of observations reported in Table 5.2) we use the monthly survey weights of the SCE in order to obtain an estimation sample that is representative of the population of U.S. household heads. See Armantier et al. (2017) for more details on the construction of these weights. We also Winsorize the top 5 percentile of the distributions of realized earnings and earnings beliefs.

beliefs (ψ) at the 95% level.[12] Then, we perform the test without covariates, where we allow for multiplicative aggregate shock and thus test H_{0S}, with $q(y;c) = y/c$.[13] Then, we use the function `test` of our companion R package RationalExp.[14] We choose the same values for the tuning parameters $b_0 = 0.3$ and $\kappa = 0.001$ as in the Monte-Carlo simulations in Section 5.4. We also set $p = 0.05$, $\epsilon = 0.05$, and $\eta = 10^{-6}$. Following Andrews and Shi (2017), the interval $\widehat{\mathcal{Y}}$ is approximated by a grid of length 100 from $\min_{i=1,\dots,n} \widetilde{Y}_i$ to $\max_{i=1,\dots,n} \widetilde{Y}_i$. Finally, we use 5,000 bootstrap simulations to compute the critical values of the test.

5.5.3 Are earnings expectations rational?

In Table 5.2 below, we report the results from the naive test of RE ($\mathbb{E}(Y) = \mathbb{E}(\psi)$), and our preferred test ("Full RE"), where we allow for multiplicative aggregate shocks. We implement the tests both on the overall population and on separate subgroups. The latter approach allows us not only to identify which groups fail to rationalize RE, but also, and importantly, to account for the possibility that aggregate shocks may in fact differ across subgroups.

Several remarks are in order. First, using our test, we reject for the whole population, at any standard level, the hypothesis that agents form rational expectations over their future earnings. Second, we also reject RE (at the 5% level) when we apply our test separately for whites (non-Hispanics) and minorities, as well as low vs. high numeracy test scores.[15]

Third, the results from our test point to beliefs formation being heterogeneous across schooling (college degree vs. no college degree) and tenure (more or less than 6 months spent in current job) levels. In particular, we cannot rule out that the beliefs

[12] We show in Table 5.4 of the Web Appendix that our results are robust to other levels of Winsorization.

[13] In our application, the parameter c is estimated using survey weights from the SCE.

[14] See Section 3 in our user's guide (D'Haultfœuille et al., 2018*a*) for details on this function.

[15] Respondents' numeracy is evaluated in the SCE through five questions involving computation of sales, interests on savings, chance of winning lottery, of getting a disease and being affected by a viral infection. Respondents are then partitioned into two categories: "High numeracy" (4 or 5 correct answers), and "low numeracy" (3 or fewer correct answers).

about future earnings of individuals with more schooling experience correspond to rational expectations with respect to some information set. Similarly, while we reject RE at any standard level for the subgroup of workers who have accumulated less than 6 months of experience in their current job, we can only marginally reject at the 10% level RE for those who have been in their current job for a longer period of time. As such, these findings complement some of the recent evidence from the economics of education and labor economics literatures that individuals have more accurate beliefs about their ability as they progress through their schooling and work careers (see, e.g., Stinebrickner and Stinebrickner, 2012; Arcidiacono, Aucejo, Maurel and Ransom, 2016).

Table 5.2: Tests of RE on annual earnings

	$\mathbb{E}(Y-\psi)/\mathbb{E}(Y)$	Naive RE (p-val)	Variance RE (p-val)	Full RE (p-val)	Number of obs. ψ	Y
All	0.034	0.23	0.71	< 0.001	1,565	1,428
Women	0.059	0.13	0.62	< 0.001	730	649
Men	0.025	0.48	0.58	0.210	835	779
White	0.032	0.31	0.67	0.021	1,200	1,097
Minorities	0.046	0.43	0.60	< 0.006	365	331
College degree	-0.001	0.96	0.50	0.130	1,106	1,053
No college degree	0.093	0.04	0.57	0.013	459	375
High numeracy	0.033	0.28	0.62	0.012	1,158	1,070
Low numeracy	0.055	0.27	0.58	0.022	407	358
Tenure $\leq$ 6 months	0.105	0.24	0.63	< 0.001	271	180
Tenure $>$ 6 months	0.007	0.81	0.65	0.091	1,294	1,248

Notes: "Naive RE" denotes the naive RE test of equality of means between Y and ψ. "Variance RE" denotes the variance RE test where the null hypothesis is the variance of Y being greater or equal than the variance of ψ, once we account for aggregate, multiplicative shocks. "Full RE" denotes the test without covariates, where we test H_{0S} with $q(y,c) = y/c$. We use 5,000 bootstrap simulations to compute the critical values of the Full RE test. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

Fourth, using the naive test of equality of means between earnings beliefs and realizations, one would instead generally not reject the null at any standard levels. The one exception is the subgroup of workers without a college degree, for whom the naive test yields rejection of RE at the 5% level. But, as discussed before, one cannot rule out that such a rejection is due to aggregate shocks.

Even though individuals in the overall sample form expectations over their earnings in the near future that are realistic, in the sense of not being significantly biased, the result from our preferred test shows that earnings expectations are nonetheless not rational. Taken together, these findings highlight the importance of incorporating the additional restrictions of rational expectations that are embedded in our test, using the distributions of subjective beliefs and realized outcomes to detect violations of rational expectations. That the variance test of RE never rejects the null at any standard levels indicates that it is important in practice to go beyond the first moments, and exploit instead the full distributions of beliefs and outcomes to detect departures from rational expectations. These results also suggest that, in order to rationalize the realized and expected earnings data, one should consider alternative models of expectation formation that primarily differ from RE in their third, or higher-order moments.

The results of the direct test of RE on the subsample of individuals who are followed over four months are reported in Table 5.3 below. While these results generally paint a similar picture to the results of our test, there are some differences. In particular, the direct test rejects RE at the 5% level for men and at 1% for individuals with tenure greater than 6 months, whereas we do not reject RE for the former group and only marginally so, at the 10% level, for the latter. The direct test also rejects with less power than our test for certain groups (low numeracy, tenure lower than 6 months, and minorities). This lower power may seem surprising given that the direct test can exploit the joint distribution of (Y, ψ), but is simply due to the important reduction in sample size when focusing on the subsample of individuals who are followed over four months results.

There are also important issues associated with the direct test, which generally warrant caution when interpreting the results from this test. Most importantly, as already discussed in Section 5.2.2, the direct test is not robust to measurement errors on the subjective beliefs ψ. As shown in Proposition S7 in the Web Appendix, it is however possible to derive a restriction on β under RE. Specifically, if ξ_ψ is positively correlated with $\varepsilon + \xi_Y$, we have, under RE,

$$\beta \geq 1 - \frac{1}{1+\underline{\lambda}}, \tag{5.3}$$

where $\underline{\lambda}$ is a lower bound on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$. Table 5.3 also reports the results of tests combining (5.3) with the restrictions on the marginal distributions used in our full RE test. Adding the restriction (5.3) does not change the results for values of signal-to-noise ratio between 5 and 20 (i.e., for noise-to-signal ratios between 5% and 20%). Overall, using the subsample of linked data (Y, ψ) through this additional restriction does not add much to our test, at least once we account for possible measurement errors on the elicited beliefs. Another significant concern with the direct test, and, more generally, the use of linked data on (Y, ψ), is that attrition may be endogenous. We discuss this issue in more details in Appendix 5.8.6.

Table 5.3: Direct test, our test, and combined test of RE on annual earnings

	β	Direct test	Full RE	Combined test		Number of obs.		
Bound on signal/noise $\underline{\lambda}$				5	20			
Implied bound on β				0.833	0.952			
		(p-val)	(p-val)	(p-value)	(p-value)	ψ	Y	(ψ, Y)
All	0.954	0.001	< 0.001	< 0.001	< 0.001	1,565	1,428	768
Women	0.956	0.002	< 0.001	< 0.001	< 0.001	730	649	356
Men	0.960	0.021	0.210	0.276	0.276	835	779	412
White	0.963	0.004	0.021	0.019	0.010	1,200	1,097	596
Minorities	0.928	0.010	0.006	0.007	0.005	365	331	172
College degree	0.974	0.060	0.130	0.182	0.182	1,106	1,053	560
No college degree	0.954	0.044	0.013	0.017	0.017	459	375	208
High numeracy	0.959	0.001	0.012	0.016	0.016	1,158	1,070	573
Low numeracy	0.954	0.094	0.022	0.030	0.030	407	358	195
Tenure $\leq$ 6 months	0.942	0.015	0.001	0.002	0.001	271	180	98
Tenure $>$ 6 months	0.956	0.001	0.091	0.094	0.094	1,294	1,248	670

Notes: "Direct test" denotes the direct test of RE when (ψ, Y) is observed. β is the coefficient of the regression of Y on ψ in that case. "Full RE" denotes the test without covariates, where we test H_{0S} with $q(y, c) = y/c$. We use 5,000 bootstrap simulations to compute the critical values of the Full RE test. "Combined RE test" denotes the test without covariates, where we test H_{0S} with $q(y, c) = y/c$, which is the "Full RE" test, combined with the additional restriction $\beta \geq 1 - 1/(1 + \underline{\lambda})$, where $\underline{\lambda}$ is an a priori bound on the signal-to-noise ratio. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

Coming back to our test, the rejection of RE for the overall population but also for most of the subpopulations are, in view of Proposition 4, unlikely to be due to data quality issues. In that sense, these results may be seen as robust evidence against the RE hypothesis for individual earnings, at least in this context. As a result, conclusions of behavioral models based on the assumption that agents form rational expectations about their future earnings may be misleading. Exploring this important question requires one to go beyond testing though, by quantifying the extent to which model predictions are actually sensitive to the violations from rational expectations that have been detected with our test. We investigate this issue in D'Haultfœuille et al. (2018*b*) in the context of a life-cycle consumption model.

5.6 Conclusion

In this paper, we develop a new test of rational expectations that can be used in a broad range of empirical settings. In particular, our test only requires having access to the marginal distributions of realizations and subjective beliefs. As such, it can be applied in frequent cases where realizations and beliefs are observed in two separate datasets, or only observed for a selected sub-population. By bypassing the need to link beliefs to future realizations, our approach also enables to test for rational expectations without having to wait until the outcomes of interest are realized and made available to researchers. We establish that whether one can rationalize rational expectations is equivalent to the distribution of realizations being a mean-preserving spread of the distribution of beliefs, a condition which can be tested using recent tools from the moment inequalities literature. We show that our test can easily accommodate covariates and aggregate shocks, and, importantly for practical purpose, is robust to some degree of measurement errors on the elicited beliefs. We apply our method to test for rational expectations about future earnings, using data from the Survey of Consumer Expectations. While individuals tend to be right on average about their future earnings, our test strongly rejects rational expectations.

Beyond testing, in this application as in any other situations where rational expectations are violated, a natural next step is to evaluate the deviations from rational expectations that one can rationalize from the available data. In the context of structural analysis, a central question then becomes to which extent the main predictions of the model are sensitive to those departures from rational expectations. We explore this important issue and propose in D'Haultfœuille et al. (2018*b*) a tractable sensitivity analysis framework on the assumed form of expectations.

5.7 Proofs of the equivalence results

5.7.1 Proof of Lemma 1

Under H_0, there exist Y', ψ' and $\mathcal{I}'$ such that $Y' \sim Y$, $\psi' \sim \psi$, $\sigma(\psi') \subset \mathcal{I}'$ and $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Then, by the law of iterated expectations,

$$\mathbb{E}[Y'|\psi'] = \mathbb{E}\left[\mathbb{E}\left[Y'|\mathcal{I}'\right]|\psi'\right] = \mathbb{E}\left[\psi'|\psi'\right] = \psi'.$$

Conversely, if there exists (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$, let $\mathcal{I}' = \sigma(\psi')$. Then $\psi' = \mathbb{E}\left[Y'|\psi'\right] = \mathbb{E}\left[Y'|\mathcal{I}'\right]$ and H_0 holds.

5.7.2 Proof of Theorem 1

(i) $\Leftrightarrow$ (iii). By Strassen's theorem (Strassen, 1965, Theorem 8), the existence of (Y, ψ) with margins equal to F_Y and F_ψ and such that $\mathbb{E}\left[Y|\psi\right] = \psi$ is equivalent to $\int f dF_\psi \leq \int f dF_Y$ for every convex function f. By, e.g., Proposition 2.3 in Gozlan et al. (2018), this is, in turn, equivalent to (iii).

(ii) $\Leftrightarrow$ (iii). By Fubini-Tonelli's theorem, $\int_{-\infty}^{y} F_Y(t)dt = \mathbb{E}\left[\int_{-\infty}^{y} \mathbb{1}\{t \geq Y\}dt\right] = \mathbb{E}\left[(y - Y)^+\right]$. The same holds for ψ. Hence, $\Delta(y) \geq 0$ for all $y \in \mathbb{R}$ is equivalent to $\mathbb{E}\left[(y - Y)^+\right] \geq \mathbb{E}\left[(y - \psi)^+\right]$ for all $y \in \mathbb{R}$. The result follows.

5.7.3 Proof of Proposition 1

First, by Jensen's inequality, we obtain

$$\mathbb{E}[(y_0 - Y)^+|\psi] \geq (y_0 - \mathbb{E}(Y|\psi))^+ = (y_0 - \psi)^+.$$

Moreover, $\Delta(y_0) = 0$ implies that $\mathbb{E}((y_0 - Y)^+) = \mathbb{E}((y_0 - \psi)^+)$. Hence, almost surely, we have

$$\mathbb{E}[(y_0 - Y)^+|\psi] = (y_0 - \psi)^+.$$

Equality in the Jensen's inequality implies that the function is affine on the support of the random variable. Therefore, for almost all u, we either have $\text{Supp}(Y|\psi = u) \subset [y_0, \infty)$ or $\text{Supp}(Y|\psi = u) \subset (-\infty, y_0]$. Because $\mathbb{E}[Y|\psi] = \psi$, $\text{Supp}(Y|\psi = u) \subset [y_0, \infty)$ for almost all $u > y_0$ and $\text{Supp}(Y|\psi = u) \subset (-\infty, y_0]$ for almost all $u < y_0$. Then, for all $\tau \in (0,1)$, $F^{-1}_{Y|\psi}(\tau|u) \geq y_0$ for almost all $u \geq y_0$ and $F^{-1}_{Y|\psi}(\tau|u) \leq y_0$ for almost all $u \leq y_0$. Thus, for all $\tau \in (0,1)$, by continuity of $F^{-1}_{Y|\psi}(\tau|\cdot)$, $F^{-1}_{Y|\psi}(\tau|y_0) = y_0$. This implies that $Y|\psi = y_0$ is degenerate.

5.7.4 Proof of Proposition 2

We first prove that H_{0X} is equivalent to the existence of (Y', ψ') such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$ and $\mathbb{E}((Y'|\psi', X) = \psi'$. First, under H_{0X}, there exists $(Y', \psi', \mathcal{I}')$ such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$, $\sigma(\psi', X) \subset \mathcal{I}'$ and $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Then

$$\mathbb{E}[Y'|\psi', X] = \mathbb{E}\left[\mathbb{E}\left[Y'|\mathcal{I}'\right]|\psi', X\right] = \mathbb{E}\left[\psi'|\psi', X\right] = \psi'.$$

Conversely, if there exists (Y', ψ') such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$ and $\mathbb{E}(Y'|\psi', X) = \psi'$, let $\mathcal{I}' = \sigma(X', \psi')$. Then $\psi' = \mathbb{E}(Y'|\psi', X) = \mathbb{E}(Y'|\mathcal{I}')$ and H_{0X} holds. The proposition then follows as Theorem 1.

5.7.5 Proof of Proposition 4

For all y, $\xi \mapsto \mathbb{E}[(y - \psi - \xi)^+]$ is decreasing and convex. Then, because F_{ξ_ψ} dominates at the second order $F_{\xi_Y + \varepsilon}$, we have

$$\int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\varepsilon + \xi_Y}(\xi) \geq \int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\xi_\psi}(\xi).$$

As a result, for all y, we obtain

$$\begin{aligned}
\mathbb{E}\left[\left(y-\widehat{Y}\right)^{+}\right] &= \int \mathbb{E}\left[(y-\psi-\varepsilon-\xi_Y)^{+}\left|\varepsilon+\xi_Y=\xi\right.\right] dF_{\varepsilon+\xi_Y}(\xi) \\
&= \int \mathbb{E}\left[(y-\psi-\xi)^{+}\right] dF_{\varepsilon+\xi_Y}(\xi) \\
&\geq \int \mathbb{E}\left[(y-\psi-\xi)^{+}\right] dF_{\xi_\psi}(\xi) \\
&= \mathbb{E}\left[(y-\widehat{\psi})^{+}\right].
\end{aligned}$$

Moreover, $\mathbb{E}\left(\widehat{Y}\right)=\mathbb{E}\left(\widehat{\psi}\right)$. By Theorem 1, $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0.

5.8 Appendices

In this web appendix, we first establish additional results with aggregate shocks. Then, we discuss the robustness of linear regressions for testing rational expectations (RE) when expectations and realizations of the variable of interest are jointly observed but measured with errors. Third, we consider tests when only rounded expectations are observed. Fourth, we develop tests when the two samples are not representative of the same population. Fifth, we present additional simulations, with covariates. Sixth, we display additional material on the application. The last section gathers all remaining proofs.

5.8.1 Additional results with aggregate shocks

Statistical tests in the presence of aggregate shocks

In this appendix, we show how to adapt the construction of the test statistic and obtain similar results as in Theorem 2 in the presence of aggregate shocks. As explained in Section 5.2.2, we mostly have to replace $\widetilde{Y}$ by $\widetilde{Y}_c = Dq\left(\widetilde{Y}, c\right) + (1-D)\psi$. Because we include covariates here, as in Section 5.3, c is actually a function of X. Also, the true function c_0 has to be estimated. We let $\widehat{c}$ denote such a nonparametric estimator, which is based on $\mathbb{E}[q(Y, c_0(X))|X] = \mathbb{E}[\psi|X]$. When $q(y,c) = y - c$ or $q(y,c) = y/c$, we get respectively $c_0(X) = \mathbb{E}(Y|X) - \mathbb{E}(\psi|X)$ and $c_0(X) = \mathbb{E}(Y|X)/\mathbb{E}(\psi|X)$, and $\widehat{c}$ is easy to compute using nonparametric estimators of $\mathbb{E}(Y|X)$ and $\mathbb{E}(\psi|X)$.

Because in Proposition 3 (ii) we do not test for a moment equality anymore, $m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)$ reduces to $m_1\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)$. We let hereafter $\overline{m}_n(g,y) = \sum_{i=1}^n m_1\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)/n$. In the test statistic T, we replace, for $(y,g) \in \mathcal{Y} \times \cup_{r\geq 1}\mathcal{G}_r$, $\overline{\Sigma}_n(g,y)$ by $\overline{\Sigma}_n(g,y) = \widehat{\Sigma}_n(g,y) + \epsilon\text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}_{\widehat{c}}\right), \widehat{\mathbb{V}}\left(\widetilde{Y}_{\widehat{c}}\right)\right)$, where $\widehat{\Sigma}_n(g,y)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}_{\widehat{c}}\right)$ are respectively the sample covariance matrix of $\sqrt{n}\overline{m}_n(g,y)$ and the empirical variance of $\widetilde{Y}_{\widehat{c}}$. The last difference with the test considered in Section 5.3 is that when using the bootstrap to compute the critical value, we also have to re-estimate c_0 in the bootstrap sample.

We obtain in this context a result similar to Theorem 2 above, under the regularity conditions stated in Assumption 4. We let hereafter $\mathcal{C}_s\left([0,1]^{d_X}\right)$ denote the space of continuously differentiable functions of order s on $[0,1]^{d_X}$ that have a finite norm $\|c\|_{s,\infty} = \max_{|\boldsymbol{k}|\leq s} \sup_{x\in[0,1]^{d_X}} \left|c^{(\boldsymbol{k})}(x)\right|$. We also let, for any function f on a set $\mathcal{G}$, $\|f\|_{\mathcal{G}} = \sup_{x\in\mathcal{G}} |f(x)|$. Finally, when the distribution of $\left(D, \widetilde{Y}, X\right)$ is F, K_F denotes the asymptotic covariance kernel of $n^{-1/2}\text{Diag}\left(\mathbb{V}\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}\overline{m}$.

Assumption 4. *(i) $\widehat{c}$ and c_0 belong to $\mathcal{C}_s\left([0,1]^{d_X}\right)$, with $s \geq d_X$. Moreover, $\|\widehat{c} - c_0\|_{[0,1]^{d_X}} = o_P(1)$.*

(ii) For all $y \in \mathcal{Y}$, q is Lipschitz on $\mathcal{Y}\times[-C,C]$ for some $C > \|c_0\|_{[0,1]^{d_X}}$. Moreover, $\sup_{(y,c)\in\mathcal{Y}\times[-C,C]} |q(y,c)| \leq M_0$;

(iii) For all $c \in \mathbb{R}$, the function $q(\cdot,c) : \mathcal{Y} \to \mathcal{Y}$ is bijective and its inverse $q^I(\cdot,c)$ is Lipschitz on $\mathcal{Y}$;

(iv) $F_{\psi|X}(\cdot|x)$, $F_{Y|X}(\cdot|x)$ are Lipschitz on $\mathcal{Y}$ uniformly in $x \in [0,1]^{d_X}$ with constants $Q_{F,1}$ satisfying $\sup_{F\in\mathcal{F}_0} Q_{F,1} \leq \overline{Q}_1 < \infty$. Also, $F_{q(\psi,c(X))}$, $F_{q(Y,c(X))}$ are Lipschitz on $[-M_0, M_0]$ with constants $Q_{F,2}$ satisfying $\sup_{F\in\mathcal{F}_0} Q_{F,2} \leq \overline{Q}_2 < \infty$;

(iv) $\inf_{F\in\mathcal{F}} \mathbb{V}_F\left[\widetilde{Y}_c^2\right] > 0$ and $\epsilon_0 \leq \inf_{F\in\mathcal{F}} \mathbb{E}_F[D] \leq \sup_{F\in\mathcal{F}} \mathbb{E}_F[D] \leq 1-\epsilon_0$ for some $\varepsilon_0 \in (0, 1/2)$. Also, $\widehat{\mathbb{V}}_F\left[\widetilde{Y}_{\widehat{c}}^2\right]$ is a consistent estimator of $\mathbb{V}_F\left[\widetilde{Y}_c^2\right]$.

Part (i) imposes some regularity conditions on c_0 and its nonparametric estimator $\widehat{c}$. It is possible to check such regularity conditions on $\widehat{c}$ with kernel or series estimators of $\mathbb{E}(Y|X)$ and $\mathbb{E}(\psi|X)$. Parts (ii) and (iii) also hold when $q(y,c) = y - c$ and $q(y,c) = q(y)/c$, by imposing in the second case that c belongs to a compact subset of $(0,\infty)$. Proposition 5 shows that under these conditions, the test has asymptotically correct size.

Proposition 5. *Suppose that $r_n \to \infty$ and that Assumptions 3 and 4 hold. Then (i) in Proposition 2 holds, replacing $\varphi_{n,\alpha}$ by $\varphi_{n,\alpha,\widehat{c}}$.*

Results like (ii) and (iii) in Proposition 2 could also be obtained under the conditions of Proposition 5, modifying directly the proof of Proposition 2.

Impossibility results with more flexible effects of aggregate shocks

We show here that restrictions in the way aggregate shocks affect the outcome are needed to be able to reject RE with F_Y and F_ψ. We consider for that purpose the following model:

$$Y = \sum_{k=0}^{K} C_k V^k + \varepsilon, \tag{5.4}$$

where V is $\mathcal{I}$-measurable and the individual shock ε satisfies $E[\varepsilon|\mathcal{I}] = 0$. The vector $C := (C_0, ..., C_K)'$ represents aggregate shocks, which is assumed to be independent of $\mathcal{I}$, with support $\mathbb{R}^{K+1}$. We also assume that $E(C) = (0, 1, 0, ..., 0)'$, so that $V = \mathbb{E}[Y|\mathcal{I}]$ and under RE, $\psi = V$. Let $Q_c(y) = \sum_{k=0}^{K} c_k y^k$. Then $\mathbb{E}(Y|C = c, \mathcal{I}) = Q_c(V)$ and under RE, we have

$$\mathbb{E}(Y|C = c, \mathcal{I}) = Q_c(\psi).$$

Hence, as in Section 5.2.2, we consider the following hypothesis:

H_{0SK} : there exist random variables (Y', ψ'), a sigma-algebra $\mathcal{I}'$ and $c \in \mathbb{R}^{K+1}$ such that

$\sigma(\psi') \subset \mathcal{I}'$, $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}\left[Y'|\mathcal{I}'\right] = Q_c(\psi')$.

The following proposition is a negative result on the possibility to test for H_{0SK}.

Proposition 6. *Suppose that F_Y and F_ψ are continuous with supports that are bounded intervals. For any $\eta > 0$, there exists $K > 0$ and F, with $\sup_{u\in\mathbb{R}} |F(u) - F_\psi(u)| < \eta$, such that H_{0SK} holds with Y and $\widetilde{\psi} \sim F$ (instead of ψ).*

Proposition 6 states that as K grows large, the set of cdfs F_Y and F_ψ satisfying H_{0SK} (and thus RE in Model (5.4)) becomes arbitrarily close, for the Kolmogorov-Smirnov metric, to the set of of cdfs F_Y and F_ψ that do not satisfy H_{0SK}. In other words, $\cup_{K\in\mathbb{N}}\mathrm{H}_{0SK}$ is dense in the set of all continuous cdfs having bounded interval as supports. When combined with Theorem 2 in Bertanha and Moreira (2020), this implies that there does not exist any almost-surely continuous test of $\cup_{K\in\mathbb{N}}\mathrm{H}_{0SK}$ that has non-trivial power.

A similar, negative result holds if aggregate shocks are allowed to vary with respect to unobserved, individual-specific variables. For instance, shocks may be sector-specific, but sectors may be unobserved in the data. To show such an impossibility result, consider the following model:

$$Y = q(C, U) + V + \varepsilon,$$

where both U and V are $\mathcal{I}-$measurable, C is an aggregate shock independent of $\mathcal{I}$ and the individual shock ε satisfies $E[\varepsilon|\mathcal{I}] = 0$. Thus, aggregate shocks affect the outcome in an additive way, but heterogeneously across individuals, depending on their U, which is assumed to be unobserved by the econometrician and can thus depend on V in a flexible way. We assume without loss of generality that $E[q(C, U)|\mathcal{I}] = 0$, so that $\psi = V$ under RE. Let us also assume that $q(u, c) = \sum_{k=0}^{K} c_k u^k$ and $U = \xi V$, with $\xi > 0$, $\xi \perp\!\!\!\perp V$ and $\mathbb{E}[\xi^k] < \infty$ for all $k \leq K$. Let $C'_k = E[\xi^k]C_k$ if $k \neq 1$, $C'_1 = E[\xi]C_1 - 1$ and $C' = (C'_0, ..., C'_K)'$. Then, under RE,

$$\mathbb{E}[Y|C' = c', \mathcal{I}] = \sum_{k=0}^{K} c'_k \psi^k.$$

Moreover, if $\text{Supp}(C) = \mathbb{R}^{K+1}$, we also have $\text{Supp}(C') = \mathbb{R}^{K+1}$, and no constraint is imposed on c'.[16] As a result, we are led again to test H_{0SK}, and the same negative result as above holds.

5.8.2 Tests based on linear regressions with measurement errors

We suppose here to observe both $(\widehat{Y}, \widehat{\psi})$ satisfying (5.1). In this framework, we study the restrictions that RE entail on the coefficient β of the (theoretical) linear regression of $\widehat{Y}$ on $\widehat{\psi}$.

Proposition 7. *1. For any values of $(\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}), Cov(\widehat{Y}, \widehat{\psi}))$ such that $\mathbb{V}(\widehat{Y}) > \mathbb{V}(\widehat{\psi})$, there exists a DGP compatible with this triple, satisfying (5.1), for which*

[16] $E[q(C, U)|\mathcal{I}] = 0$ implies that $E[C_k] = 0$ for $k = 0, ..., K$, but it does not restrict the set of possible c'_k.

RE hold and such that $\varepsilon + \xi_Y \perp\!\!\!\perp \psi$ *and* F_{ξ_ψ} *dominates at the second order* $F_{\xi_Y + \varepsilon}$.

2. *If* $\beta < 1 - 1/(1 + \underline{\lambda})$ *for some* $\underline{\lambda} \geq 0$, *there exists no DGP compatible with this value of* β, *satisfying* (5.1), *for which RE hold and such that corr*$(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ *and* $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi) \geq \underline{\lambda}$.

The first result is a negative one. It implies that without further restrictions than those already imposed in Proposition 4, the regression of $\widehat{Y}$ on $\widehat{\psi}$ does not bring any additional restriction related to RE. The second result, on the other hand, shows that if one assumes a positive correlation between ξ_ψ and $\xi_Y + \varepsilon$ and a lower bound on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$, then β is bounded from below under RE. The restriction $\text{corr}(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ seems reasonable. First, given that the shocks ε cannot be anticipated, it is natural to assume that $\text{corr}(\xi_\psi, \varepsilon) = 0$. It then follows that the assumption $\text{corr}(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ holds if the measurement errors on Y and ψ are positively correlated. This would typically happen, for instance, if individuals report their expectations and realized earnings omitting in both cases some components of their earnings, or if they instead overstate their realized earnings, and their expectations accordingly.

This proposition just focuses on the linear regression of $\widehat{Y}$ on $\widehat{\psi}$, since this regression has been very often used to test for RE. This means, however, that there may in principle be additional restrictions on the joint distribution of $(\widehat{Y}, \widehat{\psi})$ implied by RE.

5.8.3 Tests with rounding practices

We have considered in Section 5.2.2 the possibility of measurement errors on ψ. Another source of uncertainty on ψ is rounding. Rounding practices by interviewees are common. A way to interpret these practices is that in situations of ambiguity, individuals may only be able to bound the distribution of their future outcome Y (Manski, 2004). If individuals round at 5% levels, for instance, an answer $\psi = 0.05$ for the beliefs about percent increase of income should then only be interpreted as $\psi \in [0.025, 0.075]$. Another case where only bounds on ψ are observed is when questions to elicit subjective expectations take the following form: “What do you think is the percent chance that your own $[Y]$ will be below $[y]$?", for a certain grid of y. If 0 and 100 are always observed, or if we assume that the support of subjective distributions is included in $[\underline{y}, \overline{y}]$, we can still compute bounds on ψ.[17] In such cases, we only observe (ψ_L, ψ_U), with $\psi_L \leq \psi \leq \psi_U$. For a thorough discussion of this issue, and especially of how to infer rounding practices, see Manski and Molinari (2010).

In this setting, rationalizing rational expectations is less stringent than in our baseline set-up since the constraints on the distribution of ψ are weaker. Formally, the null hypothesis takes the following form:

$$\mathrm{H}_{0B}: \ \exists (Y', \psi', \mathcal{I}') : \ \sigma(\psi') \subset \mathcal{I}', \ Y' \sim Y, \ F_{\psi_U} \leq F_{\psi'} \leq F_{\psi_L} \text{ and } \mathbb{E}(Y'|\mathcal{I}') = \psi'.$$

To obtain an equivalent formulation to H_{0B}, a natural idea would be to fix a candidate cdf $F \in [F_{\psi_U}, F_{\psi_L}]$ for F_ψ and apply Theorem 1 with this F. Then, letting $\Delta_F(y) = \int_{-\infty}^{y} F_Y(t) - F(t)dt$ and $\delta_F = \mathbb{E}(Y) - \int u dF(u)$, H_{0B} would hold as long as for some $F \in [F_{\psi_U}, F_{\psi_L}]$, $\Delta_F(y) \geq 0$ for all $y \in \mathbb{R}$ and $\delta_F = 0$. In practice though, directly checking whether such a distribution exists would be very difficult. Fortunately, we show in the following proposition that it is in fact sufficient to check that these conditions hold for a specific candidate distribution. To define the cdf of this distribution, we introduce, for all $b \in \mathbb{R}$, the random variables

$$\psi^b = \psi_U \mathbb{1}\{\psi_U < b\} + \max(b, \psi_L)\mathbb{1}\{\psi_U \geq b\}.$$

We also let $\psi^{-\infty} = \psi_L$ and $\psi^{\infty} = \psi_U$. The cdf of ψ^b is then $F^b(t) = F_{\psi_U}(t)\mathbb{1}\{t < b\} + F_{\psi_L}(t)\mathbb{1}\{t \geq b\}$, for all $b \in \overline{\mathbb{R}}$. We let $\mathcal{F}_B = \{F^b, b \in \overline{\mathbb{R}}\}$ denote the set of all such cdfs.

Assumption 5. $\mathbb{E}(|Y|) < \infty$, $\mathbb{E}(|\psi_L|) < \infty$ *and* $\mathbb{E}(|\psi_U|) < \infty$.

Proposition 8. *Suppose that Assumption 5 holds. First, if* $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, *there exists a unique* $F^* \in \mathcal{F}_B$ *such that* $\delta_{F^*} = 0$. *Second, the following statements*

[17] Note however that in this case, our approach does not take into account all the information on the subjective distribution.

are equivalent:

(i) H_{0B} *holds.*

(ii) $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$ *and* $\Delta_{F^*}(y) \geq 0$ *for all* $y \in \mathbb{R}$*.*

This test shares some similarities with the test in the presence of aggregate shocks. Specifically, if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, we first identify $b_0 \in \overline{\mathbb{R}}$ such that the candidate belief ψ^{b_0}, which plays a similar role as the modified outcome $q(Y, c_0)$ in the test with aggregate shocks, satisfies the equality constraint $\mathbb{E}[\psi^{b_0}] = \mathbb{E}[Y]$. Noting that the inequality $\Delta_{F^*}(y) \geq 0$ can be rewritten as $\mathbb{E}\left[(y - Y)^+ - \left(y - \psi^{b_0}\right)^+\right] \geq 0$, it follows from (ii) that rationalizing RE in this context (i.e., H_{0B}) is then equivalent to a set of many moment inequality constraints involving the distributions of realizations Y and candidate belief ψ^{b_0}.

5.8.4 Tests with sample selection in the datasets

We consider here cases where the two samples are not representative of the same population, or formally, D is not independent of (Y, ψ). This may arise for instance because of oversampling of some subpopulations or differences in nonresponse between the two surveys that are used. We assume instead that selection is conditionally exogenous, that is to say:

$$D \perp\!\!\!\perp (Y, \psi) | X. \tag{5.5}$$

We show how to use a propensity score weighting to handle such a selection. Denote by $p(x) = P(D = 1|X = x) = \mathbb{E}\left[D|X = x\right]$ the propensity score and by

$$W(X) = \frac{D}{p(X)} - \frac{1 - D}{1 - p(X)}.$$

The law of iterated expectations combined with Proposition 2 directly yields the following proposition:

Proposition 9. *Suppose that* (5.5) *and Assumption 1 hold. Then* H_{0X} *is equivalent* *to*

$$\mathbb{E}\left[W(X)\left(y-\widetilde{Y}\right)^{+}\middle|X\right]\geq 0$$

for all $y\in\mathbb{R}$ *and* $\mathbb{E}\left[W(X)\widetilde{Y}\middle|X\right]=0.$

This proposition shows that under sample selection, we can build a statistical test of H_{0X} akin to that developed in Section 5.3, by merely estimating nonparametrically $p(X)$. We could consider for that purpose a series logit estimator, for instance. Validity of such a test would follow using very similar arguments as for the test with aggregate shocks considered above.

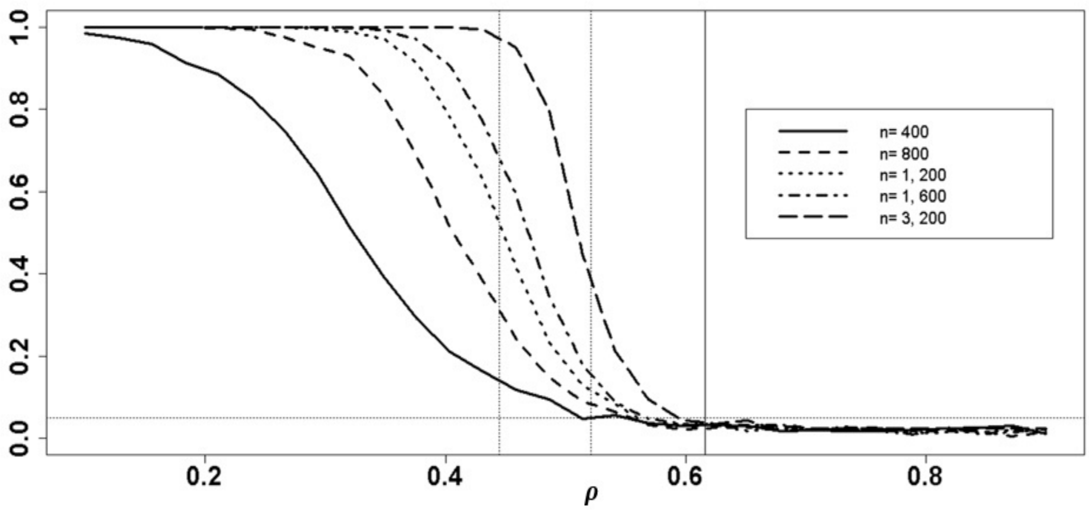
5.8.5 Simulations with covariates

We consider here simulations including covariates. The DGP is similar to that considered in Section 5.4. Specifically, we assume that $Y=\rho\psi+\sqrt{X}\varepsilon$, with $\rho\in[0,1]$, $\psi\sim\mathcal{N}(0,1)$, $X\sim\text{Beta}(0.1,10)$ and

$$\varepsilon=\zeta\left(-\mathbb{1}\{U\leq 0.1\}+\mathbb{1}\{U\geq 0.9\}\right),$$

where $\zeta\sim\mathcal{N}(2,0.1)$ and $U\sim\mathcal{U}[0,1]$. (ψ,ζ,U,X) are supposed to be mutually independent. Like in the test without covariates, we can show that the test with covariates is able to reject RE if and only if $\rho<0.616$. On the other hand, $\mathbb{E}\left[Y|X\right]=\mathbb{E}\left[\psi|X\right]$, so the naive conditional test has no power. The test based on conditional variances rejects only if $\rho<0.445$. Finally, we can show that without using X, our test has power only for $\rho<0.52$. Hence, relying on covariates allows us to gain power for $\rho\in[0.521,0.616)$.

Again, we consider $n_\psi=n_Y=n\in\{400;800;1,200;1,600;3,200\}$, use 500 bootstrap simulations to compute the critical value, and rely on 800 Monte-Carlo replications for each value of ρ and n. We use the same parameters $p=0.05$ and $b_0=0.3$ as above.



Notes: the dotted vertical lines correspond to the theoretical limit for the rejection of the null hypothesis for test based on variance ($\rho \simeq 0.445$), our test without covariates ($\rho \simeq 0.521$) and our tests with covariates ($\rho = 0.616$). The dotted horizontal line corresponds to the 5% level.

Figure 5-2: Power curves for the test with covariates.

Figure 5-2 shows that the RE test with covariates asymptotically outperforms the RE test without covariates. The test exhibits a similar behavior as that without covariates, though, as we could expect, the power converges less quickly to one as n tends to infinity.

5.8.6 Additional material on the application

Effect of the Winsorization on the RE test

Winsorization level	0.95 (p-value)	0.97 (p-value)	0.99 (p-value)
All	< 0.001	< 0.001	0.002
Women	< 0.001	< 0.001	0.001
Men	0.210	0.254	0.342
White	0.021	0.030	0.049
Minorities	0.006	0.007	0.018
College degree	0.130	0.146	0.196
No college degree	0.013	0.012	0.009
High numeracy	0.012	0.017	0.034
Low numeracy	0.022	0.026	0.029
Tenure ≤ 6 months	0.001	0.005	0.009
Tenure > 6 months	0.091	0.118	0.304

Notes: We test H_{0S} with $q(y, c) = y/c$, using 5,000 bootstrap simulations to compute the critical values. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at either the 0.95, 0.97, or 0.99 quantile.

Table 5.4: Full test of RE with different levels of Winsorization

Possibly endogenous attrition in the survey

In addition to measurement errors, another potential issue when using the linked data (Y, ψ) is that attrition may be related to Y itself. This would create a sample selection issue that would invalidate the direct test, even absent any measurement errors. To explore this possibility, Table 5.5 below reports the estimation results from a logit model of attrition on earnings beliefs, gender, race/ethnicity, college degree attainment, numeracy test score, tenure and a (linear) time trend. The main takeaway from this table is that earnings beliefs ψ are significantly associated with attrition, even after controlling for this extensive set of characteristics. This result suggests that individuals for whom we observe both earnings expectations and realizations are likely to earn more than those who are not followed across the two waves. Along the same lines, a Kolmogorov-Smirnov test rejects at the 1% level the equality of the distributions of realized earnings between the whole sample and the subsample that would be used for the direct test. Similarly, we reject the equality of the distributions of expected earnings between these two samples. These results indicate that, in this context, the direct RE test is likely to be misleading. Conversely, attrition is unlikely to be an issue with our test, since we use in each wave the observations of all respondents.[18]

	Intercept	ψ	Male	White	Coll. Degree	Low Num.	Tenure > 6	Trend
All	1.327**	-6.206e-06**	0.046	-0.311	-0.137	-0.141	-0.786**	-0.040
	(0.293)	(1.621e-06)	(0.138)	(0.222)	(0.139)	(0.162)	(0.164)	(0.033)

Notes: 1,565 observations. Significance levels: †: 10%, *: 5%, **: 1%.

Table 5.5: Logit model of attrition

5.8.7 Proofs

Notation and preliminaries

For any set $\mathcal{G}$, let us denote by $l^\infty(\mathcal{G})$ the collection of all uniformly bounded real functions on $\mathcal{G}$ equipped with the supremum norm $\|f\|_{\mathcal{G}} = \sup_{x\in\mathcal{G}} |f(x)|$. Denote by $L^2(F)$ the square integrable space with respect to the measure associated with F, and let $\|\cdot\|_{F,2}$ be the corresponding norm. We let $N(\epsilon, \mathcal{T}, L_2(F))$ denote the minimal number of ϵ-balls with respect to $\|\cdot\|_{F,2}$ needed to cover $\mathcal{T}$. An ϵ-bracket (with respect to F) is a pair of real functions (l, u) such that $l \leq u$ and $\|u - l\|_{F,2} \leq \epsilon$. Then, for any set of real functions $\mathcal{M}$, we let $N_{[]}(\epsilon, \mathcal{M}, L_2(F))$ denote the minimum number of

[18] The one assumption we need to make is that respondents in the surveys used to measure ψ (i.e., those of March and July 2015) are drawn from the same population as those from the surveys used to measure Y (i.e., those of July and November 2015). That there is no significant time trend in the attrition model (Table 5.5) suggests that this assumption is reasonable in this context.

ϵ-brackets needed to cover $\mathcal{M}$. We denote by $\mathcal{G} = (\cup_{r\geq 1}\mathcal{G}_r)$. For $x \in \mathbb{R}^d$, $d > 1$, we denote by $\|x\|_\infty = \max_{j=1,\ldots,d} |x|$.

For a sequence of random variable $(U_n)_{n\in\mathbb{N}}$ and a set $\mathcal{F}_0$, we say that $U_n = O_P(1)$ uniformly in $F \in \mathcal{F}_0$ if for any $\epsilon > 0$ there exist $M > 0$ and $n_0 > 0$ such that $\sup_{F\in\mathcal{F}_0} \mathbb{P}_F(|U_n| > M) < \epsilon$ for all $n > n_0$. Similarly we say that $U_n = o_P(1)$ uniformly in $F \in \mathcal{F}_0$ if for any $\epsilon > 0$, $\sup_{F\in\mathcal{F}_0} \mathbb{P}_F(|U_n| > \epsilon) \to 0$.

Finally, we add stars to random variables whenever we consider their bootstrap versions, as with T^* versus T. We define o_{P^*} and O_{P^*} as above, but conditional on $\left(\widetilde{Y}_i, D_i, X_i\right)_{i=1\ldots n}$. Convergence in distribution conditional on $\left(\widetilde{Y}_i, D_i, X_i\right)_{i=1\ldots n}$ is denoted by $\to_{d^*}$.

Proof of Theorem 2

(i) This is a particular case of Proposition 5 below, with $q(Y, c_0) = Y$. The proof is therefore omitted.

(ii) We show that equality holds for $F_0 \in \mathcal{F}_0$ satisfying the conditions stated in (ii). The proof is divided in three steps. We first prove convergence in distribution of T to S defined below, and conditional convergence of T^* towards the same limit. Then we show that the cdf H of S is continuous and strictly increasing in the neighborhood of its quantile of order $1-\alpha$, for any $\alpha \in (0, 1/2)$. The third step concludes.

1. Convergence in distribution of T and T^*.

Let us introduce some notation. Let $K_{j,j}$ ($j \in \{1,2\}$) be the j-th diagonal element of the covariance kernel K, $\mathcal{S} : (\nu, K) \mapsto (1-p)\left(-\nu_1/K_{1,1}^{-1/2}\right)^{+2} + p\left(\nu_2/K_{2,2}^{-1/2}\right)^2$, $q(r) = \left(r^2+100\right)^{-1}(2r)^{-d_X}$, and

$$\nu_{n,F_0}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_i, X_i, g, y\right) - \mathbb{E}_{F_0}\left[m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)\right]\right).$$

Finally, we define $k_{n,F_0}(y,g) = \sqrt{n}\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\mathbb{E}_{F_0}\left[m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)\right]$,

$$K_{n,F_0}(y,g,y',g') = \text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\widehat{\text{Cov}}\left(\sqrt{n}\overline{m}_n(y,g), \sqrt{n}\overline{m}_n(y',g')\right)\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2},$$
$$\overline{K}_{n,F_0}(y,g,y',g') = K_{n,F_0}(y,g,y',g') + \epsilon\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}\right)\right)\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2},$$

and use the notations $K_{n,F_0}(y,g) = K_{n,F_0}(y,g,y,g)$ and $\overline{K}_{n,F_0}(y,g) = \overline{K}_{n,F_0}(y,g,y,g)$.

We have, by definition of T,

$$T = \sup_{y\in\mathcal{Y}}\sum_{(a,r):r\in\{1,\ldots,r_n\},a\in A_r} q(r)\mathcal{S}\left(\nu_{n,F_0}(y,g_{a,r}) + k_{n,F_0}(y,g_{a,r}), \overline{K}_{n,F_0}(y,g_{a,r})\right).$$

To characterize the distribution of T (resp. T^*), we first prove the convergence of ν_{n,F_0} and $K_{n,F_0}(y,g_{a,r})$ (resp. ν^*_{n,F_0} and $K^*_{n,F_0}(y,g_{a,r})$). For those purposes, we use a class of functions which is a general form taken by m_1 defined in (5.2), namely, for any $0 < N_1 < M_1$,

$$\mathcal{M}_0 = \{f_{y,\phi_1,\phi_2,g}(\widetilde{y},x,d) = \left(d\phi_1(y-\widetilde{y})^+ - (1-d)\phi_2(y-\widetilde{y})^+\right)g(x),$$
$$(y,\phi_1,\phi_2,g) \in \mathcal{Y}\times[N_1,M_1]^2\times\mathcal{G}\}.$$

Remark that $\mathcal{M}_0$ is a particular case of classes $\mathcal{M}$ defined in (5.9) below. Then, by the proof of Proposition 5 below, Assumptions PS1 and PS2 in AS are satisfied. Thus, the assumptions of Lemma D.2 in AS hold as well. This entails that Assumptions PS4 and PS5 in AS hold. Namely, there exists a Gaussian process ν_{F_0} such that

- $\nu_{n,F_0} \to_d \nu_{F_0}$ and $\nu^*_{n,F_0} \to_{d^*} \nu_{F_0}$;

- For all $r \in \mathbb{N}$ and $(y,g) \in \mathcal{Y}\times\mathcal{G}_r$, $\overline{K}_{n,F_0}(y,g) \to_P K_{F_0}(y,g)+\epsilon I_2$ and $K^*_{n,F_0}(y,g) \to_{P^*} K_{F_0}(y,g) + \epsilon I_2$, where I_2 is the 2×2 identity matrix.

Moreover, letting $k_{F_0}(y,g)$ denote the limit in probability of $k_{n,F_0}(y,g)$, we have $k_{F_0}(y,g) = 0$ if $(y,g) \in \mathcal{L}_{F_0}$ and ∞ otherwise. Note that by assumption, the set $\mathcal{L}_{F_0}$ is nonempty.

Thus, using (D.11) in the proof of Theorem D.3. in AS, which is based on the uniform continuity of the function $\mathcal{S}$ in the sense of Assumption S2 therein, we have, under F_0,

$$\begin{aligned} T \to_d & \sup_{y\in\mathcal{Y}} \sum_{(a,r)\in A_r\times\mathbb{N}} \mathcal{S}\left(\nu_{F_0}(y,g_{a,r}) + k_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right) \\ &= S := \sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right), \end{aligned}$$

where the equality follows by definition of $\mathcal{S}$ and $k_{F_0}(y,g)$. Similarly, using Assumption PS5 and (D.11) in AS, replacing T by T^* and quantities $\nu_{n,F_0}(y,g_{a,r})$ and $K_{n,F_0}(y,g_{a,r})$ by their bootstrap counterparts (see the proof of Lemma D.4 in AS) we have $T^* \to_{d^*} S$.

2. The cdf H of S is continuous and strictly increasing in the neighborhood of any of its quantile of order $1-\alpha > 1/2$.

First, the cdf H of S is a convex functional of the Gaussian process ν_{F_0}. Then, as in the proof of Lemma B3 in Andrews and Shi (2013), we can use Theorem 11.1 of Davydov et al. (1998) p.75 to show that H is continuous and strictly increasing at every point of its support except $\underline{r} = \inf\{r\in\mathbb{R} : H(r) > 0\}$. Moreover, for any $r>0$,

$$\begin{aligned} H(r) &\geq \mathbb{P}\left(\sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right) < r\right) \\ &\geq \mathbb{P}\left(\sup_{j\in\{1,2\},(y,a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} \left|(K_{2,F_0,j,j}(y,g_{a,r}) + \epsilon)^{-1/2}\nu_{F_0,j}(y,g_{a,r})\right| < \frac{\sqrt{r/2}}{Q}\right) \\ &> 0, \end{aligned}$$

where $Q = \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r) < \infty$ and we use Problem 11.3 of Davydov et al. (1998) p.79 for the last inequality. This yields $r > \underline{r}$ and H is continuous and strictly increasing on $(0,\infty)$.

Then, we show that for any $\alpha \in (0,1/2)$, the quantile of order $1-\alpha$ of the distribution of S is positive. By assumption, there exists $(y_0, g_0) \in \mathcal{L}_{F_0}$ such that either $K_{F_0,11}(y_0, g_0) > 0$ or $K_{F_0,2}(y_0, g_0) > 0$. This yields

$$\begin{aligned}\mathbb{P}(S > 0) &= 1 - \mathbb{P}\left(\sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y, g_{a,r}), K_{F_0}(y, g_{a,r}) + \epsilon I_2\right) = 0\right)\\ &\geq 1 - \mathbb{P}\left(\nu_{F_0,1}(y_0, g_0) \leq 0,\, \nu_{F_0,2}(y, g_0) = 0\right)\\ &\geq 1 - \min\left\{\mathbb{P}\left(\nu_{F_0,1}(y_0, g_0) \leq 0\right), \mathbb{P}\left(\nu_{F_0,2}(y_0, g_0) = 0\right)\right\}\\ &\geq 1/2. \end{aligned} \tag{5.6}$$

The first inequality holds by definition of the supremum and because $\mathcal{S}$ is nonnegative. To obtain the last inequality, note that either $\nu_{F_0,1}(y_0, g_0)$ is non-degenerate, in which case the first probability is $1/2$ (since $\nu_{F_0,1}(y_0, g_0)$ is normal with zero mean), or $\nu_{F_0,2}(y_0, g_0)$ is non-degenerate, in which case the second probability is 0.

Finally, using that H is strictly increasing on $(0, \infty)$, (5.6) ensures that any quantile of S of order $1 - \alpha$ with $\alpha \in [0, 1/2)$ is positive. Hence, H is continuous and strictly increasing in the neighborhood of any such quantiles.

3. Conclusion.

Using $T^* \to_{d^*} S$ in distribution, Step 2 and Lemma 21.2 in Van der Vaart (2000), we have that for $\eta > 0$, $c^*_{n,\alpha} \to_{d^*} c(1 - \alpha + \eta) + \eta$, where $c(1 - \alpha + \eta)$ is the $(1-\alpha+\eta)$-th quantile of the distribution of S. Because $T \to_d S$ and H is continuous at $c(1 - \alpha + \eta) + \eta > 0$, we obtain that

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \mathbb{P}_{F_0}\left(T > c^*_{n,\alpha}\right) = \alpha.$$

Combined with the inequality of Part (i) above, this yields the result.

(iii) This results follows from Theorem E.1 in AS. First, Assumption SIG2 in AS holds for $\sigma^2_F = \mathbb{V}_F\left(\widetilde{Y}\right)$, following the proof of Lemma 7.2 (b) under Assumption 3-(ii). Second, Assumptions PS4 and PS5 are satisfied using the point (ii) above. Third, Assumptions CI, MQ, S1, S3, S4 in AS are also satisfied by construction of the statistic T. Thus, Theorem E.1 in AS yields the result. $\square$

Proof of Proposition 5

We introduce $\mathbb{E}_{F,c} = \mathbb{E}_F\left[m\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)\right]$ and

$$\nu_{n,F}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{Diag}\left(\widehat{\mathbb{V}}_F\left(\widetilde{Y}_{\widehat{c}}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_{\widehat{c},i}, X_i, g, y\right) - E_{F,\widehat{c}}\right),$$
$$\overline{\nu}_{n,F}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{Diag}\left(\mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_{c_0,i}, X_i, g, y\right) - E_{F,c_0}\right).$$

The proof is based on Theorem 5.1 in AS, hence we have to check that the corresponding assumptions PS1, PS2, and SIG1 hold. Namely, we have to ensure that

- **PS1**: for all sequence $F \in \mathcal{F}$ and all $(d, y', x, g, y, c) \in \{0,1\} \times \mathcal{Y} \times [0,1]^{d_X} \times \mathcal{G}_r \times \mathcal{Y} \times \mathcal{C}_s\left([0,1]^{d_X}\right)$

  $$\left|\frac{m(d, y', x, g, y)}{\mathbb{V}_F\left(\widetilde{Y}_{c,i}\right)}\right| \leq M(d, y', x, g, y) \text{ and } \mathbb{E}_F\left[M\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)^{2+\delta}\right] \leq C < \infty,$$

  where $\delta > 0$ and for some function M;

- **PS2**: for all sequence $F_n \in \mathcal{F}$, the i.i.d triangular array of processes

  $$\mathcal{T}_n^0 = \left\{\frac{m\left(D_i, \widetilde{Y}_{n,c(X_{n,i})}, X_{n,i}, g, y\right)}{\mathbb{V}_{F_n}\left(\widetilde{Y}_{n,c(X_{n,i})}\right)},\ (c, y, g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times \mathcal{G},\ i \leq n,\ n \geq 1\right\}$$

  is manageable with respect to some envelope function U_1 (see Pollard, 1990, p.38 for the definition of a manageable class);

- **SIG1**: for all $\zeta > 0$, $\sup_{F\in\mathcal{F}, c\in\mathcal{C}_s\left([0,1]^{d_X}\right)} \mathbb{P}\left(\left|\widehat{\mathbb{V}}_F\left(\widetilde{Y}_{i,c}\right)/\mathbb{V}_F\left(\widetilde{Y}_{i,c}\right) - 1\right| > \zeta\right) \to 0$.

We proceed in two steps, to handle the fact that c_0 and $\text{Diag}\left(\mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}$ are estimated:

1. We first show that

$$\sup_{F\in\mathcal{F}_0}\sup_{g\in\cup_{r\geq 1}\mathcal{G}_r, y\in\mathcal{Y}}\left\|\nu_{n,F}(y,g)-\overline{\nu}_{n,F}(y,g)\right\|_\infty = o_P(1), \tag{5.7}$$

$$\sup_{F\in\mathcal{F}_0}\sup_{g\in\cup_{r\geq 1}\mathcal{G}_r, y\in\mathcal{Y}}\left\|\nu^*_{n,F}(y,g)-\overline{\nu}^*_{n,F}(y,g)\right\|_\infty = o_{P^*}(1). \tag{5.8}$$

2. Next, we show that m satisfies assumptions PS1, PS2, and that SIG1 in AS also holds for $\sigma_F^2=\mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)$, where $F\in\mathcal{F}$ and $\widehat{\sigma}_n^2=n^{-1}\sum_{i=1}^n\left(\widetilde{Y}_{\widehat{c},i}-n^{-1}\sum_{j=1}^n\widetilde{Y}_{\widehat{c},j}\right)^2$.

1. Proof of (5.7)-(5.8)

We apply the uniform version over $F\in\mathcal{F}_0$ of Theorem 3 in Chen et al. (2003) to a general class of functions to which pertain the moment condition m (see (5.2), with $\widetilde{Y}$ replaced here by $\widetilde{Y}_c=Dq\left(\widetilde{Y},c\right)+(1-D)\psi$ and without the moment equality m_2). Hence, it suffices to verify that Assumptions (3.2) and (3.3) of Theorem 3 in Chen et al. (2003) are satisfied. Let us introduce, for any $0<N_1<M_1$, the classes of functions

$$\mathcal{M}_1=\left\{f_{c,y,\phi,g}\left(\widetilde{y},x\right)=\phi\left(y-q\left(\widetilde{y},c(x)\right)\right)^+g(x),(c,y,\phi,g)\in\mathcal{C}_s\left([0,1]^{d_X}\right)\times\mathcal{Y}\times[N_1,M_1]\times\mathcal{G}\right\}, \tag{5.9}$$

$$\mathcal{M}_2=\left\{f_{c,y,\phi,g}\left(\widetilde{y},x\right)=\phi\left(y-\widetilde{y}\right)^+g(x),\ (c,y,\phi,g)\in\mathcal{C}_s\left([0,1]^{d_X}\right)\times\mathcal{Y}\times[N_1,M_1]\times\mathcal{G}\right\},$$

$$\begin{aligned}\mathcal{M}=&\{f_{c,y,\phi_1,\phi_2,g}\left(\widetilde{y},x,d\right)=\left(dg_{c,y,\phi_1,g}-(1-d)q_{c,y,\phi_2,g}\right)\left(\widetilde{y},x\right),\ g\in\mathcal{M}_1,\ q\in\mathcal{M}_2,\\ &(c,y,\phi_1,\phi_2,g)\in\mathcal{C}_s\left([0,1]^{d_X}\right)\times\mathcal{Y}\times[N_1,M_1]^2\times\mathcal{G}\}.\end{aligned}$$

Note that ϕ_1, ϕ_2, and c in the class $\mathcal{M}$ denote components of m that are estimated.

Consider the space $\mathcal{C}_s\left([0,1]^{d_X}\right)\times\mathcal{Y}\times[N_1,M_1]^2\times\mathcal{G}$ equipped with the norm

$$\|(c,y,\phi_1,\phi_2,g)\|=\max\left\{\|c\|_{[0,1]^{d_X}},|y|,|\phi_1|,|\phi_2|,\|g\|_{[0,1]^{d_X}}\right\}.$$

For $v=(c,y,\phi_1,\phi_2,g),v'=(c',y',\phi_1',\phi_2',g')\in\mathcal{C}_s\left([0,1]^{d_X}\right)\times\mathcal{Y}\times[N_1,M_1]^2\times\mathcal{G}$ and $(\widetilde{y},x,d)\in\mathcal{Y}\times[0,1]^{d_X}\times\{0,1\}$, we have, by the triangular inequality and Assumptions

4-(i) and 4-(v),

$$\begin{aligned}
\left|f_v\left(\widetilde{y},x,d\right)-f_{v'}\left(\widetilde{y},x,d\right)\right| \leq & \left|g_{c,y,\phi_1,g}\left(\widetilde{y},x\right)-g_{c',y',\phi_1',g'}\left(\widetilde{y},x\right)\right| \\
& +\left|q_{c,y,\phi_2,g}\left(\widetilde{y},x\right)-q_{c',y',\phi_2',g'}\left(\widetilde{y},x\right)\right| \\
\leq & (M+M_0)\left(\left|\phi_1-\phi_2'\right|+\left|\phi_2-\phi_2'\right|\right) \\
& +2M_1\left[\left|y-y'\right|+\left|q\left(\widetilde{y},c(x)\right)-q\left(\widetilde{y},c'(x)\right)\right|\right] \\
& +2M_0M_1\Big[\left|\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y\right\}-\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y'\right\}\right| \\
& \quad +\left|\mathbb{1}\left\{q\left(\widetilde{y},c(x)\right)\leq y'\right\}-\mathbb{1}\left\{q\left(\widetilde{y},c'(x)\right)\leq y'\right\}\right| \\
& \quad +\left|g(x)-g'(x)\right|\Big].
\end{aligned}$$

Denote by $K_q>0$ the Lipschitz constant of $q(\widetilde{y},.)$. Then, by convexity of $x\mapsto x^2$, we obtain

$$\begin{aligned}
\frac{1}{7}\left|f_v\left(\widetilde{y},x,d\right)-f_{v'}\left(\widetilde{y},x,d\right)\right|^2 \leq & (M+M_0)^2\left(\left|\phi_1-\phi_1'\right|^2+\left|\phi_2-\phi_2'\right|^2\right) \\
& +4M_1^2\left[\left|y-y'\right|^2+K_q\left\|c-c'\right\|^2_{[0,1]^{d_X}}\right] \\
& +4(M_0M_1)^2\Big[\left|\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y\right\}-\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y'\right\}\right| \\
& \quad +\left|\mathbb{1}\left\{q\left(\widetilde{y},c(x)\right)\leq y'\right\}-\mathbb{1}\left\{q\left(\widetilde{y},c'(x)\right)\leq y'\right\}\right| \\
& \quad +\left\|g-g'\right\|^2_{[0,1]^{d_X}}\Big].
\end{aligned}$$

Fix $\delta>0$. If $\|v-v'\|\leq\delta$, this yields

$$\begin{aligned}
\frac{1}{7}\left|f_v\left(\widetilde{y},x,d\right)-f_{v'}\left(\widetilde{y},x,d\right)\right|^2 \leq & \delta^2\left(2(M+M_0)^2+4M_1^2(1+K_q)+4(M_0M_1)^2\right) \\
& +4(M_0M_1)^2\Big[\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y+\delta\right\}-\mathbb{1}\left\{q(\widetilde{y},c(x))\leq y-\delta\right\} \\
& \quad +\left|\mathbb{1}\left\{\widetilde{y}\leq q^I\left(y',c(x)\right)\right\}-\mathbb{1}\left\{\widetilde{y}\leq q^I\left(y',c'(x)\right)\right\}\right|\Big].
\end{aligned}$$

Next, by Assumption 4-(iv), we obtain

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left\{q\left(\widetilde{Y}, c(X)\right) \leq y+\delta\right\} - \mathbb{1}\left\{q\left(\widetilde{Y}, c(X)\right) \leq y-\delta\right\}\right] \\
&= F_{q\left(\widetilde{Y}, c(X)\right)}(y+\delta) - F_{q\left(\widetilde{Y}, c(X)\right)}(y-\delta) \\
&\leq 2\overline{Q}_2\delta.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathbb{1}\left\{Y \leq q^I\left(y', c(X)\right)\right\} - \mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c'(X)\right)\right\}\right|\right] \\
\leq&\mathbb{E}\left[\mathbb{1}\left\{Y \leq q^I\left(y', c(X)\right) - Q_{F,2}\delta\right\} - \mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c(X)\right) + Q_{F,2}\delta\right\}\right] \\
\leq&\mathbb{E}\left[F_{Y|X}\left(q^I\left(y', c(X)\right) - Q_{q^I}\delta\middle|X\right) - F_{Y|X}\left(q^I\left(y', c(X)\right) + Q_{q^I}\delta\middle|X\right)\right] \\
\leq&2Q_{F,1}Q_{q^I}\delta,
\end{aligned}$$

where Q_{q^I} is the Lipschitz constant of q^I. Thus, by Assumption 4, there exists $Q > 0$ such that

$$\sup_{F\in\mathcal{F}_0} \mathbb{E}\left[\sup_{\|v-v'\|\leq\delta}\left|f_v\left(\widetilde{Y}, X, D\right) - f_{v'}\left(\widetilde{Y}, X, D\right)\right|^2\right] \leq Q\delta. \tag{5.10}$$

Therefore the class $\mathcal{M}$ satisfies Condition (3.2) of Theorem 3 in Chen et al. (2003) uniformly in $F \in \mathcal{F}_0$. Moreover, the class $\mathcal{G}$ is manageable and thus Donsker (see Lemma 3 in Andrews and Shi, 2013). Finally, by Remark 3 (ii) in Chen et al. (2003), $\mathcal{C}_s\left([0,1]^{d_X}\right)$ is also Donsker. Then, $\mathcal{C}_s\left([0,1]^{d_X}\right)$, $\mathcal{Y}$, $[N_1, M_1]$, and $\mathcal{G}$ satisfy Condition (3.3) of Theorem 3 in Chen et al. (2003). The result follows by Theorem 3 in Chen et al. (2003).

2. m satisfies PS1 and PS2 of AS and SIG1 of AS also holds for σ_F^2 and $\widehat{\sigma}_n^2$.

From Assumption 4 (iii) and the proof of Lemma 7.2 (a) in AS, PS1 is satisfied replacing B by $\max(M, M_0)$ in the proof of Lemma 7.2-(a) in AS.

We now show that PS2 in AS also holds. As the result is uniform over $\mathcal{F}_0$, we have to consider sequences for the cdfs F_n of $(D_{n,i}, Y_{n,i}, X_{n,i})_{i=1\ldots n}$ (with $F_n \in \mathcal{F}_0$).

We also define

$$\begin{aligned}\widetilde{Y}_{n,c(X_{n,i})} &= D_{n,i} q\left(Y_{n,i}, c(X_{n,i})\right) + (1-D_{n,i})\psi_{n,i},\\ W_{n,i} &= \frac{D_{n,i}}{\mathbb{E}_{F_n}\left[D_{n,i}\right]} - \frac{1-D_{n,i}}{\mathbb{E}_{F_n}\left[1-D_{n,i}\right]},\\ \sigma^2_{F_n} &= \mathbb{V}_{F_n}\left(\widetilde{Y}_{n,c(X_{n,i})}\right).\end{aligned}$$

Note that by Assumption 3 (iii), $\sigma^2_{F_n} \geq \overline{\sigma} > 0$ for all $F_n \in \mathcal{F}$. Let $(\Omega, \mathbb{F}, F_n)$ be a probability space and let ω denote a generic element in Ω. Showing Assumption PS2 in AS then boils down to prove that for any $0 < N_1 < M_1 := 1/\inf_F \sigma^2_F$, the i.i.d triangular array of processes

$$\mathcal{T}_{1,n,\omega} = \Big\{W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+ g(X_{n,i}),\ (c,y,\phi,g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G},\\ i \leq n, n \geq 1\Big\}$$

is manageable with respect to some envelope function U_1. Lemma 3 in Andrews and Shi (2013) shows that the processes $\{g(X_{n,i}),\ g \in \mathcal{G},\ i \leq n, n \geq 1\}$ are manageable with respect to the constant function 1. Then, using Lemma D.5 in AS, it remains to show that

$$\mathcal{T}'_{1,n,\omega} = \left\{W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+,\ (c,y,\phi) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1], i \leq n, n \geq 1\right\},$$

is manageable with respect to some envelope. For such an envelope, we can consider $U'_1(\omega) = (M_0 + M)/(\overline{\sigma}\epsilon_0)$. We now prove the manageability of $\mathcal{T}'_{1,n,\omega}$. Let us define

$$\mathcal{M}' = \big\{f_{c,y,\phi_1,\phi_2}\left(\widetilde{y}, x, d\right) = d\phi_1\left(y - q\left(\widetilde{y}, c(x)\right)\right)^+ - (1-d)\phi_2\left(y - \widetilde{y}\right)^+,\\ (c,y,\phi_1,\phi_2) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2\big\}.$$

Reasoning as for the class $\mathcal{M}$ defined in (5.9), and using the last equation of the proof of Theorem 3 in Chen et al. (2003), p.1607, we have that for $\epsilon > 0$,

$$N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right) \leq N\left(\epsilon', [N_1, M_1]^2, |\cdot|\right) \times N\left(\epsilon', \mathcal{Y}, |\cdot|\right) \times N\left(\epsilon', \mathcal{C}_s\left([0,1]^{d_X}\right), \|\cdot\|_{[0,1]^{d_X}}\right),$$

with $\epsilon' = (\epsilon/(2Q))^2$ and Q defined in (5.10). Using Theorem 2.7.1 page 155 in Van der Vaart and Wellner (1996), there exists a constant Q_2 depending only on s, d_X, and $[0,1]^{d_X}$ such that

$$\ln\left(N\left(\epsilon', \mathcal{C}_s([0,1]^{d_X}), \|\cdot\|_{[0,1]^{d_X}}\right)\right) \leq Q_2 \epsilon'^{-d_X/s}.$$

Moreover, because $\mathcal{Y}$ and $[N_1, M_1]$ are compact subsets of two Euclidean spaces, there exist Q_3, Q_4 such that

$$N\left(\epsilon', [N_1, M_1]^2, |\cdot|\right) \leq Q_3 \epsilon'^{-4} \text{ and } N\left(\epsilon', \mathcal{Y}, |\cdot|\right) \leq Q_4 \epsilon'^{-2}. \tag{5.11}$$

This yields

$$\ln\left(N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right)\right) \leq (6 + Q_2) \max\left(-\ln(\epsilon'), \epsilon'^{-d_X/s}\right) + \ln(Q_3 Q_4). \tag{5.12}$$

Let $\odot$ denote element-by-element product and $\mathcal{D}\left(\epsilon\left|\alpha \odot U_1'(\omega)\right|, \alpha \odot \mathcal{T}_{1,n,\omega}'\right)$ denote random packing numbers. By (A.1) in Andrews (1994, p.2284), we have

$$\begin{aligned}\sup_{\omega\in\Omega, n\geq 1,\ \alpha\in\mathbb{R}_+^n} \mathcal{D}\left(\epsilon\left|\alpha \odot U_1'(\omega)\right|, \alpha \odot \mathcal{T}_{1,n,\omega}'\right) &\leq \sup_{F\in\mathcal{F}_0} N\left(\frac{\epsilon}{2}, \mathcal{M}', \|\cdot\|_2\right) \\ &\leq \sup_{F\in\mathcal{F}_0} N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right),\end{aligned} \tag{5.13}$$

where the second inequality follows as in e.g., Van der Vaart and Wellner (1996, p.84). Then, (5.12) ensures (see Definition 7.9 in Pollard (1990), p.38) that

$$\sup_{\omega\in\Omega, n\geq 1,\ \alpha\in\mathbb{R}_+^n} \mathcal{D}\left(\epsilon\left|\alpha \odot U_1'(\omega)\right|, \alpha \odot \mathcal{T}_{1,n,\omega}'\right) \leq \lambda(\epsilon),$$

where $\lambda(\epsilon) = \exp\left((6 + Q_2)\max\left(-2\ln\left(\epsilon/(2Q)\right), \left(\epsilon/(2Q)\right)^{-2d_X/s}\right) + \ln(Q_3 Q_4)\right)$. Moreover, by using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b \geq 0$,

$$\int_0^1 \sqrt{\ln(\lambda(\epsilon))} d\epsilon \leq \sqrt{6+Q_2} \int_0^1 \left[\max\left(-2\ln\left(\epsilon/(2Q)\right), \left(\epsilon/(2Q)\right)^{-2d_X/s}\right)\right]^{1/2} d\epsilon$$
$$+ \sqrt{\ln(Q_3 Q_4)}$$
$$< \infty.$$

Thus, $\mathcal{T}'_{1,n,\omega}$ hence $\mathcal{T}_{1,n,\omega}$ are manageable. Therefore, m satisfies PS2 in AS.

Finally, in order to show that SIG1 in AS is satisfied, we use Assumption 4 (iii) and follow the proof of Lemma 7.2 (b) in AS where we replace Y by $q(Y, c(X))$ and B by $\max(M, M_0)$. The result follows.

Proof of Proposition 6

Hereafter, we let $[\underline{\psi}, \overline{\psi}]$ (resp. $[\underline{y}, \overline{y}]$) denote the support of ψ (resp. of Y). As in Lemma 1, H_{0SK} holds if and only if there exists a pair of random variables (Y', ψ') and c such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = Q_c(\psi')$. Now, if Q_c is strictly increasing on $[\underline{\psi}, \overline{\psi}]$, we have $\mathbb{E}[Y'|\psi'] = Q_c(\psi')$ if and only if $\mathbb{E}[Y'|Q_c(\psi')] = Q_c(\psi')$. In view of Theorem 1, the latter is equivalent to F_Y being a mean-preserving spread of $F_{Q_c(\psi')}$. Therefore, the proposition holds if for any $\eta > 0$, there exists K, $c \in \mathbb{R}^{K+1}$ and F such that (i) Q_c is strictly increasing on $[\underline{\psi}, \overline{\psi}]$; (ii) $\sup_{y\in\mathbb{R}} |F_\psi(y) - F(y)| < \eta$; (iii) F_Y is mean-preserving spread of $F_{Q_c(\widetilde{\psi})}$, with $\widetilde{\psi} \sim F$.

Fix $\eta > 0$. Since F_Y is continuous on $[\underline{y}, \overline{y}]$, it is uniformly continuous on this set. Hence, there exists η' such that

$$|y - y'| < \eta' \Rightarrow |F_Y(y) - F_Y(y')| < \eta. \tag{5.14}$$

By assumption, $F_Y^{-1} \circ F_\psi$ is increasing and continuous. Then, by Theorem 9 in Mulansky and Neamtu (1998), there exists a sequence $(P_n)_{n\in\mathbb{N}}$ of increasing polynomials on $[\underline{\psi}, \overline{\psi}]$ satisfying $P_n(\underline{\psi}) = \underline{y}$ and $P_n(\overline{\psi}) = \overline{y}$ and converging uniformly to $F_Y^{-1} \circ F_\psi$.

Hence, there exists P_{n_0} such that

$$\sup_{y\in[\underline{\psi},\overline{\psi}]} |P_{n_0}(y) - F_Y^{-1} \circ F_\psi(y)| < \eta'. \tag{5.15}$$

Let K be the degree of P_{n_0} and $c \in \mathbb{R}^K$ denote the vector of coefficients of P_{n_0}, so that $Q_c = P_{n_0}$. Q_c is a non-constant polynomial, which is increasing on $[\underline{\psi}, \overline{\psi}]$. Hence, its derivative vanishes a finite number of times and Q_c is actually strictly increasing. Hence, Condition (i) above holds. Moreover, combining (5.15) with (5.14), we obtain

$$\sup_{y\in[\underline{\psi},\overline{\psi}]} |F_Y \circ Q_c(y) - F_\psi(y)| < \eta.$$

Now, let $F := F_Y \circ Q_c$ on $[\underline{\psi}, \overline{\psi}]$, $F(y) := 0$ for all $y < \underline{\psi}$ and $F(y) := 1$ for all $y > \overline{\psi}$. Then F is continuous and increasing, with limit 0 and 1 respectively at $-\infty$ and ∞. Thus, it is a cdf and Condition (ii) above holds. Finally, let $\widetilde{\psi} \sim F$. We have, for any $y \in [\underline{y}, \overline{y}]$,

$$P\left(Q_c(\widetilde{\psi}) \leq y\right) = F \circ Q_c^{-1}(y) = F_Y(y).$$

This implies that $F_{Q_c(\widetilde{\psi})}$ is a mean-preserving spread of F_Y. The result follows.

Proof of Proposition 7

1. We consider for that purpose $(\psi^*, \xi_\psi^*, \xi_Y^*, \varepsilon^*) \sim \mathcal{N}(m, \Sigma)$, potentially different from the true $(\psi, \xi_\psi, \xi_Y, \varepsilon)$, and let

$$\widehat{\psi}^* = \psi^* + \xi_\psi^*,$$
$$\widehat{Y}^* = \psi^* + \varepsilon^* + \xi_Y^*.$$

We then fix (m, Σ) so that the DGP satisfies all the restrictions specified in the propositions, and in particular, $(\mathbb{V}(\widehat{Y}^*), \mathbb{V}(\widehat{\psi}^*), \text{Cov}(\widehat{Y}^*, \widehat{\psi}^*)) = (\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}), \text{Cov}(\widehat{Y}, \widehat{\psi}))$. First, letting $m = (m_1, m_2, m_3, m_4)'$, we impose $m_2 = m_3 = m_4 = 0$, and set all the non-diagonal terms of Σ, except $\Sigma_{23} = \text{Cov}(\xi_\psi^*, \xi_Y^*)$, equal to zero. Then $(\widehat{Y}^*, \widehat{\psi}^*, \psi^*)$ satisfy (5.1) and RE hold (considering $\mathcal{I} = \sigma(\psi^*)$ and $Y^* = \psi^* + \varepsilon^*$). We fix below $\Sigma_{22} \in [0, \mathbb{V}(\widehat{\psi})]$. Then let $\Sigma_{11} = \mathbb{V}(\widehat{\psi}) - \Sigma_{22}$ and $\Sigma_{33} = \mathbb{V}(\widehat{Y}) - \mathbb{V}(\widehat{\psi}) + \Sigma_{22}$ and $\Sigma_{44} = 0$, so that $(\mathbb{V}(\widehat{Y}^*), \mathbb{V}(\widehat{\psi}^*)) = (\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}))$. Also, because $\mathbb{V}(\widehat{Y}) > \mathbb{V}(\widehat{\psi})$, $\mathbb{V}(\xi_\psi^*) < \mathbb{V}(\xi_Y^* + \varepsilon^*)$ and $F_{\xi_\psi^*}$ dominates at the second order $F_{\xi_Y^*+\varepsilon^*}$.

Now, we fix Σ_{22}. Let $a = \mathbb{V}(\widehat{Y}) - \mathbb{V}(\widehat{\psi})$ and $c = \text{Cov}(\widehat{Y} - \widehat{\psi}, \widehat{\psi})$. Then, by Cauchy-Schwarz inequality,

$$c^2 \leq \mathbb{V}(\widehat{\psi})\mathbb{V}(\widehat{Y} - \widehat{\psi}) = \mathbb{V}(\widehat{\psi})(a - 2c).$$

This means that there exists $\sigma^2 \in [0, \mathbb{V}(\widehat{\psi})]$ such that

$$c^2 \leq \sigma^2(a - 2c). \tag{5.16}$$

Let $\Sigma_{22} = \sigma^2$ and $\Sigma_{23} = c + \Sigma_{22}$. Then, by construction,

$$\begin{aligned} \text{Cov}(\widehat{Y}^*, \widehat{\psi}^*) &= \Sigma_{11} + \Sigma_{23} \\ &= \mathbb{V}(\widehat{\psi}) - \Sigma_{22} + \Sigma_{22} + c \\ &= \text{Cov}(\widehat{Y}, \widehat{\psi}). \end{aligned}$$

Moreover, in view of (5.16) and by definition of Σ_{22} and Σ_{33},

$$\begin{aligned} \Sigma_{23}^2 &= c^2 + 2c\Sigma_{22} + \Sigma_{22}^2 \\ &\leq (a - 2c)\Sigma_{22} + 2c\Sigma_{22} + \Sigma_{22}^2 \\ &= \Sigma_{33}\Sigma_{22}. \end{aligned}$$

In other words, Σ is a proper covariance matrix.

2. Let $\lambda = \mathbb{V}(\psi)/\sigma_{\xi_\psi}^2$. If (5.1) and RE hold, $\text{Cov}(\xi_\psi, \varepsilon + \xi_Y) \geq 0$ and $\lambda \geq \underline{\lambda}$, we

obtain

$$\beta - 1 = \frac{\text{Cov}(\widehat{Y} - \widehat{\psi}, \widehat{\psi})}{\mathbb{V}(\widehat{\psi})}$$
$$= \frac{\text{Cov}(\varepsilon + \xi_Y - \xi_\psi, \xi_\psi)}{\sigma^2_{\xi_\psi}(1+\lambda)}$$
$$\geq -\frac{1}{1+\underline{\lambda}}.$$

The result follows.

Proof of Proposition 8

We first prove that if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, there exists a unique $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. First, suppose that $F^b \neq F^{b'}$ and, without loss of generality, $b > b'$. Then $\psi^b \leq \psi^{b'}$, implying that $F^b(y) \leq F^{b'}(y)$ for all y. Moreover, the inequality is strict for at least one y. As a result, $\mathbb{E}(\psi^b) > \mathbb{E}(\psi^{b'})$. In other words, there is at most one $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. If $\mathbb{E}[\psi_L] = \mathbb{E}[Y]$ or $\mathbb{E}[\psi_U] = \mathbb{E}[Y]$, such a solution also exists by taking $b = -\infty$ and $b = \infty$, respectively. Now, suppose that $\mathbb{E}[\psi_L] < \mathbb{E}[Y] < \mathbb{E}[\psi_U]$. For all $\infty > b > b' > -\infty$,

$$\psi^b - \psi^{b'} = (\psi_U - \max(\psi_L, b'))\,\mathbb{1}\{\psi_U \in [b', b)\} + (b - b')\mathbb{1}\{\psi_L < b', \psi_U \geq b\}$$
$$+ (b - \psi_L)\mathbb{1}\{\psi_L \in [b', b), \psi_U \geq b\}.$$

As a result, $|\psi^b - \psi^{b'}| \leq |b - b'|$. This implies that $\widetilde{\delta} : b \mapsto \mathbb{E}[\psi^b]$ is continuous. Moreover, $\lim_{b \to -\infty} \widetilde{\delta}(b) = \mathbb{E}[\psi_L] < \mathbb{E}(Y)$ and $\lim_{b \to \infty} \widetilde{\delta}(b) = \mathbb{E}[\psi_U] > \mathbb{E}(Y)$. By the intermediate value theorem, there exists b^* such that $\widetilde{\delta}(b^*) = \mathbb{E}(Y)$. Hence, there exists $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. The first part of Proposition 8 follows.

Let us turn to the second part of the proposition. First, if (ii) holds, there exists $b_0 \in \overline{\mathbb{R}}$ such that $F^* = F^{b_0}$. Then, by construction and Theorem 1, Y and ψ^{b_0} satisfy H_0. Moreover, $F^{b_0} \in [F_{\psi_U}, F_{\psi_L}]$. Therefore, H_{0B} holds as well.

Now, let us prove that (i) implies (ii). Let us denote by $\mathcal{D}$ the set of all the cdfs for ψ such that H_{0B} holds. By Theorem 1, these are cdfs F satisfying $F_{\psi_U} \leq F \leq F_{\psi_L}$,

$\delta_F = 0$ and dominating at the second order F_Y. We show below that all $F \in \mathcal{D}$ are dominated at the second order by F^*. Then, because $F_{\psi_U} \le F^* \le F_{\psi_L}$ and $\int y dF^*(y) = \int y dF_Y(y)$, $\mathcal{D}$ is not empty only if F^* dominates at the second order F_Y. The result then follows by Theorem 1.

Thus, we have to show that for all $t \in \mathbb{R}$,

$$F^* = \text{argmin}_{F_\psi \in \mathcal{D}} \int_{-\infty}^{t} F_\psi(y)dy. \tag{5.17}$$

First, if $F^* = F^{-\infty}$, we have for all $F \neq F^*$, $F(y) \le F_{\psi_L}(y) = F^*(y)$ for all y, with strict inequality for some y. Then $\delta_F > \delta_{F^*} = 0$ and $\mathcal{D} = \{F^*\}$, implying that (5.17) holds. Similarly, (5.17) holds if $F^* = F^{\infty}$.

Suppose now that $F^* = F^{b_0}$ for some $b_0 \in \mathbb{R}$. Because $F_{\psi_U}(y) \le F_\psi(y)$ for all $y < b_0$ and all $F_\psi \in \mathcal{D}$, (5.17) holds for all $t < b_0$. We now prove that (5.17) holds also for $t \ge b_0$. First suppose that $t \ge \max(b_0, 0)$. For all $F_\psi \in \mathcal{D}$, $\int y dF_Y(y) = \int y dF_\psi(y)dy$. As a result, by Fubini's theorem,

$$\begin{aligned}
&-\int_{-\infty}^{0} F^*(y)dy + \int_{0}^{t} (1 - F^*(y))\, dy + \int_{t}^{\infty} (1 - F^*(y))\, dy \\
&= -\int_{-\infty}^{0} F_\psi(y)dy + \int_{0}^{t} (1 - F_\psi(y))\, dy + \int_{t}^{\infty} (1 - F_\psi(y))\, dy.
\end{aligned}$$

Because $F_\psi \le F_{\psi_L} = F^*$ on $[b_0, \infty]$, this implies that

$$-\int_{-\infty}^{0} F^*(y)dy + \int_{0}^{t} (1 - F^*(y))\, dy \ge -\int_{-\infty}^{0} F_\psi(y)dy + \int_{0}^{t} (1 - F_\psi(y))\, dy$$

and thus (5.17) holds for $t \ge \max(b_0, 0)$. Now, if $b_0 < 0$ and $t \in (b_0, 0)$, we have

$$\begin{aligned}
&-\left(\int_{-\infty}^{t} F^*(y)dy + \int_{t}^{0} F^*(y)dy\right) + \int_{0}^{\infty} (1 - F^*(y))\, dy \\
&= -\left(\int_{-\infty}^{t} F_\psi(y)dy + \int_{t}^{0} F_\psi(y)dy\right) + \int_{0}^{\infty} (1 - F_\psi(y))\, dy.
\end{aligned}$$

Using again $F_\psi \leq F_{\psi_L} = F^*$ on $[t, \infty)$ yields

$$-\int_t^0 F^*(y)dy + \int_0^\infty (1 - F^*(y))\, dy \leq -\int_t^0 F_\psi(y)dy + \int_0^\infty (1 - F_\psi(y))\, dy.$$

Therefore, the result also follows in this case.

Chapter 6

Estimates for the SVD of the Truncated Fourier Transform on $L^2(\cosh(b|\cdot|))$ and Stable Analytic Continuation

Joint with Eric Gautier, Toulouse School of Economics.

Preprint available here.

Abstract

The Fourier transform truncated on $[-c, c]$ is usually analyzed when acting on $L^2(-1/b, 1/b)$ and its right-singular vectors are the prolate spheroidal wave functions. This paper considers the operator acting on the larger space $L^2(\cosh(b|\cdot|))$ on which it remains injective. We give nonasymptotic upper and lower bounds on the singular values with similar qualitative behavior in m (the index), b, and c. The lower bounds are used to obtain rates of convergence for stable analytic continuation of possibly non-bandlimited functions whose Fourier transform belongs to $L^2(e^{b|\cdot|})$. We also derive bounds on the sup-norm of the singular functions. Finally, we propose a numerical

method to compute the SVD and apply it to stable analytic continuation when the function is observed with error on an interval. In the application we consider cases where the function to extrapolate is not bandlimited and when it is bandlimited but the bandlimits are unknown.

Keywords: Analytic continuation, Nonbandlimited functions, Heavy tails, Uniform estimates, Extrapolation, Singular value decomposition, Truncated Fourier transform, Singular Sturm Liouville Equations, Superresolution.

6.1 Introduction

Extrapolating an analytic square integrable function f from its observation with error on $[-c, c]$ to $\mathbb{R}$ has a wide range of applications, for example in imaging and signal processing Gosse (2010*b*), in geostatistics and with big data Coifman and Lafon (2006), and finance Gosse (2010*a*). A researcher may want to estimate a density from censored data. This means that the data is only available on a smaller set than the one of interest (see, *e.g.*, Belitser (1998); Berman (2007)). When the function is a Fourier transform, this is a type of super-resolution in image restoration Bertero and Boccacci (1998); Bertero et al. (1984); Gerchberg (1974) which can be achieved under auxiliary information such as information on the support of the object. The related problem of out-of-band extrapolation (see, *e.g.*, Alibaud et al. (2009); Bertero and Boccacci (1998)) consists in recovering a function from partial observation of its Fourier transform. A particular instance of this framework appears in the analysis of the random coefficients linear model (see, *e.g.* Gaillac and Gautier (2019*c*)). There, the model takes the form

$$Y = \alpha + \beta^\top X,$$

where $(\alpha, \beta^\top) \in \mathbb{R}^{p+1}$ and $X \in \mathbb{R}^p$ are independent random vectors, and the researcher has at her disposal an independent and identically distributed sample of $(Y, X^\top)$ from which she can estimate the Fourier transform of the density of the coefficients $(\alpha, \beta^\top)$ on $\{(t, tx) : \ (t, x) \in \mathbb{R} \times \mathcal{X}\}$, where $\mathcal{X} \subseteq \mathbb{R}^p$ is the support of X and the object of interest is the density the coefficients.

It is customary to rely on analytic functions and use Hilbert space techniques. For the extrapolation problem, one can restrict attention to bandlimited functions which are square integrable functions whose Fourier transforms have support in $[-1/b, 1/b]$. For out-of-band extrapolation, one can work with square-integrable functions whose support is a subset of $[-1/b, 1/b]$ in which case their Fourier transform is analytic by the Paley-Wiener theorem (see Reed and Simon (1980)). Prolate spheroidal wave functions (henceforth PSWF, see Osipov et al. (2013); Slepian (1965)) are the right-singular functions of the truncated Fourier transform restricted to functions with support in $[-1/b, 1/b]$. The truncated Fourier transform maps functions to their Fourier transform on $[-c, c]$. The PSWF form an orthonormal basis of the space $L^2(-c, c)$ of square-integrable functions on $(-c, c)$, are restrictions of square integrable orthogonal analytic functions on $\mathbb{R}$, and form a complete system of the bandlimited functions with bandlimits $[-1/b, 1/b]$. Hence, a bandlimited function on the whole line is simply the series expansion on the PSWF basis, sometimes called Slepian series, whose coefficients only depend on the function on $(-c, c)$, almost everywhere on $\mathbb{R}$. This makes sense if we understand the PSWF functions as their extension to $\mathbb{R}$. In this framework, analytic continuation is an inverse problem in the sense that the solution does not depend continuously on the data, more specifically severely ill-posed (see, *e.g.*, Fu et al. (2008); Grabovsky and Hovsepyan (2020); Shapiro (1986); Zhang et al. (2011)), and many methods have been proposed (see, *e.g.*, Batenkov et al. (2018); Bertero et al. (1979); Chen (2010); Coifman and Lafon (2006); Drouiche et al. (2001); Landau (1986); Miller (1970); Trefethen (2019)). To obtain precise error bounds, it is useful to obtain nonasymptotic upper and lower bounds on the singular values of the truncated Fourier transform rather than the more usual asymptotic estimates on a logarithmic scale. In several applications, uniform estimates on the right singular functions are useful as well. This occurs for example to show that certain nonparametric statistical procedures involving series are adaptive (see, *e.g.*, Chagny (2015)). This means that an estimator with a data-driven smoothing parameter reaches the optimal minimax rate of convergence. Importantly, such a program providing nonasymptotic bounds on the singular values and right singular functions has been carried recently in relation to bandlimited functions in Bonami et al. (2018); Bonami and Karoui (2016, 2017); Osipov (2013); Osipov et al. (2013); Rokhlin and Xiao (2007). A second important aspect is the access to efficient methods to obtain the singular value decomposition (henceforth SVD). While numerical solutions to the inverse problems have for a long time relied on the Tikhonov or iterative methods such as the Landweber method (Gerchberg method for out-of band extrapolation, see Bertero and Boccacci (1998)) to avoid using the SVD, recent developments have made it possible to approximate efficiently the PSWF and the SVD (see Section 8 of Osipov et al. (2013)).

Assuming that the function observed on an interval is the restriction of a bandlimited function can be questionable. For example, in the case of censored data, the observed function is a truncated density and the underlying function a density, and none of the usual families are bandlimited. Moreover, even if the function were bandlimited, one would require an upper bound on $1/b$ which might not be available in practice (see Slepian (1983)). For this reason, this paper considers the larger class of functions whose Fourier transforms belong to the space $L^2(e^{b|\cdot|})$ of square-integrable functions with weight function $e^{b|\cdot|}$. This is the largest space that we can consider to extrapolate a function with Hilbert space techniques because, for $a > 0$, $\left\{f \in L^2(\mathbb{R}) : \ \forall b < a, \ \mathcal{F}[f] \in L^2\left(e^{b|\cdot|}\right)\right\}$ is the set of square-integrable functions which have an analytic continuation on $\{z \in \mathbb{C} : \ |\mathrm{Im}(z)| < a/2\}$ (see Theorem IX.13 in Reed and Simon (1980)). The broader class $L^2(e^{b|\cdot|})$ has rarely been used in this context and, unlike the PSWF, much fewer results are available, with the notable exception of Morrison (1962); Widom (1964). It is considered in Belitser (1998) in the case of censored data and in Gaillac and Gautier (2019*c*) for the problem of estimating the density of random coefficients in the linear random coefficients model. There it is meaningful to assume the Laplace transform of the density is finite near 0 or equivalently that it does not have heavy-tails.

In this paper, we use the weight $\cosh(b\cdot)$ rather than $e^{b|\cdot|}$ because the Fourier transform of $\mathrm{sech} = 1/\cosh$ is essentially itself and, though with a different scalar product, $L^2\left(\cosh(b\cdot)\right) = L^2\left(e^{b|\cdot|}\right)$. Theorem II in Widom (1964) provides, for given $b, c > 0$

and a value of the index m going to infinity, an equivalent of the logarithm of the singular values of the truncated Fourier transform acting on $L^2(\cosh(b\cdot))$. Such an equivalent is important but this result is silent on the polynomial preexponential factors, their dependence with respect to c and b, and to deduce upper and lower bounds on the singular values which hold for all m, c, and b. This behavior is important in Gaillac and Gautier (2019c) where we integrate the bounds over c in intervals $[a, b]$ where a can be arbitrarily close to 0. This paper provides nonasymptotic upper and lower bounds on the singular values, with similar qualitative behavior, and applies the lower bounds to error bounds for stable analytic continuation using the spectral cut-off method. There, the nonasymptotic lower bounds are important to obtain a tight polynomial rates of convergence for "supersmooth functions". We also analyze a differential operator which commutes with a symmetric integral operator obtained by applying the truncated Fourier operator to its adjoint. The corresponding eigenvalue problem involves singular Sturm-Liouville equations. This allows to prove uniform estimates on the right-singular functions. Solving numerically singular differential equations allows to obtain these functions, hence all the SVD. Working with the differential operator is useful because its eigenvalues increase quadratically while those of the integral operator decrease exponentially. Finally, we illustrate numerically the proposed method for stable analytic continuation by spectral cut-off. We propose an adaptive method to select the cut-off. When the function is bandlimited and the researcher knows an interval which contains the bandlimits, we rely on the PSWF and efficient methods to compute the SVD. We also illustrate the proposed method involving numerical schemes for singular differential equations when the researcher does not have prior information on the bandlimits and when she questions the fact that the function can be bandlimited as in the statistical applications presented above.

6.2 Preliminaries

We use $\mathbb{N}_0$ for the set of nonnegative natural numbers, $a \vee b$ for the maximum of a and b, *a.e.* for almost everywhere, and $f(\cdot)$ for a function f of some generic argument. We denote, for $a > 0$, by $L^2(-a,a)$ and $L^2(\mathbb{R})$ the usual L^2 spaces of complex-valued functions equipped with the Hermitian inner product, for example $\langle f,g\rangle_{L^2(-a,a)} = \int_{-a}^{a} f(x)\overline{g}(x)dx$, by $L^2(W)$ for a positive function W on $\mathbb{R}$ the weighted L^2 spaces equipped with $\langle f,g\rangle_{L^2(W)} = \int_{\mathbb{R}} f(x)\overline{g}(x)W(x)dx$, and by $S^{\perp}$ the orthogonal complement of the set S in a Hilbert space. We denote by $\|f\|_{L^\infty([a,b])}$ the sup-norm of the function f on $[a,b]$. $\mathcal{E}$ is the operator which extends a function in $L^2(-1,1)$ to $L^2(\mathbb{R})$ by assigning the value 0 outside $[-1,1]$ and $\mathcal{R}L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is such that $\mathcal{R}f = f(-\cdot)$. The inverse of a mapping f, when it exists, is denoted by f^I. We denote, for $b,c > 0$, by

$$\begin{array}{cccc} \mathcal{C}_c: & L^2(\mathbb{R}) & \to & L^2(\mathbb{R}) \\ & f & \mapsto & cf(c\cdot) \end{array}, \qquad \begin{array}{cccc} \mathcal{F}_{b,c}: & L^2(\cosh(b\cdot)) & \to & L^2(-1,1) \\ & f & \mapsto & \mathcal{F}[f](c\cdot) \end{array}, \tag{6.1}$$

by $\mathcal{F}[f] = \int_{\mathbb{R}} e^{ix\cdot} f(x)dx$ the Fourier transform of f in $L^1(\mathbb{R})$ and also use the notation $\mathcal{F}[f]$ for the Fourier transform in $L^2(\mathbb{R})$. $\mathcal{O}^*$ denotes the Hermitian adjoint of an operator $\mathcal{O}$. Recovering f such that for b small enough $f \in L^2(\cosh(b\cdot))$ based on its Fourier transform on $[-c,c]$ amounts to inverting $\mathcal{F}_{b,c}$. This can be achieved using the SVD. Define the finite convolution operator

$$\begin{array}{cccc} \mathcal{Q}_c: & L^2(-1,1) & \to & L^2(-1,1) \\ & h & \mapsto & \int_{-1}^{1} \pi c \operatorname{sech}\left(\frac{\pi c}{2}(\cdot - y)\right) h(y)dy. \end{array} \tag{6.2}$$

It is compact, symmetric, and positive on spaces of real and complex valued functions. Denote by $(\rho_m^c)_{m\in\mathbb{N}_0}$ its positive real eigenvalues in decreasing order and repeated according to multiplicity and by $(g_m^c)_{m\in\mathbb{N}_0}$ its eigenfunctions which can be taken to be real valued. The next proposition relies on the fact that, for all $c > 0$, $\mathcal{F}[\operatorname{sech}(c\cdot)](\star) = (\pi/c)\operatorname{sech}(\pi\star/(2c))$.

Proposition 1. *For $b,c > 0$, we have $c\mathcal{F}_{b,c}\mathcal{F}_{b,c}^* = \mathcal{Q}_{c/b}$.*

Proof. Because $\mathcal{F}_{b,c} = \mathcal{F}\mathcal{C}_{c^{-1}} = c^{-1}\mathcal{C}_c\mathcal{F}$, $\mathcal{R}\mathcal{F}_{b,c} = \mathcal{F}_{b,c}\mathcal{R}$,

$$\mathcal{F}_{b,c}^* = \operatorname{sech}(b\cdot)\mathcal{R}\mathcal{F}_{b,c}\mathcal{E}, \tag{6.3}$$

and $\operatorname{sech}(b\cdot)$ is even, we obtain $\mathcal{F}_{b,c}^* = \mathcal{R}\left[\operatorname{sech}(b\cdot)\mathcal{F}_{b,c}\mathcal{E}\right]$ and

$$\begin{aligned} c\mathcal{F}_{b,c}\mathcal{F}_{b,c}^* &= c\mathcal{R}\mathcal{F}_{b,c}\left[\operatorname{sech}(b\cdot)\mathcal{F}_{b,c}\mathcal{E}\right] \\ &= 2\pi\mathcal{F}^I\left[\mathcal{C}_{c^{-1}}\left[\operatorname{sech}(b\cdot)\mathcal{C}_c\mathcal{F}\mathcal{E}\right]\right] \\ &= 2\pi c\mathcal{F}^I\left[\mathcal{C}_{c^{-1}}\left[\operatorname{sech}(b\cdot)\right]\mathcal{F}\mathcal{E}\right], \end{aligned}$$

where, for $a.e.$ $x \in \mathbb{R}$,

$$2\pi c\mathcal{F}^I\left[\mathcal{C}_{c^{-1}}\left[\operatorname{sech}(b\cdot)\right]\right](x) = \int_{\mathbb{R}} e^{-itx}\operatorname{sech}\left(\frac{bt}{c}\right)dt = \frac{\pi c}{b}\operatorname{sech}\left(\frac{\pi c}{2b}x\right).$$

As a result, we have, for $f \in L^2(-1,1)$,

$$c\mathcal{F}_{b,c}\mathcal{F}_{b,c}^*[f] = \mathcal{C}_{\pi c/(2b)}\left[2\operatorname{sech}\right] * \mathcal{E}[f] = \mathcal{Q}_{c/b}[f].$$

□

Proposition 1 yields that $(g_m^{c/b})_{m\in\mathbb{N}_0}$ are the right singular functions of $\mathcal{F}_{b,c}$. The SVD of $\mathcal{F}_{b,c}$, denoted by $\left(\sigma_m^{b,c}, \varphi_m^{b,c}, g_m^{c/b}\right)_{m\in\mathbb{N}_0}$, is such that, for $m \in \mathbb{N}_0$,

$$\sigma_m^{b,c} = \sqrt{\frac{\rho_m^{c/b}}{c}} \tag{6.4}$$

and $\varphi_m^{b,c} = \mathcal{F}_{b,c}^* g_m^{c/b}/\sigma_m^{b,c}$. It yields, for all $f \in L^2\left(\cosh(b\cdot)\right)$,

$$f = \sum_{m\in\mathbb{N}_0} \frac{1}{\sigma_m^{b,c}}\left\langle \mathcal{F}_{b,c}\left[f\right], g_m^{c/b}\right\rangle_{L^2(-1,1)} \varphi_m^{b,c}. \tag{6.5}$$

(6.5) is a core element to approximate a function from partial observations of its Fourier transform when the signal f does not have compact support.

Proposition 2. *For all $b, c > 0$, $\mathcal{F}_{b,c}$ is injective and $\left(\varphi_m^{b,c}\right)_{m\in\mathbb{N}_0}$ is a basis of $L^2(\cosh(b\cdot))$.*

Proof. We use that, for every $h \in L^2(\cosh(b\cdot))$, if we do not restrict the argument in the definition of $\mathcal{F}_{b,c}[h]$ to $[-1,1]$, $\mathcal{F}_{b,c}[h]$ can be defined as a function in $L^2(\mathbb{R})$. In what follows, for simplicity, we use $\mathcal{F}_{b,c}[h]$ for both the function in $L^2(-1,1)$ and in $L^2(\mathbb{R})$.

Let us show that $\mathcal{F}_{b,c}$ defined in (6.1) is injective. Take $h \in L^2(\cosh(b\cdot))$ such that $\mathcal{F}_{b,c}[h]$ is zero on $[-1,1]$. Then, using Theorem IX.13 in Reed and Simon (1980), $\mathcal{F}_{b,c}[h]$ is zero on $\mathbb{R}$. Thus, $\mathcal{F}[h]$ hence h are zero $a.e.$ on $\mathbb{R}$.

The second part of Proposition 2 holds by Theorem 15.16 in Kress (1999) and the injectivity of $\mathcal{F}_{b,c}$. □

Theorem II in Widom (1964) provides the equivalent

$$\log\left(\rho_m^c\right) \underset{m\to\infty}{\sim} -\pi m \frac{K(\operatorname{sech}(\pi c))}{K(\tanh(\pi c))}, \tag{6.6}$$

where $K(r) = \int_0^{\pi/2}\left(1 - r^2\sin(x)^2\right)^{-1/2} dx$ is the complete elliptic integral of the first kind. This paper provides nonasymptotic upper and lower bounds on the eigenvalues and upper bounds on the sup-norm of the functions $(g_m^c)_{m\in\mathbb{N}_0}$.

The proofs of this paper sometimes rely on the following operator

$$\begin{aligned} \mathcal{F}_c^{W_{[-1,1]}}: \quad L^2\left(W_{[-1,1]}\right) &\to L^2(-1,1), \\ f &\to \mathcal{F}[f](c\cdot) \end{aligned} \tag{6.7}$$

where $W_{[-1,1]} = \mathbb{1}\{[-1,1]\} + \infty\,\mathbb{1}\{[-1,1]^c\}$, for which we use the notations $\rho_m^{W_{[-1,1]},t_m}$ for the m^{th} eigenvalue of $\mathcal{Q}_c^{W_{[-1,1]}} = c\mathcal{F}_c^{W_{[-1,1]}}\left(\mathcal{F}_c^{W_{[-1,1]}}\right)^*$.

6.3 Lower bounds on the eigenvalues of $\mathcal{Q}_c$ and an application

6.3.1 Lower bounds on the eigenvalues of $\mathcal{Q}_c$

Lemma 1. *For all $m \in \mathbb{N}_0$, $c \in (0,\infty) \mapsto \rho_m^c$ is nondecreasing.*

Proof. Take $m \in \mathbb{N}_0$. Using the maximin principle (see Theorem 5 page 212 in Birman and Solomjak (2012)), the $m+1$-st eigenvalue ρ_m^c satisfies

$$\rho_m^c = \max_{V\in S_{m+1}} \min_{f\in V\setminus\{0\}} \frac{\langle \mathcal{Q}_c f, f\rangle_{L^2(-1,1)}}{\|f\|_{L^2(-1,1)}^2},$$

where S_{m+1} is the set of $m+1$-dimensional vector subspaces of $L^2(-1,1)$. Using (6.3) and Proposition 1, we obtain

$$\begin{aligned}
\langle \mathcal{Q}_c f, f\rangle_{L^2(-1,1)} &= c\left\langle \mathcal{F}_{1,c}\mathcal{F}_{1,c}^*[f], f\right\rangle_{L^2(-1,1)} \\
&= c\left\langle \mathcal{F}_{1,c}^*[f], \mathcal{F}_{1,c}^*[f]\right\rangle_{L^2(\cosh)} \\
&= c\left\|\operatorname{sech} \times \mathcal{F}_{1,c}\left[\mathcal{E}\left[f\right]\right]\right\|_{L^2(\cosh)}^2 \\
&= c\int_{\mathbb{R}} \operatorname{sech}(x)\left|\int_{\mathbb{R}} e^{ictx}\mathcal{E}\left[f\right](t)dt\right|^2 dx \\
&= \int_{\mathbb{R}} \operatorname{sech}\left(\frac{x}{c}\right)\left|\mathcal{F}\left[\mathcal{E}\left[f\right]\right](x)\right|^2 dx
\end{aligned} \tag{6.8}$$

hence

$$\rho_m^c = \max_{V\in S_{m+1}} \min_{f\in V\setminus\{0\}} \frac{2\pi\int_{\mathbb{R}} \operatorname{sech}(x/c)\left|\mathcal{F}\left[\mathcal{E}\left[f\right]\right](x)\right|^2 dx}{\left\|\mathcal{F}\left[\mathcal{E}\left[f\right]\right]\right\|_{L^2(\mathbb{R})}^2}. \tag{6.9}$$

Then, using that $t \mapsto \cosh(t)$ is even, nondecreasing, and positive, we obtain that, for all $0 < c_1 \le c_2$ and $x \in \mathbb{R}$, $\operatorname{sech}(x/c_2) \ge \operatorname{sech}(x/c_1)$ hence that $\rho_m^{c_1} \le \rho_m^{c_2}$. □

Theorem 1. *For all $m \in \mathbb{N}_0$, we have*

$$\forall\ 0 < c \le \frac{\pi}{4},\ \rho_m^c \ge \frac{2\sin(2c)^2}{e^2 c}\exp\left(-2\log\left(\frac{7e^2\pi}{2c}\right)m\right) \tag{6.10}$$

$$\forall c > 0,\ \rho_m^c \ \ge \pi\exp\left(-\frac{\pi(m+1)}{2c}\right). \tag{6.11}$$

(6.10) is valid for $0 < c \le \pi/4$ and more precise than (6.11) for c close to 0. (6.11) is uniformly valid. To prove it, we show that $\rho_m^c \ge \operatorname{sech}(t_m/c)\,\rho_m^{W_{[-1,1]},t_m}$ for well chosen t_m and rely on a lower bound on $\rho_m^{W_{[-1,1]},t_m}$. The proof of (6.10) uses similar arguments as those in Bonami et al. (2018) Section 5.1 and a lower bound on the best constant $\Gamma(m,\epsilon)$ such that for all interval $I \subseteq [-\pi,\pi]$ of length $2\epsilon > 0$ and all polynomial of degree at most $m \in \mathbb{N}_0$,

$$\left\|P\left(e^{i\cdot}\right)\right\|_{L^2(I)}^2 \geq \Gamma(m,\epsilon)\left\|P\left(e^{i\cdot}\right)\right\|_{L^2(-\pi,\pi)}^2.$$

We use the lower bound in Nazarov (2000) page 240

$$\Gamma(m,\epsilon) \geq \left(\frac{14e\pi}{\epsilon}\right)^{-2m}\frac{\epsilon}{\pi} \tag{6.12}$$

for $\epsilon = 4c$. It is argued in Nazarov (2000) that it cannot be significantly improved for small ϵ which is precisely the regime for which (6.10) is used to bound the eigenvalues from below.

Proof. Let $m \in \mathbb{N}_0$, $c > 0$, and $M = (m+1)/(2c)$. For $R > 0$, we denote by $PW(R)$ the Paley-Wiener space of functions whose Fourier transform has support in $[-R,R]$ and by $S_{m+1}(R)$ the set of $m+1$-dimensional subspaces of $PW(R)$. Using (6.9), we have

$$\rho_m^c = \max_{V\in S_{m+1}(1)}\min_{g\in V\setminus\{0\}}\frac{2\pi\int_{\mathbb{R}}\operatorname{sech}(x/c)\left|g(x)\right|^2dx}{\|g\|_{L^2(\mathbb{R})}^2}.$$

Then, for $g \in PW(1)$, the function $g_{Mc}: \ x \in \mathbb{R} \mapsto (Mc)^{1/2}g(Mcx)$ satisfies $\|g\|_{L^2(\mathbb{R})}^2 = \|g_{Mc}\|_{L^2(\mathbb{R})}^2$ and belongs to $PW(Mc)$. Using

$$\int_{\mathbb{R}}\operatorname{sech}\left(\frac{x}{c}\right)\left|g(x)\right|^2dx = \int_{\mathbb{R}}\operatorname{sech}(Mx)\left|g_{Mc}(x)\right|^2dx,$$

we have, for $V \in S_{m+1}(Mc)$,

$$\rho_m^c \geq \min_{g\in V\setminus\{0\}}\frac{2\pi\int_{\mathbb{R}}\operatorname{sech}(Mx)\left|g(x)\right|^2dx}{\|g\|_{L^2(\mathbb{R})}^2}. \tag{6.13}$$

Let us now choose a convenient such space V defined, for $\varphi: t \in \mathbb{R} \mapsto \sin(t/2)/(\pi t)$, as

$$V = \left\{\sum_{k=0}^m P_k e^{i(k-m/2)\cdot}\varphi(\cdot), \quad (P_k)_{k=0}^m \in \mathbb{C}^{m+1}\right\}.$$

The Fourier transform of an element of V is of the form $\sum_{k=0}^m P_k\mathcal{F}[\varphi](\cdot - k + m/2)$

and, because $\mathcal{F}[\varphi](\cdot) = \mathbb{1}\{|\cdot| \leq 1/2\}$, it has support in $[-1/2 - m/2, 1/2 + m/2] = [-Mc, Mc]$. This guarantees that $V \in S_{m+1}(Mc)$.

We now obtain a lower bound on the right-hand side of (6.13). Let $g \in V$, defined via the coefficients $(P_k)_{k=0}^m$, and, for $x \in \mathbb{R}$, let $P(x) = \sum_{k=0}^m P_k x^k$. Let $0 < x_0 \leq \pi/2$. We have, using $\forall x \in [0, 2x_0)$, $\sin(x/2)/x \geq \sin(x_0)/(2x_0)$ for the last display,

$$\begin{aligned}\int_{\mathbb{R}} \operatorname{sech}(Mx) |g(x)|^2 dx &\geq \int_{-2x_0}^{2x_0} \operatorname{sech}(Mx) \left| \sum_{k=0}^m P_k e^{ikx} \right|^2 |\varphi(x)|^2 dx \\ &\geq \frac{1}{\cosh(2Mx_0)} \min_{x \in [-2x_0, 2x_0]} |\varphi(x)|^2 \int_{-2x_0}^{2x_0} \left| \sum_{k=0}^m P_k e^{ikx} \right|^2 dx \\ &\geq \frac{\sin(x_0)^2}{(2\pi x_0)^2} e^{-2Mx_0} \left\| P\left(e^{i\cdot}\right) \right\|_{L^2(-2x_0, 2x_0)}^2 .\end{aligned}$$

Now, using that, for $k \in \mathbb{N}_0$, $t \mapsto \mathcal{F}[\varphi](t - k + m/2)$ have disjoint supports, we obtain

$$\begin{aligned}\|g\|_{L^2(\mathbb{R})}^2 &= \frac{1}{2\pi} \|\mathcal{F}[g]\|_{L^2(\mathbb{R})}^2 \\ &= \frac{1}{2\pi} \sum_{k=0}^m |P_k|^2 \|\mathcal{F}[\varphi]\|_{L^2(\mathbb{R})}^2 \\ &= \frac{1}{(2\pi)^2} \left\| P\left(e^{i\cdot}\right) \right\|_{L^2(-\pi, \pi)}^2 ,\end{aligned}$$

hence, by (6.12),

$$\rho_m^c \geq \frac{4 \sin(x_0)^2}{x_0} e^{-2Mx_0} \left(\frac{7e\pi}{x_0} \right)^{-2m} .$$

We obtain, for $0 < x_0 \leq \pi/2$ and $m \in \mathbb{N}_0$,

$$\rho_m^c \geq \frac{4 \sin(x_0)^2}{x_0} \exp\left(-\frac{x_0}{c}(m+1) - 2 \log\left(\frac{7e\pi}{x_0} \right) m \right) .$$

Thus, we have, for all $m \in \mathbb{N}_0$,

$$\rho_m^c \geq 4 e^{-2 \log(7e\pi) m} \sup_{x_0 \in (0, \pi/2]} \frac{\sin(x_0)^2}{x_0} e^{-x_0/c} \exp\left(-\left(\frac{x_0}{c} - 2 \log(x_0) \right) m \right) . \tag{6.14}$$

Using that if $2c < \pi$, $x_0 \mapsto x_0/c - 2\log(x_0)$ admits a minimum at $x_0 = 2c$, we obtain, for all $0 < c \leq \pi/4$,

$$\rho_m^c \geq \frac{2\sin(2c)^2}{e^2 c} \exp\left(-2\log\left(\frac{7e^2\pi}{2c}\right) m\right).$$

We now prove the second bound on ρ_m^c. Let $m \in \mathbb{N}_0$ and $t_m = \pi(m+1)/2$. For all $x \in \mathbb{R}$, we have $\operatorname{sech}(x/c) \geq \operatorname{sech}(t_m/c)\, \mathbb{1}\{|x| \leq t_m\}$, hence, by (6.9), we have

$$\begin{aligned}
\rho_m^c &= \max_{V \in S_{m+1}} \min_{f \in V \setminus \{0\}} \int_{\mathbb{R}} \operatorname{sech}\left(\frac{x}{c}\right) |\mathcal{F}[\mathcal{E}[f]](x)|^2 dx \frac{1}{\|f\|^2_{L^2(-1,1)}} \\
&\geq \operatorname{sech}\left(\frac{t_m}{c}\right) \max_{V \in S_{m+1}} \min_{f \in V \setminus \{0\}} \int_{\mathbb{R}} \mathbb{1}\{|x| \leq t_m\} |\mathcal{F}[\mathcal{E}[f]](x)|^2 dx \frac{1}{\|f\|^2_{L^2(-1,1)}} \\
&\geq \operatorname{sech}\left(\frac{t_m}{c}\right) \rho_m^{W_{[-1,1]}, t_m}. \qquad (6.15)
\end{aligned}$$

Using that $m = 2t_m/\pi - 1$ and (5.2) in Bonami et al. (2018) (with a difference by a factor $1/(2\pi)$ in the normalisation of $\mathcal{Q}_c^{W_{[-1,1]}}$), we have $\rho_m^{W_{[-1,1]}, t_m} \geq \pi$ hence, for all $m \in \mathbb{N}_0$,

$$\begin{aligned}
\rho_m^c &\geq \exp\left(-\frac{t_m}{c}\right) \rho_m^{W_{[-1,1]}, t_m} \quad \text{(by (6.15))} \\
&\geq \pi \exp\left(-\frac{\pi(m+1)}{2c}\right).
\end{aligned}$$

□

The best lower bound in terms of the factor in the exponential is (6.10) for $c \leq c_0$, where $c_0 = 0.12059$, and (6.11) for larger c (see Figure 6-1). This yields

Corollary 1. *For all $c > 0$,*

$$\forall m \in \mathbb{N}_0,\ \rho_m^c \geq \theta(c) e^{-2\beta(c)m}, \qquad (6.16)$$

where

$$\beta : c \mapsto \log\left(\frac{7e^2\pi}{2c}\right)1\!\!1\{c \le c_0\} + \frac{\pi}{4c}1\!\!1\{c > c_0\},$$
$$\theta : c \mapsto \frac{2\sin(2c)^2}{e^2 c}1\!\!1\{c \le c_0\} + \frac{\pi}{e^{\pi/(2c)}}1\!\!1\{c > c_0\}.$$

Clearly, because $c_0 \le \pi/4$ *and* $x \mapsto \sin(x)/x$ *is decreasing on* $(0, \pi/2]$*, the lower bound holds when we replace* θ *by*

$$\widetilde{\theta} : \ c \mapsto \frac{2\sin(2c_0)^2 c}{(ec_0)^2}1\!\!1\{c \le c_0\} + \frac{\pi}{e^{\pi/(2c)}}1\!\!1\{c > c_0\}.$$

6.3.2 Application: Error bounds for stable analytic continuation of functions whose Fourier transform belongs to $L^2(\cosh(b\cdot))$

In this section, we consider the problem where we observe the function f with error on $(x_0 - c, x_0 + c)$, for $c > 0$ and $x_0 \in \mathbb{R}$,

$$f_\delta(cx+x_0) = f(cx+x_0)+\delta\xi(x), \quad \text{for } a.e.\ x \in (-1,1), \quad \mathcal{F}[f] \in L^2(\cosh(b\cdot)), \tag{6.17}$$

where $\xi \in L^2(-1,1)$, $\|\xi\|_{L^2(-1,1)} \le 1$, and $\delta > 0$. We consider the problem of approximating $f_0 = f$ on $L^2(\mathbb{R})$ from f_δ on $(x_0 - c, x_0 + c)$. This is a classical problem for which an approach based on PSWF is prone to criticism when the researcher does not have a priori information on the bandlimits or when she questions the bandlimited assumption. As we have stressed before such an assumption makes little sense for probability densities.

Noting that, for a.e. $x \in (-1,1)$,

$$\frac{1}{2\pi}\mathcal{F}_{b,c}\left[\mathcal{F}\left[f(x_0 - \cdot)\right]\right](x) = f(cx + x_0) \tag{6.18}$$

suggests the two steps regularising procedure:

1. approximate $\mathcal{F}\left[f(x_0-\cdot)\right]/(2\pi) \in L^2(\cosh(b\cdot))$ by the spectral cut-off regularization,
$$F_\delta^N = \sum_{m\le N} \frac{1}{\sigma_m^{b,c}} \left\langle f_\delta(c\cdot+x_0), g_m^{c/b}(\cdot)\right\rangle_{L^2(-1,1)} \varphi_m^{b,c}, \tag{6.19}$$

2. take the inverse Fourier transform and define
$$f_\delta^N(\cdot) = 2\pi\mathcal{F}^I\left[F_\delta^N\right](x_0-\cdot). \tag{6.20}$$

These steps require numerical approximations of an inner product, of an inverse Fourier transform over $\mathbb{R}$, and of the singular functions. Sections 6.7 and 6.8 address these issues. The lower bounds on the eigenvalues of $\mathcal{Q}_{c/b}$ of Theorem 1 are useful to obtain rates of convergence when $\mathcal{F}[f]$, which appears on the left-hand side of (6.18), satisfies a source condition: $f \in \mathcal{H}_{\omega,x_0}^{b,c}(M)$, where

$$\mathcal{H}_{\omega,x_0}^{b,c}(M) = \left\{ f: \quad \sum_{m\in\mathbb{N}_0} \omega_m^2 \left| \left\langle \mathcal{F}[f(x_0-\cdot)], \varphi_m^{b,c}\right\rangle_{L^2(\cosh(b\cdot))}\right|^2 \le M^2 \right\}$$

for a given sequence $(\omega_m)_{m\in\mathbb{N}_0}$. The set can also be written as

$$\mathcal{H}_{\omega,x_0}^{b,c}(M) = \left\{ f: \quad \sum_{m\in\mathbb{N}_0} \left(2\pi\frac{\omega_m}{\sigma_m^{b,c}}\right)^2 \left| \left\langle f(c\cdot+x_0), g_m^{b,c}\right\rangle_{L^2(-1,1)}\right|^2 \le M^2 \right\}.$$

This amounts to smoothness of $f(c\cdot+x_0)$ on $(-1,1)$. When $\omega_m = 1$ for all m this corresponds to analyticity of f in $\mathbb{R}$. We consider below the case where we have a preexponential polynomial or exponential sequence ω_m. Theorem 1 in Bonami and Karoui (2017) provides a comparison between the smoothness in terms of a summability condition involving the coefficients on the PSWF basis and Sobolev smoothness on $(-1,1)$. Such a result is not available when the PSWF basis is replaced by $(g_m^{b,c})_{m\in\mathbb{N}_0}$ and requires further investigation.

Theorem 2. *Take $M>0$ and define β as in (6.16), then we have*

1. *for* $(\omega_m)_{m\in\mathbb{N}_0} = (m^\sigma)_{m\in\mathbb{N}_0}$*,* $\sigma > 1/2$*,* $N = \lfloor \overline{N} \rfloor$*, and* $\overline{N} = \ln(1/\delta)/(2\beta(c/b))$*,*

$$\sup_{f\in\mathcal{H}^{b,c}_{\omega,x_0}(M),\|\xi\|_{L^2(-1,1)}\le 1} \left\|f^N_\delta - f\right\|_{L^2(\mathbb{R})} \underset{\delta\to 0}{=} O((-\log(\delta))^{-\sigma}), \tag{6.21}$$

2. *for* $(\omega_m)_{m\in\mathbb{N}_0} = (e^{\kappa m})_{m\in\mathbb{N}_0}$*,* $\kappa > 0$*,* $N = \lfloor \overline{N} \rfloor$*, and* $\overline{N} = \ln(1/\delta)/(\kappa + \beta(c/b))$*,*

$$\sup_{f\in\mathcal{H}^{b,c}_{\omega,x_0}(M),\|\xi\|_{L^2(-1,1)}\le 1} \left\|f^N_\delta - f\right\|_{L^2(\mathbb{R})} \underset{\delta\to 0}{=} O\left(\delta^{\kappa/(\kappa+\beta(c/b))}\right). \tag{6.22}$$

Proof. We have, using the Plancherel equality for the first equality,

$$\begin{aligned}
\left\|f^N_\delta - f\right\|^2_{L^2(\mathbb{R})} =& \frac{1}{2\pi}\left\|\mathcal{F}\left[f^N_\delta\right] - \mathcal{F}[f]\right\|^2_{L^2(\mathbb{R})}\\
=& \frac{1}{2\pi}\left\|\mathcal{F}\left[f^N_\delta(x_0-\cdot)\right] - \mathcal{F}[f(x_0-\cdot)]\right\|^2_{L^2(\mathbb{R})}\\
\le& \frac{1}{2\pi}\left\|\mathcal{F}\left[f^N_\delta(x_0-\cdot)\right] - \mathcal{F}[f(x_0-\cdot)]\right\|^2_{L^2(\cosh(b\cdot))}\\
\le& \frac{1}{\pi}\left\|\mathcal{F}\left[f^N_\delta(x_0-\cdot)\right] - \mathcal{F}\left[f^N_0(x_0-\cdot)\right]\right\|^2_{L^2(\cosh(b\cdot))}\\
&+ \frac{1}{\pi}\left\|\mathcal{F}\left[f^N_0(x_0-\cdot)\right] - \mathcal{F}[f(x_0-\cdot)]\right\|^2_{L^2(\cosh(b\cdot))}.
\end{aligned} \tag{6.23}$$

Using (6.19) for the first equality, the Cauchy-Schwarz inequality and (6.4) for the first inequality, and (6.16) for the second inequality, we obtain

$$\begin{aligned}
&\left\|\mathcal{F}\left[f^N_\delta(x_0-\cdot)\right] - \mathcal{F}\left[f^N_0(x_0-\cdot)\right]\right\|^2_{L^2(\cosh(b\cdot))}\\
&= \left\|\sum_{m\le N}\frac{2\pi}{\sigma^{b,c}_m}\left\langle (f_\delta - f)(c\cdot + x_0), g^{c/b}_m(\cdot)\right\rangle_{L^2(-1,1)}\varphi^{b,c}_m(\cdot)\right\|^2_{L^2(\cosh(b\cdot))}\\
&= \sum_{m\le N}\left(\frac{2\pi}{\sigma^{b,c}_m}\right)^2\left|\left\langle (f_\delta - f)(c\cdot + x_0), g^{c/b}_m(\cdot)\right\rangle_{L^2(-1,1)}\right|^2\\
&\le (2\pi)^2\left\|(f_\delta - f)(c\cdot + x_0)\right\|^2_{L^2(-1,1)}\sum_{m\le N}\frac{c}{\rho^{c/b}_m}\\
&\le \frac{(2\pi)^2 c\delta^2}{\theta(c)}\|\xi\|^2_{L^2(-1,1)}\sum_{m\le N} e^{2\beta(c/b)m}\\
&\le \frac{(2\pi)^2 c\delta^2}{\theta(c)\left(1 - e^{-2\beta(c/b)}\right)} e^{2\beta(c/b)N}.
\end{aligned} \tag{6.24}$$

Using (6.20), we have

$$\begin{aligned}\mathcal{F}\left[f_0^N(x_0-\cdot)\right](\star) &= \sum_{m\le N}\frac{2\pi}{\sigma_m^{b,c}}\left\langle f(c\cdot+x_0), g_m^{c/b}(\cdot)\right\rangle_{L^2(-1,1)}\varphi_m^{b,c}(\star)\\ &= \sum_{m\le N}\frac{2\pi}{\sigma_m^{b,c}}\left\langle \mathcal{F}_{b,c}\left[\frac{1}{2\pi}\mathcal{F}\left[f(x_0-\cdot)\right]\right], g_m^{c/b}\right\rangle_{L^2(-1,1)}\varphi_m^{b,c}(\star)\\ &= \sum_{m\le N}\frac{1}{\sigma_m^{b,c}}\left\langle \mathcal{F}\left[f(x_0-\cdot)\right], \mathcal{F}_{b,c}^{*}\left[g_m^{c/b}\right]\right\rangle_{L^2(\cosh(b\cdot))}\varphi_m^{b,c}(\star)\\ &= \sum_{m\le N}\left\langle \mathcal{F}\left[f(x_0-\cdot)\right], \varphi_m^{c/b}\right\rangle_{L^2(\cosh(b\cdot))}\varphi_m^{b,c}(\star).\end{aligned}$$

Thus, using Proposition 2 and Pythagoras' theorem, we obtain

$$\begin{aligned}\left\|\mathcal{F}\left[f_0^N(x_0-\cdot)\right]-\mathcal{F}\left[f(x_0-\cdot)\right]\right\|_{L^2(\cosh(b\cdot))}^2 &= \sum_{m>N}\left|\left\langle\mathcal{F}[f(x_0-\cdot)],\varphi_m^{b,c}(\cdot)\right\rangle_{L^2(\cosh(b\cdot))}\right|^2\\ &\le \sum_{m\in\mathbb{N}_0}\left(\frac{\omega_m}{\omega_N}\right)^2\left|\left\langle\mathcal{F}[f(x_0-\cdot)],\varphi_m^{b,c}(\cdot)\right\rangle_{L^2(\cosh(b\cdot))}\right|^2\\ &\le \frac{M^2}{\omega_N^2}\quad\left(\text{using } f\in\mathcal{H}_{\omega,x_0}^{b,c}(M)\right). \qquad (6.25)\end{aligned}$$

Finally, using (6.23)-(6.25) yields

$$\left\|f_\delta^N-f\right\|_{L^2(\mathbb{R})}^2 \le \frac{1}{\pi}\left(\frac{(2\pi)^2c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\delta^2e^{2\beta(c/b)N}+\frac{M^2}{\omega_N^2}\right). \qquad (6.26)$$

Consider case (1). Take δ small enough so that $\overline{N}\ge 2$ and $\log\left(\delta\log(1/\delta)^{2\sigma}\right)\le 0$. By (6.26) and the definition of $(\omega_N)_{n\in\mathbb{N}_0}$ in the first display below, $\overline{N}-1\le N\le\overline{N}$ in the second display, and $\overline{N}\ge 2$ in the third display, we obtain

$$\begin{aligned}\left\|f_\delta^N-f\right\|_{L^2(\mathbb{R})}^2 &\le \frac{N^{-2\sigma}}{\pi}\left(\frac{(2\pi)^2c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\delta^2e^{2\beta(c/b)N}N^{2\sigma}+M^2\right)\\ &\le \frac{\overline{N}^{-2\sigma}\left(1-1/\overline{N}\right)^{-2\sigma}}{\pi}\left(\frac{(2\pi)^2c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\delta^2e^{2\beta(c/b)\overline{N}}\overline{N}^{2\sigma}+M^2\right)\\ &\le \frac{\overline{N}^{-2\sigma}2^{2\sigma}}{\pi}\left(\frac{(2\pi)^2c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\delta^2e^{2\beta(c/b)\overline{N}}\overline{N}^{2\sigma}+M^2\right).\end{aligned}$$

Using that

$$\delta^2 \exp\left(2\beta\left(\frac{c}{b}\right)\overline{N}\right)\overline{N}^{2\sigma} = \exp\left(2\sigma\log\left(\frac{1}{2\beta(c/b)}\right) + \log\left(\delta\log\left(\frac{1}{\delta}\right)^{2\sigma}\right)\right) \le \left(\frac{1}{\beta(c/b)}\right)^{2\sigma},$$

yields

$$\left\|f_\delta^N - f\right\|_{L^2(\mathbb{R})}^2 \le \frac{1}{\pi}\left(4\beta\left(\frac{c}{b}\right)\right)^{2\sigma}\left(\frac{(2\pi)^2 c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\left(\frac{1}{\beta(c/b)}\frac{\sigma}{e}\right)^{2\sigma} + M^2\right)(-\log(\delta))^{-2\sigma}, \tag{6.27}$$

hence the result.

Consider now case (2).

Using $\overline{N} - 1 \le N \le \overline{N}$ in the first display and $\delta^2 \exp\left(2\left(\beta\left(\frac{c}{b}\right)+\kappa\right)\overline{N}\right) = 1$ and the definition of $\overline{N}$ in the second display, yields

$$\begin{aligned}\left\|f_\delta^N - f\right\|_{L^2(\mathbb{R})}^2 &\le \frac{e^{-2\kappa\left(\overline{N}-1\right)}}{\pi}\left(\frac{(2\pi)^2 c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)}\delta^2 e^{2(\beta(c/b)+\kappa)\overline{N}} + M^2\right)\\ &\le \frac{e^{2\kappa}}{\pi}\left(\frac{(2\pi)^2 c}{\theta(c)\left(1-e^{-2\beta(c/b)}\right)} + M^2\right)\delta^{2\kappa/(\kappa+\beta(c/b))},\end{aligned}$$

hence the result. □

The rate in (6.21) does not depend on c but the constant blows up as $c \to 0$ (see (6.27)). In contrast, the rate in (6.22) deteriorates for small values of c. The result (6.22) is related to those obtained for the so-called "2exp-severely ill-posed problems" (see Cavalier et al. (2004) for a survey and Tsybakov (2000) which obtains similar polynomial rates) where the singular values decay exponentially and the functions are supersmooth.

The proof of Theorem 2 requires an upper bound on a sum involving the singular values for small m in the denominator. Theorem 1 allows to obtain (6.24). Without it, one could at best obtain, instead of (6.24), the upper bound $(2\pi)^2 c\delta^2(N+1)/\rho_N^{b,c}$. Because (6.6) is an equivalent of the logarithm we are unable to obtain a polynomial rate of convergence as sharp as in (6.22).

6.4 Upper bounds on the eigenvalues of $\mathcal{Q}_c$

Theorem 3. *For $m \in \mathbb{N}_0$ and $0 < c < 1$, we have*

$$\rho_m^c \leq \frac{2ec^{2m+1}}{\sqrt{2m+1}(1-c^2)}.$$

The proof of this result uses arguments which have been used in the proof of Theorem 3.1 in Bonami et al. (2018) for the PSWF.

Proof. Using the minimax principle (see Theorem 4 page 212 in Birman and Solomjak (2012)), the $m+1$-th eigenvalue ρ_m^c satisfies

$$\rho_m^c = \min_{V \in S_m} \max_{f \in V^\perp} \frac{\langle \mathcal{Q}_c f, f \rangle_{L^2(-1,1)}}{\|f\|_{L^2(-1,1)}^2},$$

where S_m is the set of m-dimensional vector subspaces of $L^2(-1,1)$. We use (6.8), which yields

$$\rho_m^c = \min_{V \in S_m} \max_{f \in V^\perp} \frac{c \left\langle \mathcal{F}_{1,c}^*[f], \mathcal{F}_{1,c}^*[f] \right\rangle_{L^2(\cosh)}}{\|f\|_{L^2(-1,1)}^2}.$$

We denote by $(P_m)_{m \in \mathbb{N}_0}$ the Legendre polynomials with the normalization $P_m(1) = 1$. They are such that $\left(\sqrt{m+1/2} P_m\right)_{m \in \mathbb{N}_0}$ is an orthonormal basis of $L^2(-1,1)$. Let V be the vector space spanned by $P_0, \dots, P_{m-1}$. Take $f \in V^\perp$ of norm 1. It is of the form $f = \sum_{k=m}^\infty a_k \sqrt{k+1/2} P_k$, where $\sum_{k=m}^\infty |a_k|^2 = 1$. The Cauchy-Schwarz inequality yields, for *a.e.* $x \in \mathbb{R}$,

$$\left|\mathcal{F}_{1,c}^* f(x)\right|^2 \leq \left(\sum_{k=m}^\infty |a_k|^2\right) \left(\sum_{k=m}^\infty \left(k + \frac{1}{2}\right) \left|\mathcal{F}_{1,c}^* P_k(x)\right|^2\right)$$

and after integration

$$\left\|\mathcal{F}_{1,c}^* f\right\|_{L^2(\cosh)}^2 \leq \sum_{k=m}^\infty \left(k + \frac{1}{2}\right) \left\|\mathcal{F}_{1,c}^* P_k\right\|_{L^2(\cosh)}^2.$$

Thus, we have

$$\rho_m^c \leq c \sum_{k=m}^{\infty} \left(k + \frac{1}{2}\right) \left\|\mathcal{F}_{1,c}^* P_k\right\|_{L^2(\cosh)}^2. \tag{6.28}$$

Then, using (18.17.19) in Olver et al. (2010), we obtain, for *a.e.* x and $c > 0$,

$$\begin{aligned}\mathcal{F}_{1,c}^*\left[P_k\right](x) &= \operatorname{sech}(x)\, \mathcal{F}_{1,c}\left[\mathcal{E}\left[P_k\right]\right](-x) \\ &= \operatorname{sech}(x)\, i^{-k} \sqrt{\frac{2\pi}{c\,|x|}} J_{k+1/2}(c\,|x|),\end{aligned}$$

where $J_{k+1/2}$ is the Bessel function of order $k+1/2$. Using (9.1.62) in Abramowitz and Stegun (1965) in the first display, $\Gamma(k+3/2) = (k+1/2)\Gamma(k+1/2)$ and $\Gamma(k+1/2) \geq \sqrt{2\pi}e^{-k-1/2}(k+1/2)^k$ (see (1.4) in Mortici and Chen (2011)) in the second display, we obtain

$$\begin{aligned}\left|\mathcal{F}_{1,c}^*\left[P_k\right](x)\right| &\leq \operatorname{sech}(x)\, \sqrt{\pi} \frac{|cx/2|^k}{\Gamma(k+3/2)}. \\ &\leq \operatorname{sech}(x) \sqrt{\frac{e}{2}} \frac{1}{k+1/2} \left(\frac{ec}{2(k+1/2)}\right)^k |x|^k.\end{aligned}$$

Thus, we have

$$\begin{aligned}\left\|\mathcal{F}_{1,c}^*\left[P_k\right]\right\|_{L^2(\cosh)}^2 &\leq \frac{e}{2(k+1/2)^2} \left(\frac{ec}{2(k+1/2)}\right)^{2k} \int_{\mathbb{R}} x^{2k} \operatorname{sech}(x)\, dx \\ &\leq \frac{e}{(k+1/2)^2} \left(\frac{ec}{2(k+1/2)}\right)^{2k} \int_0^{\infty} x^{2k} e^{-x} dx \\ &\leq \frac{e}{(k+1/2)^2} \left(\frac{ec}{2(k+1/2)}\right)^{2k} \Gamma(2k+1).\end{aligned}$$

Then, by (6.28) for the first inequality and using (1.3) in Mortici and Chen (2011) for the second one, we have

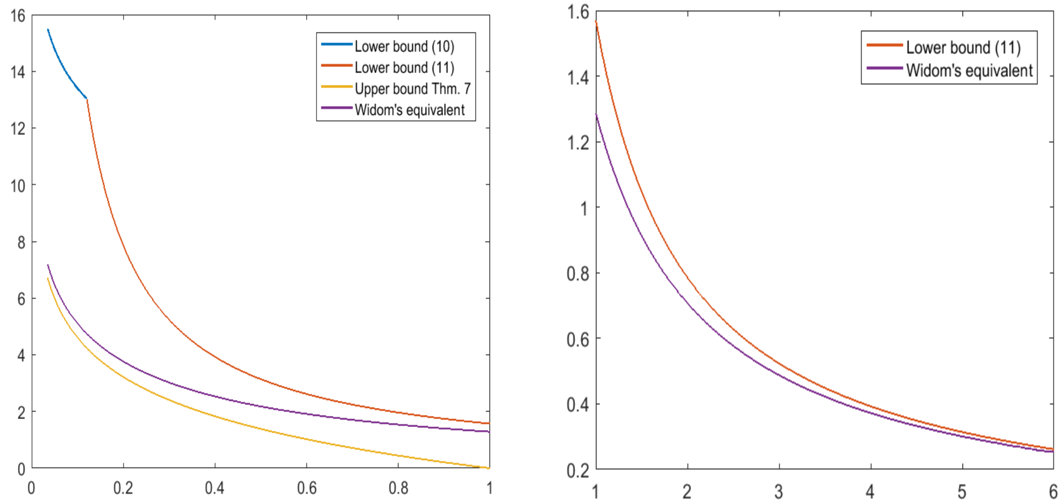
$$\begin{aligned}
\rho_m^c &\leq ec \sum_{k=m}^{\infty} \frac{1}{k+1/2} \left(\frac{ec}{2(k+1/2)} \right)^{2k} \Gamma(2k+1) \\
&\leq ec \sum_{k=m}^{\infty} \frac{1}{k+1/2} \left(\frac{ec}{2(k+1/2)} \right)^{2k} (2k+1)^{2k+1/2} e^{-2k} \\
&= 2ec \sum_{k=m}^{\infty} \frac{1}{\sqrt{2k+1}} c^{2k} \\
&\leq \frac{2ec}{\sqrt{2m+1}} \sum_{k=m}^{\infty} c^{2k}
\end{aligned}$$

hence, for $0 < c < 1$, this yields the result. □

Theorem 3 holds for a limited range of values of c but this range is enough to construct the so-called test functions to prove the minimax lower bounds in Gaillac and Gautier (2019c). Note also that, using Corollary 1 and Theorem 3, we have, for all $0 < c \leq c_0 < 1$ and $m \in \mathbb{N}$,

$$\frac{2\sin(2c_0)^2 c}{(ec_0)^2} \exp\left(-2 \left(\log\left(\frac{1}{c}\right) + 2 + \log\left(\frac{7\pi}{2}\right) \right) m \right) \leq \rho_m^c \leq \frac{2ec_0 \exp\left(-2\log\left(\frac{1}{c}\right) m \right)}{\sqrt{2m+1}(1-c_0^2)}.$$

The exponential factors in these upper and lower bounds have a similar behavior as c approaches 0 (see also Figure 6-1).



(a) $c < 1$ (b) $c \geq 1$

Figure 6-1: Bounds on $\lim_{m\to\infty} -\log(\rho_m^c)/m$ and Widom's equivalent (6.6) as a function of c.

6.5 Properties of a differential operator which commutes with $\mathcal{Q}_c$

In this section, we consider differential operators $\mathcal{L}[\psi] = -(p\psi')' + q\psi$ on $L^2(-1,1)$, with (1) $p(x) = \cosh(4c) - \cosh(4cx)$ and $q(x) = 3c^2\cosh(4cx)$, (2) $p(x) = 1 - x^2$ and $q(x) = q^c(x)$ where, for $Y(x) = \sin(X(x))$, $X(x) = (\pi/U(c))\int_0^x p(\xi)^{-1/2}d\xi$, and

$$U(c) = \int_{-1}^{1} p(\xi)^{-1/2}d\xi,$$

$$q^c(Y(x)) = \frac{1}{2} + \frac{1}{4}\tan(X(x))^2 - \left(\frac{U(c)c}{\pi}\right)^2\left(\cosh(4cx) + \frac{\sinh^2(4cx)}{p(x)}\right), \qquad (6.29)$$

and (3) $p(x) = 1 - x^2$ and $q(x) = 0$. By Widom (1964) (see also Morrison (1962)), the eigenfunctions of $\mathcal{Q}_c$ are those of the differential operator in case (1) with domain $\mathcal{D} \subset \mathcal{D}_{\max} = \{\psi \in L^2(-1,1) : \mathcal{L}[\psi] \in L^2(-1,1)\}$ with boundary conditions of continuity at ± 1. This is an important property for the asymptotic analysis in Widom (1964) and to obtain bounds on the sup-norm of these functions in Section 6.6 and numerical approximations of them in Section 6.7. To study $\mathcal{L}$ in case (1), Widom (1964) uses the changes of variable and function, for all $x \in (-1,1)$ and $\psi \in \mathcal{D}_{\max}$,

$$y = Y(x), \tag{6.30}$$

$$\forall y \in (-1,1),\ \Gamma(y) = F(y)\psi\left(Y^{-1}(y)\right),\ F(y) = \left(\frac{p(Y^{-1}(y))}{1-y^2}\right)^{1/4}, \tag{6.31}$$

where Y is a $C^\infty-$diffeomorphism on $(-1,1)$. This relates an eigenvalue problem for (1) to an eigenvalue problem for (2) and it is useful to view the operator in case (2) as a perturbation of the operator in case (3). In the three cases, $1/p$ and q are holomorphic on a simply connected open set $(-1,1) \subseteq E \subseteq \mathbb{C}$. The spectral analysis involves the solutions to (H_λ): $-(p\psi')' + (q-\lambda)\psi = 0$ with $\lambda \in \mathbb{C}$ which are holomorphic on E and span a vector space of dimension 2 (see Sections IV 1 and 10 in Hartman (1987)). So they are infinitely differentiable on $(-1,1)$, have isolated zeros in $(-1,1)$, and the condition of continuity (or boundedness) at ± 1 makes sense.

We now present a few useful estimates.

Lemma 2. *We have, for all $c > 0$,*

$$\frac{\sqrt{2}e^{2c}}{\sinh(4c)} < U(c) < \pi\frac{\sqrt{2}e^{2c}}{\sinh(4c)}.$$

Proof. By the second equation page 229 of Widom (1964)

$$U(c) = \frac{1}{c(1+\cosh(4c))^{1/2}} K\left(\frac{e^{4c}-1}{e^{4c}+1}\right),$$

and the result follows from the fact that, by Corollary 3.3 in Anderson et al. (1992),

$$\frac{ce^{2c}}{\sinh(2c)} < K\left(\frac{e^{4c}-1}{e^{4c}+1}\right) < \frac{\pi ce^{2c}}{\sinh(2c)}.$$

We obtain the final expressions by classical relations between hyperbolic functions. □

We make use of the identity, for all $x \in [-1,1]$,

$$p(x) = 4c\sinh(4c)(1-x)(1+u(x)), \tag{6.32}$$

$$u(x) = \int_x^1 \frac{4c\cosh(4ct)}{\sinh(4c)(1-x)}(x-t)dt, \tag{6.33}$$

which is obtained by Taylor's theorem with remainder in integral form

$$\cosh(4cx) = \cosh(4c) + (x-1)4c\sinh(4c) + \int_1^x 16c^2\cosh(4ct)(x-t)dt.$$

Also, u is increasing on $[-1,1]$ because, for all $x \in [-1,1]$,

$$u'(x) = \frac{4c}{\sinh(4c)(1-x)^2}\int_x^1 \cosh(4ct)(1-t)dt > 0 \tag{6.34}$$

and, for all $x \in [0,1]$,

$$-1 + \frac{1}{4c\sinh(4c)}\left(\cosh(4c)-1\right) \le u(x) \le 0. \tag{6.35}$$

Lemma 3. *We have, for all $c > 0$ and $x \in [0,1]$,*

$$\frac{c\sinh(4c)}{1-x} - \frac{8c^3\sinh(4c)\cosh(4c)}{3\left(\cosh(4c)-1\right)} \le \left(\int_x^1 p(\xi)^{-1/2}d\xi\right)^{-2} \le \frac{c\sinh(4c)}{1-x}.$$

Proof. We have

$$\begin{aligned}
&\left(\int_x^1 p(\xi)^{-1/2}d\xi\right)^{-2}\\
&= \left(\frac{1}{(4c\sinh(4c))^{1/2}}\left(2(1-x)^{1/2} - \int_x^1\int_1^\xi \frac{u'(t)}{2\sqrt{1-\xi}(1+u(t))^{3/2}}d\xi dt\right)\right)^{-2}\\
&= \frac{c\sinh(4c)}{1-x}\frac{1}{(1+\widetilde{u}(x))^2} \qquad (6.36)\\
&= \frac{c\sinh(4c)}{1-x} - \frac{c\sinh(4c)(2+\widetilde{u}(x))\widetilde{u}(x)}{(1-x)(1+\widetilde{u}(x))^2}, \qquad (6.37)
\end{aligned}$$

where

$$\widetilde{u}(x) = \int_x^1 \frac{1}{4\sqrt{(1-\xi)(1-x)}} \left(\int_\xi^1 \frac{u'(t)}{(1+u(t))^{3/2}} dt \right) d\xi. \tag{6.38}$$

The upper bound in Lemma 3 uses that, for all $x \in [0,1]$, $\widetilde{u}(x) \geq 0$. We now consider the lower bound. By (6.34), u is a C^1-diffeomorphism and

$$\int_x^1 \frac{u'(t)dt}{(1+u(t))^2} dt = -\frac{u(x)}{1+u(x)}. \tag{6.39}$$

Now, by (6.33), we have, for all $x \in [0,1]$,

$$\begin{aligned} -u(x) &\leq \frac{4c\cosh(4c)}{\sinh(4c)(1-x)} \int_x^1 (t-x)dt \\ &= \frac{2c\cosh(4c)}{\sinh(4c)}(1-x), \end{aligned}$$

and, by (6.35),

$$\int_x^1 \frac{u'(t)dt}{(1+u(t))^2} dt \leq 8c^2 \frac{\cosh(4c)}{\cosh(4c)-1}(1-x). \tag{6.40}$$

Now, using that $\widetilde{u}(x) \geq 0$ and that $g: \ t \mapsto (2+t)/(1+t)^2$ is decreasing on $[0,\infty)$ hence $g(t)t \leq 2t$ for $t \geq 0$, we have

$$\begin{aligned} \frac{(2+\widetilde{u}(x))\widetilde{u}(x)}{(1+\widetilde{u}(x))^2} &\leq \int_x^1 \frac{1}{2\sqrt{(1-\xi)(1-x)}} \left(\int_\xi^1 \frac{u'(t)}{(1+u(t))^2} dt \right) d\xi \quad \text{(by (6.38))} \\ &\leq \frac{4c^2\cosh(4c)}{\cosh(4c)-1} \int_x^1 \sqrt{\frac{1-\xi}{1-x}} d\xi \quad \text{(by (6.40))} \\ &\leq \frac{8c^2\cosh(4c)}{3\,(\cosh(4c)-1)}(1-x). \end{aligned} \tag{6.41}$$

□

Proposition 3. *F is such that*

$$\|F\|^4_{L^\infty([-1,1])} \leq 2\pi^2 e^{4c} c^2 \tag{6.42}$$

$$\|1/F\|^4_{L^\infty([-1,1])} \leq \frac{\pi^2 e^{-4c}}{4c} \left(1 + \frac{4c^2}{3}\right)^2 \coth(2c). \tag{6.43}$$

For all $c > 0$ and $\lambda \in \mathbb{C}$, the change of variables and function (6.30)-(6.31) maps a solution of $\left(H_{U(c)^2\lambda/\pi^2}\right)$ in case (2) to a solution of (H_λ) in case (1) and reciprocally the inverse transformation maps a solution of (H_λ) in case (1) to a solution of $\left(H_{U(c)^2\lambda/\pi^2}\right)$ in case (2) and is a bijection of $\mathcal{D}$. Also, q^c can be extended by continuity to $[-1,1]$ and, for all $y \in [-1,1]$,

$$\frac{1}{2} - \left(\frac{U(c)c}{\pi}\right)^2 - R(c) \leq q^c(y) \leq \frac{1}{2} - \left(\frac{U(c)c}{\pi}\right)^2, \tag{6.44}$$

where

$$R(c) = \frac{2}{\pi^2} + \left(\frac{U(c)c}{\pi}\right)^2 \left(\left(\cosh(4c)\left(1 + \frac{c}{3}\coth(2c)\right) - 1\right) + 2c\sinh(4c)\right).$$

Proof. To prove (6.42) and (6.43) it is sufficient, by parity, to consider $x \in [0,1]$. (6.42) is a obtained by the following sequence of inequalities

$$\begin{aligned}
\frac{p(x)}{1-Y(x)^2} &= \frac{p(x)}{\sin^2\left(\pi\int_x^1 p(\xi)^{-1/2}d\xi/U(c)\right)} \\
&\leq \left(\frac{U(c)}{2}\right)^2 \frac{p(x)}{\left(\int_x^1 p(\xi)^{-1/2}d\xi\right)^2} \quad (\text{because } \sin(x) \geq 2x/\pi) \\
&\leq \left(\frac{U(c)}{2}\right)^2 p(x)\frac{c\sinh(4c)}{1-x} \quad (\text{by Lemma 3}) \\
&\leq \left(\frac{U(c)}{2}\right)^2 4c^2\sinh(4c)^2(1+u(x)) \quad (\text{by (6.32)}) \\
&\leq \frac{\pi^2 e^{4c}c^2\sinh(4c)^2}{\sinh(2c)^2(1+\cosh(4c))} \quad (\text{by Lemma 2 and (6.35)}).
\end{aligned}$$

We obtain (6.43) by the inequalities below. Using for the first display that, for $x \in [0,\pi/2]$, $\sin(x) \leq x$, (6.32) and (6.36) for the second display, (6.35) and (6.41) for the third, and Lemma 2 for the fourth, we obtain, for all $x \in [0,1)$,

$$\begin{aligned}\frac{1-Y(x)^2}{p(x)} &\leq \left(\frac{\pi}{U(c)}\right)^2 \frac{\left(\int_x^1 p(\xi)^{-1/2} d\xi\right)^2}{p(x)} \\ &\leq \left(\frac{\pi}{U(c)}\right)^2 \frac{2\left(1+\widetilde{u}(x)\right)^2}{(4c\sinh(4c))^2} \frac{1}{1+u(x)} \\ &\leq \left(\frac{\pi}{U(c)}\right)^2 \frac{(1+4c^2/3)^2}{2c\sinh(4c)^2} \frac{\sinh(4c)}{\cosh(4c)-1} \\ &\leq \frac{\pi^2 e^{-4c}}{4c} \frac{(1+4c^2/3)^2 \sinh(4c)}{\cosh(4c)-1}.\end{aligned}$$

Classical relations between hyperbolic functions yield the final expressions for (6.42) and (6.43).

Let Γ and ψ related via (6.31). Buy the above if one function is in $\mathcal{D}$ the other is s well. Moreover, by (6.31), we have

$$F'(y) = \frac{F(y)}{4}\left(\frac{p'}{pY'}\left(Y^{-1}(y)\right) + \frac{2y}{1-y^2}\right)$$

$$(1-y^2)\Gamma'(y) = \frac{F(y)}{4}\left((1-y^2)\left(\frac{p'\psi}{pY'} + 4\frac{\psi'}{Y'}\right)\left(Y^{-1}(y)\right) + 2y\psi\left(Y^{-1}(y)\right)\right)$$

so differentiating a second time and injecting the above inequality, yields

$$\begin{aligned}&\left((1-y^2)\Gamma'\right)'(y) \\ &= \frac{F(y)}{4}\Big(\frac{1}{4}\left(\frac{p'}{pY'}\left(Y^{-1}(y)\right) + \frac{2y}{1-y^2}\right)\left((1-y^2)\left(\frac{p'\psi}{pY'} + 4\frac{\psi'}{Y'}\right)\left(Y^{-1}(y)\right) + 2y\psi\left(Y^{-1}(y)\right)\right) \\ &- 2y\left(\frac{p'\psi}{pY'} + 4\frac{\psi'}{Y'}\right)\left(Y^{-1}(y)\right) + (1-y^2)\left[\frac{1}{Y'}\left(\frac{p'\psi}{pY'} + 4\frac{\psi'}{Y'}\right)'\right]\left(Y^{-1}(y)\right) \\ &+ 2\left(\psi\left(Y^{-1}(y)\right) + y\frac{\psi'}{Y'}\left(Y^{-1}(y)\right)\right)\Big).\end{aligned}$$

Dividing by $F(y)/4$ and using (6.30), Γ is solution of $\left(H_{U(c)^2\lambda/\pi^2}\right)$ iff ψ is solution on

$(-1,1)$ of

$$\frac{1}{4p(x)}\left(p'(x)+\frac{2YY'p}{1-Y^2}(x)\right)\left(\frac{1-Y^2}{(Y')^2\,p}(x)\left(p'\psi+4p\psi'\right)(x)+2\frac{Y}{Y'}(x)\psi\left(x\right)\right)$$
$$-2\frac{Y}{Y'}(x)\left(\frac{p'\psi}{p}+4\psi'\right)(x)+\frac{1-Y^2}{Y'}(x)\left(\frac{p'\psi}{pY'}+4\frac{\psi'}{Y'}\right)'(x)+2\left(\psi\left(x\right)+\frac{Y}{Y'}(x)\psi'\left(x\right)\right)$$
$$=4\left(q^c\left(Y(x)\right)-\frac{U(c)^2\lambda}{\pi^2}\right)\psi\left(x\right).$$

We now use, for all $x\in(-1,1)$,

$$Y'(x)=\frac{\pi}{U(c)p(x)^{1/2}}\cos(X(x)), \tag{6.45}$$

which yields the equality between C^∞ functions: $(1-Y^2)/\left((Y')^2p\right)=(U(c)/\pi)^2$ and

$$\left(1+2\frac{Y}{p'Y'}\left(\frac{\pi}{U(c)}\right)^2\right)\left(\left(\frac{U(c)}{\pi}\right)^2\left(\frac{(p')^2}{4p}\psi+p'\psi'\right)+\frac{Y}{2pp'Y'}\left((p')^2\,\psi\right)\right)$$
$$-2\frac{Y}{p'Y'}\left(\frac{(p')^2}{p}\psi+4p'\psi'\right)+\left(\frac{U(c)}{\pi}\right)^2pY'\left(\frac{p'\psi}{pY'}+4\frac{\psi'}{Y'}\right)'+2\frac{Y}{p'Y'}p'\psi'$$
$$=4\left(q^c\left(Y\right)-\frac{1}{2}-\frac{U(c)^2\lambda}{\pi^2}\right)\psi.$$

The term in factor of ψ on the left-hand side of the above equality is

$$\left(\frac{U(c)}{\pi}\right)^2\left(1+2\frac{Y}{p'Y'}\left(\frac{\pi}{U(c)}\right)^2\right)^2\frac{(p')^2}{4p}-2\frac{Y}{p'Y'}\frac{(p')^2}{p}+\left(\frac{U(c)}{\pi}\right)^2\frac{pp''Y'-(p')^2\,Y'-pp'Y''}{pY'}$$

Using $-2pY''=p'Y'+2\left(\pi/U(c)\right)^2Y$ which is obtained from (6.45), this becomes

$$\left(\frac{U(c)}{\pi}\right)^2\left(1+2\frac{Y}{p'Y'}\left(\frac{\pi}{U(c)}\right)^2\right)^2\frac{(p')^2}{4p}-\frac{Y}{Y'}\frac{p'}{p}+\left(\frac{U(c)}{\pi}\right)^2\left(p''-\frac{(p')^2}{2p}\right)$$
$$=\left(\frac{Y}{Y'}\right)^2\left(\frac{\pi}{U(c)}\right)^2\frac{1}{p}+\left(\frac{U(c)}{\pi}\right)^2\left(p''-\frac{(p')^2}{4p}\right)$$
$$=\tan\left(X(x)\right)^2+\left(\frac{U(c)}{\pi}\right)^2\left(p''-\frac{(p')^2}{4p}\right).$$

hence

$$
\begin{aligned}
&4\left(\frac{U(c)}{\pi}\right)^2 (p\psi')' \\
&= 4\left(q^c\left(Y\right) - \frac{1}{2} - \frac{1}{4}\tan\left(X(x)\right)^2 + \frac{1}{4}\left(\frac{U(c)}{\pi}\right)^2\left(\frac{(p')^2}{4p} - p''\right) - \left(\frac{U(c)}{\pi}\right)^2\lambda\right)\psi
\end{aligned}
$$

and ψ is solution of (H_λ) in case (1).

We now obtain upper and lower bounds on the even function $q^c(Y(x))$, for $x \in [0,1]$, and start with the lower bound. To bound $\tan(X)^2$ in (6.29), we use

$$
\tan\left(\frac{\pi}{U(c)}\int_0^x p(\xi)^{-1/2}d\xi\right)^2 = \left(\tan\left(\frac{\pi}{U(c)}\int_x^1 p(\xi)^{-1/2}d\xi\right)\right)^{-2}, \tag{6.46}
$$

and (96) in Yang et al. (2014) in the first display and Lemma 3 and the fact that $(a-b)^2 \geq a^2 - 2ab$ for $a, b > 0$ in the second display. We obtain

$$
\begin{aligned}
\tan\left(X(x)\right)^2 &\geq \left(\frac{U(c)}{\pi}\left(\int_x^1 p(\xi)^{-1/2}d\xi\right)^{-1} - \frac{4}{\pi U(c)}\int_x^1 p(\xi)^{-1/2}d\xi\right)^2 \\
&\geq \left(\frac{U(c)c}{\pi}\right)^2\left(\frac{\sinh(4c)}{c(1-x)} - \frac{8c\sinh(4c)\cosh(4c)}{3\left(\cosh(4c)-1\right)}\right) - \frac{8}{\pi^2}.
\end{aligned}
$$

To bound the second term in the bracket in (6.29) we proceed as follows. We have

$$
\begin{aligned}
\frac{4c\sinh(4cx)^2}{p(x)} &= \frac{\sinh(4c)}{1-x}\,\frac{1}{1+u(x)}\,\frac{\sinh(4cx)^2}{\sinh(4c)^2} \quad \text{(by (6.32))} \\
&= \frac{\sinh(4c)}{1-x}\left(1 + \int_x^1 \frac{u'(t)dt}{(1+u(t))^2}\right) \quad \text{(by (6.39))} \\
&\leq \frac{\sinh(4c)}{1-x}\left(1 + 8c^2(1-x)\right) \quad \text{(by (6.40))},
\end{aligned} \tag{6.47}
$$

hence

$$
\begin{aligned}
q^c(Y(x)) \geq& \frac{1}{2} - \frac{2}{\pi^2} - \left(\frac{U(c)c}{\pi}\right)^2\left(\cosh(4c)\left(1 + \frac{2c\sinh(4c)}{3\left(\cosh(4c)-1\right)}\right) + 2c\sinh(4c)\right) \\
\geq& \frac{1}{2} - \left(\frac{U(c)c}{\pi}\right)^2 - R(c).
\end{aligned}
$$

Consider the upper bound on q^c. For $x \in [0,1]$, by (6.46) and $0 < z \leq \tan(z)$ on $(0, \pi/2]$, we have

$$q^c(Y(x)) \leq \frac{1}{2} + \left(\frac{U(c)c}{\pi}\right)^2 \left(\frac{1}{4c^2\left(\int_x^1 p(\xi)^{-1/2}d\xi\right)^2} - \frac{\sinh(4cx)^2}{p(x)} - \cosh(4cx)\right).$$

Using Lemma 3, (6.47), and (6.34), we have

$$q^c(Y(x)) \leq \frac{1}{2} - \left(\frac{U(c)c}{\pi}\right)^2. \tag{6.48}$$

□

The unbounded operator $\mathcal{L}$ on domain $\mathcal{D}$ in case (3) is self-adjoint. Indeed, it is shown page 571 of Niessen and Zettl (1992) that $\mathcal{D}$ is the domain of the self-adjoint Friedrichs extension of the minimal operator corresponding to the differential operator on $L^2(-1,1)$ on the domain $\mathcal{D}_{\min}$ (the subset of $\mathcal{D}_{\max}$ of functions with support in $(-1,1)$, see page 173 in Zettl (2005), we removed one condition on $\mathcal{D}_{\max}$ which is automatically satisfied). By Proposition 3, the multiplication defined, for $\psi \in \mathcal{D}_{\max}$, by $\psi \to q^c\psi$ is bounded and symmetric on $L^2(-1,1)$. Thus, by the Kato-Rellich theorem (see, *e.g.*, Reed and Simon (1975)), the unbounded operator $\mathcal{L}$ on domain $\mathcal{D}$ in case (2) is self-adjoint. Denote by $((U(c)/\pi)^2\chi_m^c)_{m\in\mathbb{N}_0}$ the eigenvalues of the unbounded operator $\mathcal{L}$ on domain $\mathcal{D}$ in case (2) arranged in increasing order and repeated according to multiplicity. They are real and, because the operator is bounded below, they are bounded below by the same constant. Moreover, Proposition 3 yields that $(\chi_m^c)_{m\in\mathbb{N}_0}$ are the eigenvalues of the unbounded operator $\mathcal{L}$ on domain $\mathcal{D}$ in case (1). The following result gives exact constants and a behavior uniform over m which is coherent with the asymptotic result on page 14 of Widom (1964).

Theorem 4. *We have, for all $m \in \mathbb{N}_0$ and $c > 0$,*

$$\left(\frac{\pi}{U(c)}\right)^2 \left(m(m+1) + \frac{1}{2} - R(c)\right) + c^2 \leq \chi_m^c \leq \left(\frac{\pi}{U(c)}\right)^2 \left(m(m+1) + \frac{1}{2}\right) - c^2.$$

Proof. This follows from the min-max theorem and (6.44). □

6.6 Uniform estimates on the singular functions g_m^c

Theorem 5. *We have, for all $m \in \mathbb{N}_0$ and $c > 0$,*

$$\|g_m^c\|_{L^\infty([-1,1])}$$
$$\leq \frac{\pi}{e^{2c}}\left(1+\frac{4c^2}{3}\right)^{1/2}\left(\frac{\sinh(4c)}{4c}\right)^{1/4}\cosh(2c)^{1/2}\left(\frac{2R(c)}{m+1/2}+\left(1+\sqrt{\frac{2}{3}}\frac{R(c)}{m+1/2}\right)\sqrt{m+\frac{1}{2}}\right).$$

The proof of this result uses similar ideas as in the proof of Proposition 5 in Bonami and Karoui (2016). The important additional ingredients are the change of variables and functions and Proposition 3.

Proof. Using in the first display the change of variables (6.30) and the change of functions (6.31) with $\psi = g_m^c$, and denoting by $\Gamma_m^c(\cdot) = F(\cdot) g_m^c\left(Y^{-1}(\cdot)\right)$ and $\widetilde{\Gamma}_m^c = \Gamma_m^c \sqrt{U(c)/\pi}$, which is real valued, and (6.45) and (6.31) in the second display, we obtain

$$\begin{aligned}\int_{-1}^{1}\left|\widetilde{\Gamma}_m^c(y)\right|^2 dy &= \frac{U(c)}{\pi}\int_{-1}^{1} Y'(x)\left|F(Y(x))\right|^2 \left|g_m^c(x)\right|^2 dx\\ &= \int_{-1}^{1}\frac{\cos(X(x))}{\sqrt{1-\sin(X(x))^2}}\left|g_m^c(x)\right|^2 dx = 1.\end{aligned}$$

Also, by Proposition 3, for all $y \in (-1,1)$,

$$\left((1-y^2)\left(\widetilde{\Gamma}_m^c\right)'\right)'(y)+m(m+1)\widetilde{\Gamma}_m^c(y) = \left(m(m+1)-\left(\frac{U(c)}{\pi}\right)^2\chi_m^c+q^c(y)\right)\widetilde{\Gamma}_m^c(y). \tag{6.49}$$

We obtain, by the method of variation of constants and knowledge of the solutions to the homogenous equation corresponding to the left-hand side of (6.49), that there exist $A, B \in \mathbb{R}$ such that, for $y \in (-1,1)$,

$$\widetilde{\Gamma}_m^c(y) = A\overline{P}_m(y) + BQ_m(y) + \frac{1}{m+1/2}\int_y^1 L_m(y,z)\sqrt{1-z^2}G_c(z)\widetilde{\Gamma}_m^c(z)dz, \quad (6.50)$$

where $\overline{P}_m$ is the Legendre polynomial of degree m and norm 1 in $L^2(-1,1)$, Q_m is the Legendre function of the second kind, $G_c(y) = m(m+1) - (U(c)/\pi)^2\chi_m^c + q^c(y)$, and $L_m(y,z) = \sqrt{1-z^2}\left(\overline{P}_m(y)Q_m(z) - \overline{P}_m(z)Q_m(y)\right)$. By Theorem 4 and Proposition 3, we have $\|G_c\|_{L^\infty([-1,1])} \le R(c)$. Because $\Gamma_m^c(1)$ is finite, $\overline{P}_m$ is bounded but $\lim_{y\to 1} Q_m(y) = \infty$, we know that $B = 0$. By the result after Lemma 9 in Bonami and Karoui (2016), for all $0 \le y \le z \le 1$, $|L_m(y,z)| \le 1$. Hence, by the Cauchy-Schwarz inequality, we have, for all $y \in (1,1)$,

$$\begin{aligned}\left|\widetilde{\Gamma}_m^c(y) - A\overline{P}_m(y)\right| &\le \frac{1}{m+1/2}\left(\int_y^1 (L_m(y,z))^2(1-z^2)dz\right)^{1/2}\left(\int_y^1 G_c(z)^2\widetilde{\Gamma}_m^c(z)^2dz\right)^{1/2},\\ &\le \frac{R(c)}{m+1/2}(1-y) \qquad (6.51)\end{aligned}$$

so

$$\int_{-1}^1 \left|\widetilde{\Gamma}_m^c(y) - A\overline{P}_m(y)\right|^2 dy \le \frac{2R(c)^2}{3(m+1/2)^2}$$

and, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\int_{-1}^1 \left|\widetilde{\Gamma}_m^c(y) - A\overline{P}_m(y)\right|^2 dy \ge& 1 + A^2 - 2|A|\int_{-1}^1 \left|\widetilde{\Gamma}_m^c(y)\right|^2 dy \int_{-1}^1 \left|\overline{P}_m(y)\right|^2 dy\\ \ge&(1-|A|)^2,\end{aligned}$$

hence

$$|A| \le 1 + \sqrt{\frac{2}{3}}\frac{R(c)}{m+1/2}. \qquad (6.52)$$

Also, by (6.43) and Lemma 2, we have

$$\|1/F\|_{L^\infty([-1,1])}\sqrt{\frac{\pi}{U(c)}} \le \pi e^{-2c}\left(1+\frac{4c^2}{3}\right)^{1/2}\left(\frac{\sinh(4c)}{4c}\right)^{1/4}\cosh(2c)^{1/2},$$

and we obtain the result by (6.51), (6.52), and $\left\|\overline{P}_m\right\|_{L^\infty([-1,1])} \le \sqrt{m+1/2}$. □

Corollary 2. *For all $m \in \mathbb{N}_0$ and $c > 0$,*

$$\|g_m^c\|_{L^\infty([-1,1])} \leq H(c)\sqrt{m+\frac{1}{2}}, \tag{6.53}$$

where

$$H(c) = \pi\sqrt{1+\frac{4c^2}{3}}\left(1+2\sqrt{2}\left(2+\frac{1}{\sqrt{3}}\right)\left(\frac{2}{\pi^2}+\frac{8}{3}(1+2c)\left(c^2+\frac{9c}{8}+\frac{1}{2}\right)\right)\right).$$

Proof. By the above results, (6.53) holds with

$$\pi e^{-2c}\left(1+\frac{4c^2}{3}\right)^{1/2}\left(\frac{\sinh(4c)}{4c}\right)^{1/4}\cosh(2c)^{1/2}\left(1+2\sqrt{2}R(c)\left(2+\frac{1}{\sqrt{3}}\right)\right)$$

in place of $H(c)$ and

$$R(c) < \frac{2}{\pi^2}+2\left(\frac{ce^{2c}}{\sinh(4c)}\right)^2\left(\left(\cosh(4c)\left(1+\frac{c}{3}\coth(2c)\right)-1\right)+2c\sinh(4c)\right).$$

hence, using that $e^c \geq 1+c$ which implies $c\coth(c) \leq c+2$,

$$R(c) < \frac{2}{\pi^2}+\frac{8ce^{4c}}{3\sinh(4c)}\left(c^2+\frac{9c}{8}+\frac{1}{2}\right) < \frac{2}{\pi^2}+\frac{8}{3}(1+2c)\left(c^2+\frac{9c}{8}+\frac{1}{2}\right).$$

We obtain the result, using

$$\begin{aligned} e^{-2c}\left(\frac{\sinh(4c)}{4c}\right)^{1/4}\cosh(2c)^{1/2} &= e^{-2c}\left(\frac{\sinh(2c)}{2c}\right)^{1/4}\cosh(2c)^{3/4} \\ &= \left(\frac{1-e^{-4c}}{4c}\right)^{1/4}\left(\frac{1+e^{-4c}}{2}\right)^{3/4} \leq 1. \end{aligned}$$

□

As a result we have, for a constant C_0,

$$\|g_m^c\|_{L^\infty([-1,1])} \leq C_0\,(c \vee 1)^4\sqrt{m+\frac{1}{2}}. \tag{6.54}$$

6.7 Numerical method to obtain the SVD of $\mathcal{F}_{b,c}$

In recent years, efficient numerical methods to obtain the SVD of the truncated Fourier transform acting on the space of bandlimited functions have been developped. This allows to go beyond the usual toolbox based on the Tikhonov or iterative methods such as the Landweber method (Gerchberg method for out-of-band extrapolation, see Bertero and Boccacci (1998)). The strategy that we implement in the next section is to first compute a numerical approximation of the right singular functions (the PSWF). We use that the first coefficients of the decomposition of the PSWF on the Legendre polynomials can be obtained by solving for the eigenvectors of two tridiagonal symmetric Toeplitz matrices (for even and odd values of m, see Section 2.6 in Osipov et al. (2013)). We can then compute their image by $\mathcal{F}_c^{W_{[-1,1]}*}$ (see (6.7) for the definition of $\mathcal{F}_c^{W_{[-1,1]}}$) because, by Gaillac and Gautier (2019c), $\mathcal{F}_c^{W_{[-1,1]}*} = \mathcal{R}\left[\mathbb{1}\{[-1,1]\}\mathcal{F}_c^{W_{[-1,1]}}\mathcal{E}\right]$ applied to the Legendre polynomials has a closed form involving the Bessel functions of the first kind (see (18.17.19) in Olver et al. (2010)).

For nonbandlimited functions, we propose to rely on the differential operator $\mathcal{L}$ in case (1) at the beginning of Section 6.5. We have used that because $\mathcal{Q}_c$ commutes with $\mathcal{L}$, $(g_m^c)_{m\in\mathbb{N}_0}$ are the eigenfunctions of $\mathcal{L}$. To obtain a numerical approximation of these functions, we use $\mathcal{L}$, whose eigenvalues are of the order of m^2 (see Theorem 4), rather than $\mathcal{Q}_c$, whose eigenvalues decay to zero exponentially. This is achieved by solving numerically for the eigenfunctions of a singular Sturm-Liouville operator. We approximate the values of the eigenfunctions on a grid on $[-1,1]$ using the MATLAB package MATSLISE 2 (it implements constant perturbation methods for limit point nonoscillatory singular problems, see Ledoux (2007) chapters 6 and 7 for the method and an analysis of the numerical approximation error). By Proposition A.1 in Gaillac and Gautier (2019c), we have $\varphi_m^{b,c}(\cdot) = \varphi_m^{1,c/b}(b\cdot)\sqrt{b}$ for all $m\in\mathbb{N}_0$. Finally, we use $\mathcal{F}_{1,c/b}^*\left[g_m^{c/b}\right] = \sigma_m^{1,c/b}\varphi_m^{1,c/b}$ and that $\varphi_m^{1,c/b}$ has norm 1 to obtain the remaining of the SVD. $\mathcal{F}_{1,c/b}^*\left[g_m^{c/b}\right]$ is computed using the fast Fourier transform.

6.8 Illustration: application to analytic continuation

We solve for f in (6.17) in Case (a) $f = 0.5/\cosh(2\cdot)$, which is not bandlimited, and Case (b) $f = \mathrm{sinc}(2\cdot)/6$ which is bandlimited, when $c = 0.5$, $x_0 = 0$, and $\xi = \cos(50\cdot)$. We use approximation f_δ^N described in Section 6.3.2 with $b = 1$ for Case (a), $b = 1/6.5$ for Case (b).

By analogy with the statistical problem where $\delta\xi$ is random rather than bounded, we use the terminology estimator. We select the value for the parameter $N = \widehat{N}$ based on a type of Goldenshluger-Lepski method (see Goldenshluger and Lepski (2014)):

$$\widehat{N} \in \underset{N' \in \{0,\dots,N_{\max}\}}{\operatorname{argmin}} \; B(N) + \Sigma(N),$$

$$B(N) = \sup_{N \le N' \le N_{\max}} \left(\left\| F_\delta^{N' \vee N} - F_\delta^N \right\|^2_{L^2(\cosh(b\cdot))} + \Sigma(N') \right)_+, \;\; \Sigma(N) = \frac{2\pi c \delta^2 e^{2\beta(c/b)N}}{1 - e^{-2\beta(c/b)}},$$

and $N_{\max} = \lfloor \log(1/\delta) \rfloor$. Performing analytic continuation using (6.19) requires the approximation of the scalar products on $[-1, 1]$ of the observed function f_δ with $g_m^{c/b}$. We use the package MATSLISE 2 to compute the value of the functions $\left(g_m^{c/b}\right)_{m=0}^{N_{\max}}$ at the n first Gauss-Legendre quadrature nodes. Results are presented in figures 6-2 and 6-3, where we use a 2^{12} resolution in the Fast Fourier transform, $n = 15000$, and precision of 10^{-10} for the computation of the eigenvalues in MATSLISE 2, which also controls the precision of the computation of the eigenfunctions in the function computeEigenfunction of MATSLISE 2 despite that this is not explicitely computed (see sections 7.2.3 and 5.2 in Ledoux (2007) for examples).

We compare f_δ^N to a similar estimator based on (6.19) but with the PSWF instead of g_m^c in Case (b). This approach can only be used to perform analytic continuation of bandlimited functions when the researcher knows an interval which contains the bandlimits. In contrast, even for bandlimited functions, using the estimator based on $g_m^{c/b}$ allows to perform analytic continuation without the knowledge of an interval containing the support of the Fourier transform of the function. Importantly, Figure 6-4 shows that f_δ^N performs almost as well as the unfeasible method using the PSWF.

For the sake of conciseness, this paper does not study the effect of the various discretizations which can be carried out with arbitrary precision. Rather, we used in the numerical illustration conservative choices for those. This paper also does not consider the statistical problem, prove minimax lower bounds for it, and the adaptivity of the data-driven rule giving $N = \widehat{N}$. This is the object of future work. The interested reader can refer to Gaillac and Gautier (2019*c*) for the full statistical analysis for estimation of the density of random coefficients in the linear random coefficients model.

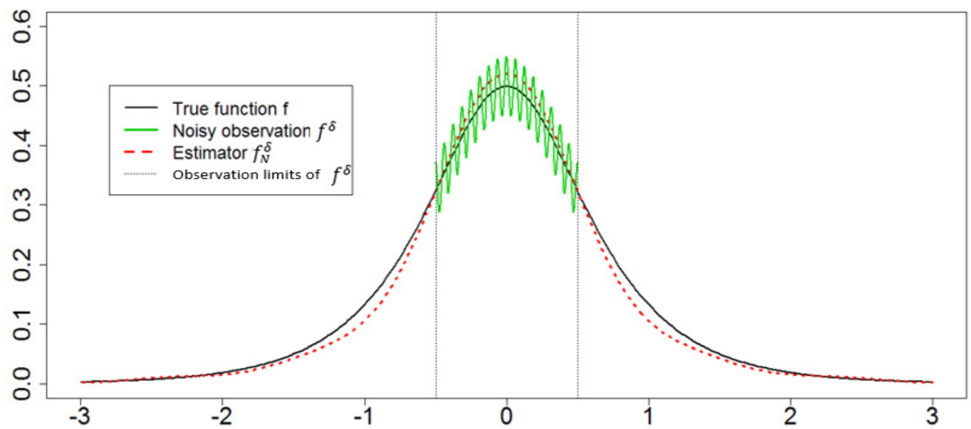


Figure 6-2: Case (a) with noise ($\delta = 0.05$), where F_{δ}^{N} in (6.19) uses $g_m^{c/b}$.

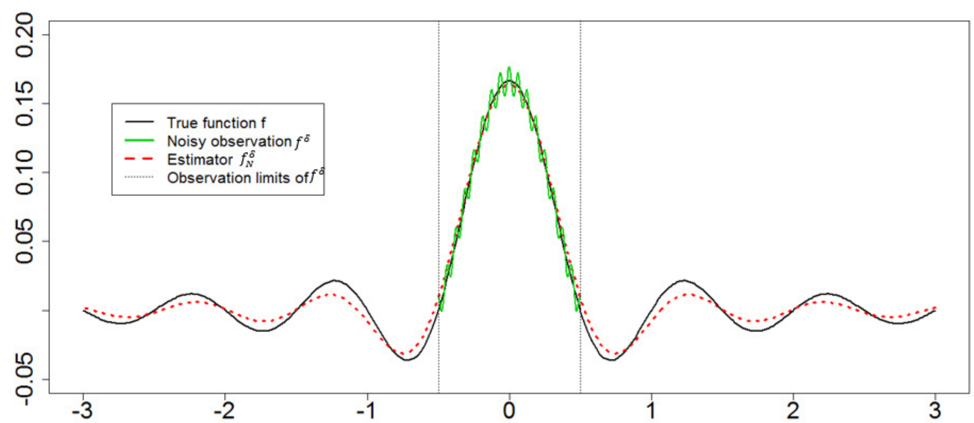


Figure 6-3: Case (b) with noise ($\delta = 0.01$), where F_{δ}^{N} in (6.19) uses $g_m^{c/b}$.

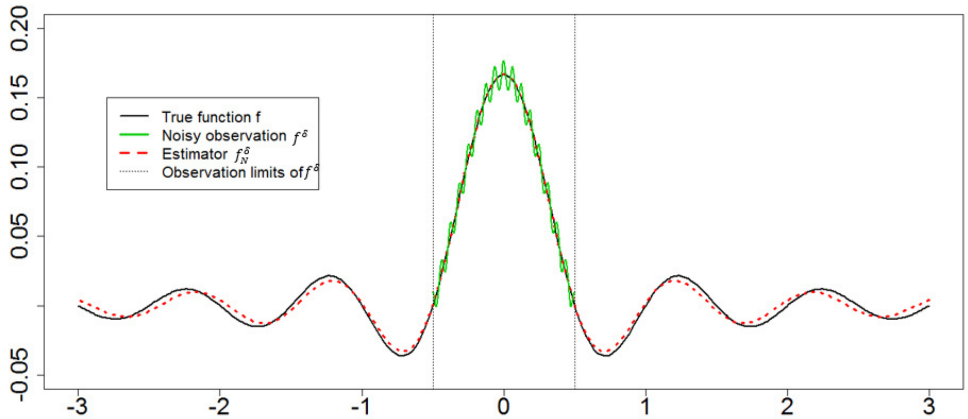


Figure 6-4: Case (b) with noise ($\delta = 0.01$), where F_δ^N in (6.19) uses the PSWF.

Bibliography

Abramowitz, M. and Stegun, I. (1965), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Dover Publications.

Aguirregabiria, V. and Mira, P. (2010), 'Dynamic discrete choice structural models: A survey', *Journal of Econometrics* **156**, 38–67.

Ajmi, N., Jouini, A. and Rieusset, P. G. L. (2011), 'Wavelet bases on a triangle', *Applied mathematics and computation* **217**(22), 9387–9396.

Alesina, A. and Rosenthal, H. (1995), *Partisan politics, divided government, and the economy*, Cambridge University Press.

Alibaud, N., Maréchal, P. and Saesor, Y. (2009), 'A variational approach to the inversion of truncated Fourier operators', *Inverse Probl.* **25**, 045002.

Alquier, P., Gautier, E. and Stoltz, G. (2011*a*), *Inverse problems and high-dimensional estimation: stats in the Château Summer School, August 31-September 4, 2009*, Vol. 203, Springer Science & Business Media.

Alquier, P., Gautier, E. and Stoltz, G. (2011*b*), *Inverse problems and high-dimensional estimation: Stats in the Château summer school, August 31-September 4, 2009*, Springer.

Anderson, G., Vamanamurthy, M. and Vuorinen, M. (1992), 'Functional inequalities for hypergeometric functions and complete elliptic integrals', *SIAM J. Math. Anal.* **23**, 512–524.

Andrews, D. (1994), 'Empirical process methods in econometrics', *Handbook of econometrics* **4**, 2247–2294.

Andrews, D. and Shi, X. (2013), 'Inference based on conditional moment inequalities', *Econometrica* **81**(2), 609–666.

Andrews, D. and Shi, X. (2017), 'Inference based on many conditional moment inequalities', *Journal of Econometrics* **196**(2), 275–287.

Andrews, I., Gentzkow, M. and Shapiro, J. M. (2017), 'Measuring the sensitivity of parameter estimates to estimation moments', *The Quarterly Journal of Economics* **132**(4), 1553–1592.

Arcidiacono, P., Aucejo, E., Maurel, A. and Ransom, T. (2016), College attrition and the dynamics of information revelation. NBER Working Paper No. 22325.

Arcidiacono, P., Hotz, J. and Kang, S. (2012), 'Modeling college major choices using elicited measures of expectations and counterfactuals', *Journal of Econometrics* **166**(1), 3–16.

Arcidiacono, P., Hotz, J. V., Maurel, A. and Romano, T. (2014), Recovering ex ante returns and preferences for occupations using subjective expectations data. NBER Working Paper No. 20626.

Armantier, O., Topa, G., Van der Klaauw, W. and Zafar, B. (2017), 'An overview of the survey of consumer expectations', *Economic Policy Review* **23**(2), 51–72.

Batenkov, D., Demanet, L. and Mhaskar, H. N. (2018), 'Stable soft extrapolation of entire functions', *Inverse Probl.* **35**(1), 015011.

Beare, B. and Shi, X. (2019), 'An improved bootstrap test of density ratio ordering', *Econometrics and Statistics* **10**, 9–26.

Belisle, C., Massé, J.-C. and Ransford, T. (1997), 'When is a probability measure determined by infinitely many projections?', *Annals of Probability* **25**, 767–786.

Belitser, E. (1998), 'Efficient estimation of analytic density under random censorship', *Bernoulli* **4**, 519–543.

Beran, R. (1995), 'Prediction in random coefficient regression', *J. Stat. Plan. Infer.* **43**, 205–213.

Beran, R., Feuerverger, A. and Hall, P. (1996*a*), 'On nonparametric estimation of intercept and slope distributions in random coefficient regression', *Annals of Statistics* **24**, 2569–2592.

Beran, R., Feuerverger, A. and Hall, P. (1996*b*), 'On nonparametric estimation of intercept and slope distributions in random coefficient regression', *Annals of Statistics* **24**, 2569–2592.

Beran, R. and Hall, P. (1992), 'Estimating coefficient distributions in random coefficient regressions', *Annals of Statistics* **20**, 1970–1984.

Beran, R. and Millar, W. (1994), 'Minimum distance estimation in random coefficient regression models', *Annals of Statistics* **22**, 1976–1992.

Berman, S. (2007), 'Legendre polynomial kernel estimation of a density function with censored observations and an application to clinical trials', *Comm. Pure Appl. Math.* **60**(8), 1238–1259.

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* pp. 841–890.

Berry, S. T. and Haile, P. A. (2009), Nonparametric identification of multinomial choice demand models with heterogeneous consumers, Technical report, National Bureau of Economic Research.

Bertanha, M. and Moreira, M. J. (2020), 'Impossible inference in econometrics: Theory and applications', *Journal of Econometrics* **218**, 247–270.

Bertero, M. and Boccacci, P. (1998), *Introduction to inverse problems in imaging*, CRC press.

Bertero, M., De Mol, C., Pike, E. and Walker, J. (1984), 'Resolution in diffraction-limited imaging, a singular value analysis: IV. The case of uncertain localization or non-uniform illumination of the object', *Opt. Acta* **31**, 923–946.

Bertero, M., De Mol, C. and Viano, G. A. (1979), 'On the problems of object restoration and image extrapolation in optics', *J. Math. Phys.* **20**(3), 509–521.

Birman, M. and Solomjak, M. Z. (2012), *Spectral theory of self-adjoint operators in Hilbert space*, Springer Science & Business Media.

Biroli, P., Boneva, T., Raja, A. and Rauh, C. (2020), 'Parental beliefs about returns to child health investments', *Journal of Econometrics* **Forthcoming**.

Bissantz, N., Hohage, T., Munk, A. and Ruymgaart, F. (2007), 'Convergence rates of general regularization methods for statistical inverse problems and applications', *SIAM J NUMER ANAL* **45**(6), 2610–2636.

Blundell, R. (2017), 'What have we learned from structural models?', *American Economic Review: Papers and Proceedings* **107**(5), 287–292.

Bochnak, J., Coste, M. and Roy, M.-F. (1998), *Real algebraic geometry*, Springer.

Bonami, A., Jaming, P. and Karoui, A. (2018), 'Non-asymptotic behaviour of the spectrum of the sinc kernel operator and related applications'. Preprint hal-01756828.

Bonami, A. and Karoui, A. (2014*a*), 'Spectral decay of the sinc kernel operator and approximations by Prolate Spheroidal Wave Functions'. Preprint arXiv:1012.3881.

Bonami, A. and Karoui, A. (2014*b*), 'Uniform bounds of prolate spheroidal wave functions and eigenvalues decay', *CR. Acad. Sci. I-Math.* **352**, 229–234.

Bonami, A. and Karoui, A. (2016), 'Uniform approximation and explicit estimates for the prolate spheroidal wave functions', *Constructive Approximation* **43**, 15–45.

Bonami, A. and Karoui, A. (2017), 'Approximations in sobolev spaces by prolate spheroidal wave functions', *Applied and Computational Harmonic Analysis* **42**(3), 361–377.

Boneva, T. and Rauh, C. (2018), 'Parental beliefs about returns to educational investments - the later the better?', *Journal of the European Economic Association* **16**(6), 1669–1711.

Bonhomme, S. and Weidner, M. (2019), 'Posterior average effects', *Preprint arXiv arXiv:1906.06360* .

Bound, J., Brown, C. and Mathiowetz, N. (2001), 'Measurement error in survey data', *Handbook of Econometrics* **5**, 3705–3843.

Box, G. E. and Tiao, G. C. (1973), *Bayesian inference in statistical analysis*, Vol. 40, John Wiley.

Breunig, C. (2019), 'Varying random coefficient models'. Preprint arXiv:1804.03110.

Breunig, C. (2020), 'Varying random coefficient models', *Journal of Econometrics* .

Breunig, C. and Hoderlein, S. (2018), 'Specification testing in random coefficient models', *Quant. Econ.* **9**(3), 1371–1417.

Brualdi, R. A. (2006), *Combinatorial matrix classes*, Vol. 13, Cambridge University Press.

Buchinsky, M., Li, F. and Liao, Z. (2019), 'Estimation and inference of semiparametric models using data from several sources', *Journal of Econometrics* **Forthcoming**.

Burden, B. C. and Kimball, D. C. (1998), 'A new approach to the study of ticket splitting', *American Political Science Review* **92**(3), 533–544.

Cai, T. T. (2002), 'On adaptive wavelet estimation of a derivative and other related linear inverse problems', *Journal of Statistical Planning and Inference* **108**(1-2), 329–349.

Cantoni, E. and Pons, V. (2018), 'Does context outweigh individual characteristics in driving voting behavior? evidence from relocations within the u.s', *Harvard Business School Working Paper* .

Carrasco, M. and Florens, J.-P. (2011), 'A spectral method for deconvolving a density', *Econometric Theory* **27**, 546–581.

Carrasco, M., Florens, J.-P. and Renault, E. (2007), 'Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization', *Handbook of Econometrics* **6**, 5633–5751.

Carroll, R. J. and Hall, P. (1988), 'Optimal rates of convergence for deconvolving a density', *Journal of the American Statistical Association* **83**(404), 1184–1186.

Cater, F. S. (1984), 'Differentiable, nowhere analytic functions', *American Mathematical Monthly* **91**, 618–624.

Cavalier, L. (2000), 'Efficient estimation of a density in a problem of tomography', *Annals of Statistics* **28**, 630–647.

Cavalier, L., Golubev, Y., Lepski, O. and Tsybakov, A. (2004), 'Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems', *Theory Probab. Appl.* **48**, 426–446.

Chagny, G. (2015), 'Adaptive warped kernel estimators', *Scandinavian Journal of Statistics* **42**, 336–360.

Chalendar, I. and Partington, J. R. (2007), 'Multivariable approximate Carleman-type theorems for complex measures', *Annals of Probability* **35**, 384–396.

Chen, W. (2010), 'Some aspects of band-limited extrapolations', *IEEE Trans. Signal Process.* **58**, 2647–2653.

Chen, X., Linton, O. and Van Keilegom, I. (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica* **71**(5), 1591–1608.

Chen, X. and Reiss, M. (2011), 'On rate optimality for ill-posed inverse problems in econometrics', *Econometric Theory* **27**, 497–521.

Chernozhukov, V., Fernández-Val, I. and Newey, W. K. (2019), 'Nonseparable multinomial choice models in cross-section and panel data', *Journal of Econometrics* **211**(1), 104–116.

Chung, S.-Y. (2002), 'Quasianalyticity of positive definite continuous functions', *Publications of the Research Institute for Mathematical Sciences* **38**, 725–733.

Cohen, A., Daubechies, I. and Vial, P. (1993), 'Wavelets on the interval and fast wavelet transforms', *Applied and Computational Harmonic Analysis* .

Coifman, R. R. and Lafon, S. (2006), 'Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions', *Appl. Comput. Harmon. Anal.* **21**(1), 31–52.

Compiani, G. (2019), Market counterfactuals and the specification of multi-product demand: a nonparametric approach, Technical report, Working paper.

Comte, F. and Genon-Catalot, V. (2018), 'Regression function estimation as a partly inverse problem', *Ann. Inst. Statist. Math.* pp. 1–32.

Comte, F., Genon-Catalot, V. and Samson, A. (2013), 'Nonparametric estimation for stochastic differential equations with random effects', *Stochastic Process. Appl.* **123**(7), 2522–2551.

Comte, F. and Lacour, C. (2011), 'Data-driven density estimation in the presence of additive noise with unknown distribution', *Journal of the Royal Statistical Society: Series B* **73**, 601–627.

Conlon, J. J., Philossoph, L., Wiswall, M. and Zafar, B. (2018), Labor market search with imperfect information and learning. NBER Working Paper No. 24988.

Cooprider, J., Hoderlein, S. and Meister, A. (2020), 'A panel data estimator for the distribution and quantiles of marginal effects in nonlinear structural models with an application to the demand for junk food', *Available at SSRN 3545485* .

Cross, P. J. and Manski, C. F. (2002*a*), 'Regressions, short and long', *Econometrica* **70**(1), 357–368.

Cross, P. J. and Manski, C. F. (2002*b*), 'Regressions, short and long', *Econometrica* **70**(1), 357–368.

Cuesta-Albertos, J. A., Fraiman, R. and Ransford, T. (2007), 'A sharp form of the Cramer-Wold theorem', *Journal of Theoretical Probability* **20**, 201–209.

Cunha, F. and Heckman, J. J. (2007), 'Identifying and estimating the distributions of ex post and ex ante returns to schooling', *Labour Economics* **14**(6), 870–93.

Cuturi, M. (2012), 'Positivity and transportation', *Preprint arXiv arXiv:1209.2655* .

Da Prato, G. and Zabczyk, J. (2014), *Stochastic equations in infinite dimensions, second ed.*, Cambridge.

Daubechies, I. (1992), *Ten lectures on wavelets*, Vol. 61, SIAM.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998), *Local properties of distributions of stochastic functionals*, American Mathematical Society.

De Jeu, M. (2003), 'Determinate multidimensional measures, the extended Carleman theorem and quasi-analytic weights', *Annals of Probability* **20**, 1205–1227.

De Jeu, M. (2004), 'Subspaces with equal closure', *Constructive Approximation* **20**, 93–157.

de Paula, A., Shapira, G. and Todd, P. E. (2014), 'How beliefs about hiv status affect risky behaviors: Evidence from malawi', *Journal of Applied Econometrics* **29**(6), 944–964.

Delavande, A. (2008), 'Pill, patch, or shot? subjective expectations and birth control choice', *International Economic Review* **49**(3), 999–1042.

DellaVigna, S. and Gentzkow, M. (2010), 'Persuasion: empirical evidence', *Annual Review of Economics* **2**(1), 643–669.

DellaVigna, S., List, J. A., Malmendier, U. and Rao, G. (2016), 'Voting to tell others', *The Review of Economic Studies* **84**(1), 143–181.

Dent, W. T. and Hildreth, C. (1977), 'Maximum likelihood estimation in random coefficient models', *Journal of the American Statistical Association* **72**(357), 69–72.

D'Haultfœuille, X., Gaillac, C. and Maurel, A. (2018*a*), Rationalexp: Tests of and deviations from rational expectations. Working paper.

D'Haultfœuille, X., Gaillac, C. and Maurel, A. (2018*b*), Rationalizing rational expectations? Tests and deviations. NBER Working Paper No. 25274.

D'Haultfoeuille, X., Gaillac, C. and Maurel, A. (2020), 'Rationalizing Rational Expectations? Characterization and Tests', *Preprint arxiv arXiv:2003.11537* .

Dion, C. (2014), 'New adaptive strategies for nonparametric estimation in linear mixed models', *Journal of Statistical Planning and Inference* **150**, 30–48.

Dominitz, J. (1998), 'Earnings expectations, revisions, and realizations', *Review of Economics and statistics* **80**(3), 374–388.

Donoho, D. L. and Johnstone, I. M. (1995), 'Adapting to unknown smoothness via wavelet shrinkage', *Journal of the American Statistical Association* **90**(432), 1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995), 'Wavelet shrinkage: asymptopia?', *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(2), 301–337.

Drouiche, K., Kateb, D. and Noiret, C. (2001), 'Regularization of the ill-posed problem of extrapolation with the malvar-wilson wavelets', *Inverse Probl.* **17**(5), 1513.

Dunker, F., Eckle, K., Proksch, K. and Schmidt-Hieber, J. (2017), 'Tests for qualitative features in the random coefficients model'. Preprint arXiv:1704.01066.

Evdokimov, K. and White, H. (2012), 'Some extensions of a lemma of Kotlarski', *Econometric Theory* **28**, 925–932.

Fan, Y., Sherman, R. and Shum, M. (2014), 'Identifying treatment effects under data combination', *Econometrica* **82**(2), 811–822.

Fan, Y., Sherman, R. and Shum, M. (2016), 'Estimation and inference in an ecological inference model', *Journal of Econometric Methods* **5**(1), 17–48.

Florens, J.-P., Heckman, J. J., Meghir, C. and Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica* **76**(5), 1191–1206.

Fox, J. (2017), 'A note on nonparametric identification of distributions of random coefficients in multinomial choice models', *NBER working paper series* .

Frikel, J. (2013), 'Sparse regularization in limited angle tomography', *Appl. Comput. Harmon. Anal.* **34**, 117–141.

Frogner, C. and Poggio, T. (2019), Fast and flexible inference of joint distributions from their marginals, *in* 'International Conference on Machine Learning', pp. 2002–2011.

Fu, C.-L., Dou, F.-F., Feng, X.-L. and Qian, Z. (2008), 'A simple regularization method for stable analytic continuation', *Inverse Probl.* **24**.

Fuchs, W. (1964), 'On the eigenvalues of an integral equation arising in the theory of band-limited signals', *J. Math. Anal. Appl.* **9**, 317–330.

Fuster, A., Kaplan, G. and Zafar, B. (2020), 'What would you do with $500? Spending responses to gains, losses, news and loans', *Review of Economic Studies* **Forthcoming**.

Gaillac, C. (2020), 'Inference in two data combination problems using entropic penalization of the support function', *Working paper* .

Gaillac, C. and Gautier, E. (2019*a*), 'Estimates for the SVD of the truncated fourier transform on $L^2(\exp(b|\cdot|))$ and stable analytic continuation'. Preprint arXiv:1905.11338v3.

Gaillac, C. and Gautier, E. (2019*b*), 'Identification in some random coefficients models when regressors have limited variation', *Working paper* .

Gaillac, C. and Gautier, E. (2019*c*), 'The linear random coefficients model when regressors have limited variation'. Preprint arXiv:1905.06584.

Gaillac, C. and Gautier, E. (2019*d*), 'RandomCoefficients: Adaptive estimation in the linear random coefficients model'. RandomCoefficients vignette.

Gautier, E. and Hoderlein, S. (2015), 'A triangular treatment effect model with random coefficients in the selection equation'. Preprint arXiv:1109.0362v4.

Gautier, E. and Kitamura, Y. (2013), 'Nonparametric estimation in random coefficients binary choice models', *Econometrica* **81**, 581–607.

Gautier, E. and Le Pennec, E. (2018), 'Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding', *Electronic Journal of Statistics* **12**, 277–320.

Gautschi, W. (1959), 'Some elementary inequalities relating to the Gamma and incomplete Gamma function', *J. Math. Phys.* **38**, 77–81.

Gennaioli, N., Ma, Y. and Shleifer, A. (2015), Expectations and investment, *in* 'NBER Macroeconomics Annual 2015, Volume 30', University of Chicago Press.

Gennaioli, N., Ma, Y. and Shleifer, A. (2016), 'Expectations and investment', *NBER Macroeconomics Annual* **30**(1), 379–431.

Gerber, A. S. and Green, D. P. (2000), 'The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment', *American Political Science Review* pp. 653–663.

Gerchberg, R. (1974), 'Super-resolution through error energy reduction', *J. Mod. Opt.* **21**, 709–720.

Giné, E. and Nickl, R. (2016), *Mathematical foundations of infinite-dimensional statistical models*, Vol. 40, Cambridge University Press.

Goldenshluger, A. and Lepski, O. (2008), 'Universal pointwise selection rule in multivariate function estimation', *Bernoulli* **14**(4), 1150–1190.

Goldenshluger, A. and Lepski, O. (2014), 'On adaptive minimax density estimation on $\mathbb{R}^d$', *Probability Theory Related Fields* **159**, 479–543.

Goodman, L. A. (1959), 'Some alternatives to ecological correlation', *American Journal of Sociology* **64**(6), 610–625.

Gosse, L. (2010*a*), 'Analysis and short-time extrapolation of stock market indexes through projection onto discrete wavelet subspaces', *Nonlinear Anal. Real World Appl.* **11**(4), 3139–3154.

Gosse, L. (2010*b*), 'Effective band-limited extrapolation relying on Slepian series and l1 regularization', *Compt. Math. Appl.* **60**, 1259–1279.

Gourieroux, C. and Pradel, J. (1986), 'Direct test of the rational expectation hypothesis', *European Economic Review* **30**(2), 265–284.

Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y. and Tetali, P. (2018), 'Characterization of a class of weak transport-entropy inequalities on the line', *Annales de l'IHP* **54**(3), 1667–1693.

Grabovsky, Y. and Hovsepyan, N. (2020), 'Explicit power laws in analytic continuation problems via reproducing kernel hilbert spaces', *Inverse Problems* **36**(3).

Green, D. P. and Gerber, A. S. (2019), *Get out the vote: How to increase voter turnout*, Brookings Institution Press.

Greiner, D. J. and Quinn, K. M. (2010), 'Exit polling and racial bloc voting: Combining individual-level and r x c ecological data', *The Annals of Applied Statistics* pp. 1774–1796.

Griffiths, W. E., Drynan, R. G. and Prakash, S. (1979), 'Bayesian estimation of a random coefficient model', *Journal of Econometrics* **10**(2), 201–220.

Gutknecht, D., Hoderlein, S. and Peters, M. (2018), 'Constrained information processing and individual income expectations'. Working paper.

Hahn, M. G. and Quinto, E. T. (1985), 'Distances between measures from 1-dimensional projections as implied by continuity of the inverse radon transform', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **70**(3), 361–380.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (2012), *Wavelets, approximation, and statistical applications*, Vol. 129, Springer Science & Business Media.

Hartman, P. (1987), *Ordinary differential equations*, SIAM.

Heckman, J. and Singer, B. (1984), 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica* **52**, 271–320.

Hirschman, I. I. (1950), 'On the distribution of the zeros of functions belonging to certain quasi-analytic classes', *American Journal of Mathematics* **72**, 396–406.

Hoderlein, S., Holzmann, H. and Meister, A. (2017), 'The triangular model with random coefficients', *Journal of Econometrics* **201**, 144–169.

Hoderlein, S., Klemelä, J. and Mammen, E. (2010), 'Analyzing the random coefficient model nonparametrically', *Econometric Theory* **26**, 804–837.

Hoffman, M. and Burks, S. V. (2020), 'Worker overconfidence: Field evidence and implications for employee turnover and returns from training', *Quantitative Economics* **11**, 315–348.

Hohmann, D. and Holzmann, H. (2015), 'Weighted angle radon transform: Convergence rates and efficient estimation', *Stat. Sinica* .

Holzmann, H. and Meister, A. (2019), 'Rate-optimal nonparametric estimation for random coefficient regression models'. Preprint arXiv:1902.05261.

Hoorfar, A. and Hassani, M. (2007), 'Approximation of the Lambert W function and hyperpower function', *Research report* **10**.

Hsu, Y.-C. (2016), 'Consistent tests for conditional treatment effects', *The Econometrics Journal* **20**(1), 1–22.

Ichimura, H. and Thompson, T. S. (1998), 'Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution', *Journal of Econometrics* **86**, 269–295.

Imai, K. (2005), 'Do get-out-the-vote calls reduce turnout? the importance of statistical methods for field experiments', *American Political Science Review* pp. 283–300.

Imai, K. and Khanna, K. (2016), 'Improving ecological inference by predicting individual ethnicity from voter registration records', *Political Analysis* **24**(2), 263–272.

Imai, K., King, G. and Velasco Rivera, C. (2020), 'Do nonpartisan programmatic policies have partisan electoral effects? evidence from two large-scale experiments', *The Journal of Politics* **82**(2), 714–730.

Imai, K., Lu, Y. and Strauss, A. (2008), 'Bayesian and likelihood inference for 2× 2 ecological tables: an incomplete-data approach', *Political Analysis* **16**(1), 41–69.

Infusino, M. (2016), Quasi-analyticity and determinacy of the full moment problem from finite to infinite dimensions, *in* 'Stochastic and Infinite Dimensional Analysis', Springer, pp. 161–194.

Ivaldi, M. (1992), 'Survey evidence on the rationality of expectations', *Journal of Applied Econometrics* **7**(3), 225–241.

Jähnisch, M. and Nussbaum, M. (2003), 'Asymptotic equivalence for a model of independent non identically distributed', *Statistics & Decisions* **21**, 197–218.

Jiang, W., King, G., Schmaltz, A. and Tanner, M. A. (2020), 'Ecological regression with partial identification', *Political Analysis* **28**(1), 65–86.

Karoui, A. and Moumni, T. (2008), 'New efficient methods of computing the prolate spheroidal wave functions and their corresponding eigenvalues', *Appl. Comput. Harmon. Anal.* **24**, 269–289.

Katz, J. N. and King, G. (1999), 'A statistical model for multiparty electoral data', *American Political Science Review* **93**(1), 15–32.

Kerkyacharian, G. and Picard, D. (2004), 'Regression in random design and warped wavelets', *Bernoulli* **10**, 1053–1105.

King, G. (1997), *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*, Princeton University Press.

Kleiber, C. and Stoyanov, J. (2013), 'Multivariate distributions and the moment problem', *Journal of Multivariate Analysis* **113**, 7–18.

Knight, B. G. (2014), An econometric evaluation of competing explanations for the midterm gap, Technical report, National Bureau of Economic Research.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* pp. 33–50.

Kotlarski, I. (1967), 'On characterizing the Gamma and the normal distribution', *Pacific Journal of Mathematics* **20**, 69–76.

Kress, R. (1999), *Linear integral equations*, Springer.

Kuchler, T. and Zafar, B. (2019), 'Personal experiences and expectations about aggregate outcomes', *Journal of Finance* **74**, 2491–2542.

Lacour, C. (2006), 'Rates of convergence for nonparametric deconvolution', *CR. Acad. Sci. I-Math.* **342**, 877–882.

Lacour, C. and Massart, P. (2016), 'Minimal penalty for Goldenshluger–Lepski method', *Stochastic Processes and their Applications* **126**(12), 3774–3789.

Landau, H. (1986), 'Extrapolating a band-limited function from its samples taken in a finite interval', *IEEE Trans. Inform. Theory* **32**(4), 464–470.

Landau, H. and Pollak, H. (1962), 'Prolate Spheroidal Wave Functions, Fourier analysis and uncertainty?III: The dimension of the space of essentially time-and band-limited signals', *Bell System Technical Journal* **41**, 1295–1336.

Lazarsfeld, P. F., Berelson, B. and Gaudet, H. (1944), *The people's choice.*, Duell, Sloan & Pearce.

Le Pennec, C. and Pons, V. (2019), 'How do campaigns shape vote choice? multi-country evidence from 62 elections and 56 tv debates', *NBER Working Paper* **26572**.

Ledoux, V. (2007), Study of special algorithms for solving Sturm-Liouville and Schrödinger equations, PhD thesis, Ghent University.

Li, Q. and Racine, J. S. (2008), 'Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data', *Journal of Business & Economic Statistics* **26**(4), 423–434.

Lin, G. D. (1997), 'On the moment problems', *Statistics and Probability Letters* **35**, 85–90.

Lindberg, J. (2012), 'Mathematical concepts of optical superresolution', *J. Opt.* **14**, 083001.

Linton, O., Song, K. and Whang, Y.-J. (2010), 'An improved bootstrap test of stochastic dominance', *Journal of Econometrics* **154**, 186–202.

Lovell, M. C. (1986), 'Tests of the rational expectations hypothesis', *American Economic Review* **76**(1), 110–124.

Mair, B. A. and Ruymgaart, F. H. (1996), 'Statistical inverse estimation in hilbert scales', *SIAM J. Appl. Math.* **56**(5), 1424–1444.

Mallat, S. (1999), *A wavelet tour of signal processing*, Elsevier.

Mandelbrojt, S. (1952), *Séries adhérentes, régularisation des suites, applications*, Gauthiers-Villars.

Manski, C. (2004), 'Measuring expectations', *Econometrica* **72**(5), 1329–1376.

Manski, C. F. (2018), 'Credible ecological inference for medical decisions with personalized risk assessment', *Quantitative Economics* **9**(2), 541–569.

Manski, C. and Molinari, F. (2010), 'Rounding probabilistic expectations in surveys', *Journal of Business & Economic Statistics* **28**(2), 219–231.

Masten, M. A. (2017), 'Random coefficients on endogenous variables in simultaneous equations models', *The Review of Economic Studies* **85**(2), 1193–1250.

Masten, M. A. and Torgovitsky, A. (2013), 'Instrumental variables estimation of a generalized correlated random coefficients model', *Preprint arXiv* *arXiv:1310.6643* .

Matzkin, R. (2007*a*), 'Nonparametric identification', *Handbook of Econometrics* **6**, 5307–5368.

Matzkin, R. L. (2007*b*), 'Nonparametric identification', *Handbook of econometrics* **6**, 5307–5368.

McCall, J. J. (1970), 'Economics of information and job search', *The Quarterly Journal of Economics* pp. 113–126.

McCarty, N., Poole, K. T. and Rosenthal, H. (2009), 'Does gerrymandering cause polarization?', *American Journal of Political Science* **53**(3), 666–680.

Mebane, W. R. (2000), 'Coordination, moderation, and institutional balancing in american presidential and house elections', *American Political Science Review* **94**(1), 37–57.

Meghir, C. and Pistaferri, L. (2011), Earnings, consumption and life cycle choices, *in* 'Handbook of labor economics', Vol. 4, Elsevier, pp. 773–854.

Meister, A. (2007), 'Deconvolving compactly supported densities', *Mathematical Methods of Statistics* **16**, 63–76.

Meister, A. (2011), 'Asymptotic equivalence of functional linear regression and a white noise inverse problem', *The Annals of Statistics* **39**(3), 1471–1495.

Miller, K. (1970), 'Least squares methods for ill-posed problems with a prescribed bound', *SIAM J. Math. Anal.* **1**(1), 52–74.

Molinari, F. and Peski, M. (2006*a*), 'Generalization of a result on Regressions, short and long', *Econometric Theory* **22**(1), 159–163.

Molinari, F. and Peski, M. (2006*b*), 'Generalization of a result on "regressions, short and long"', *Econometric Theory* **22**(1), 159–163.

Morimoto, M. (1962), *Analytic functionals on the sphere*, American Mathematical Society.

Morrison, J. (1962), 'On the commutation of finite integral operators, with difference kernels, and linear self-adjoint differential operators', *Notices Amer. Math. Soc.* **9**, 119.

Mortici, C. and Chen, C.-P. (2011), 'New sharp double inequalities for bounding the gamma and digamma function', *Analele Universitatii de Vest din Timisoara* **49**.

Mulansky, B. and Neamtu, M. (1998), 'Interpolation and approximation from convex sets', *Journal of approximation theory* **92**(1), 82–100.

Muth, J. F. (1961), 'Rational expectations and the theory of price movements', *Econometrica* **29**(3), 315–335.

Muzellec, B., Nock, R., Patrini, G. and Nielsen, F. (2017), Tsallis regularized optimal transport and ecological inference, *in* 'Thirty-First AAAI Conference on Artificial Intelligence'.

Natterer, F. (1986), *The mathematics of computerized tomography*, Vol. 32, SIAM.

Nazarov, F. L. (2000), Complete version of Turán's lemma for trigonometric polynomials on the unit circumference, *in* 'Complex Analysis, Operators, and Related Topics', Springer, pp. 239–246.

Newey, W. K. and Stouli, S. (2018), 'Heterogenous coefficients, discrete instruments, and identification of treatment effects', *Preprint arXiv arXiv:1811.09837* .

Newey, W. and Stouli, S. (2020), 'Control variables, discrete instruments, and identification of structural functions', *Journal of Econometrics* .

Niessen, H.-D. and Zettl, A. (1992), 'Singular Sturm-Liouville problems: the Friedrichs extension and comparison of eigenvalues', *P. Lond. Math. Soc.* **3**, 545–578.

Olver, F., Lozier, D., Boisvert, R. and Clark, C. (2010), *NIST handbook of mathematical functions*, Cambridge University Press.

Osipov, A. (2013), 'Certain inequalities involving prolate spheroidal wave functions and associated quantities', *Appl. Comput. Harmon. Anal.* **35**, 359–393.

Osipov, A., Rokhlin, V. and Xiao, H. (2013), *Prolate spheroidal wave functions of order zero*, Springer.

Pakes, A. G., Hung, W. and Wu, J. (2001), 'Criteria for the unique determination of probability distributions by moments', *Australian and New Zealand Journal of Statistics* **43**, 101–111.

Patton, A. J. and Timmermann, A. (2012), 'Forecast rationality tests based on multi-horizon bounds', *Journal of Business & Economic Statistics* **30**(1), 1–17.

Pesaran, M. H. (1987), *The limits to rational expectations*, Basil Blackwell.

Petersen, L. C. (1982), 'On the relation between the multidimensional moment problem and the one-dimensional moment problem', *Mathematica Scandinavia* **51**, 361–366.

Pollard, D. (1990), Empirical processes: theory and applications, *in* 'NSF-CBMS regional conference series in probability and statistics', Institute of Mathematical Statistics and the American Statistical Association, pp. i–86.

Pons, V. (2018), 'Will a five-minute discussion change your mind? a countrywide experiment on voter choice in france', *American Economic Review* **108**(6), 1322–63.

Pons, V. and Liegey, G. (2019), 'Increasing the electoral participation of immigrants: Experimental evidence from france', *The Economic Journal* **129**(617), 481–508.

Prior, M. (2006), 'The incumbent in the living room: The rise of television and the incumbency advantage in us house elections', *The journal of Politics* **68**(3), 657–673.

Reed, M. and Simon, B. (1975), *Methods of Modern Mathematical Physics, II: Fourier Analysis, Self-adointness*, Academic Press.

Reed, M. and Simon, B. (1980), *Methods of Modern Mathematical Physics, I: Functional analysis*, Academic Press.

Reiß, M. (2008), 'Asymptotic equivalence for nonparametric regression with multivariate and random design', *Ann. Stat.* **36**(4), 1957–1982.

Ridder, G. and Moffitt, R. (2007), 'The econometrics of data combination', *Handbook of Econometrics* **6**, 5469–5547.

Robinson, W. S. (1950), 'Ecological correlations and the behavior of individuals', *American Sociological Review* **15**(3), 351–357.

Rokhlin, V. and Xiao, H. (2007), 'Approximate formulae for certain prolate spheroidal wave functions valid for large values of both order and band-limit', *Appl. Comput. Harmon. Anal.* **22**, 105–123.

Rosen, O., Jiang, W., King, G. and Tanner, M. A. (2001), 'Bayesian and frequentist inference for ecological inference: The R×C case', *Statistica Neerlandica* **55**(2), 134–156.

Rothschild, M. and Stiglitz, J. (1970), 'Increasing risk: I. a definition', *Journal of Economic Theory* **2**(3), 225–243.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies.', *Journal of educational Psychology* **66**(5), 688.

Rudin, W. (1973), *Real and complex analysis*, McGraw-Hill.

Rudin, W. (1980), *Function theory in the unit ball of* $\mathbb{C}^n$, Springer.

Rullgård, H. and Quinto, E. T. (2010), 'Local sobolev estimates of a function by means of its radon transform', *Inverse Problems and Imaging* **4**(4), 721–734.

Salzmann, H. and Zeller, K. (1955), 'Singularitäten unendlich oft differenzierbarer funktionen', *Mathematische Zeitschrift* **62**, 354–367.

Schmets, J. and Valdivia, M. (1991), 'On the extent of (non) quasi-analytic classes', *Archiv der Mathematik* **56**, 593–600.

Shabat, B. V. (1992), *Introduction to complex analysis: functions of several variables, Part 2*, AMS.

Shapiro, H. S. (1986), 'Reconstructing a function from its values on a subset of its domain? A Hilbert space approach', *J. Approx. Theory* **46**(4), 385–402.

Slepian, D. (1965), 'Some asymptotic expansions for prolate spheroidal wave functions', *J. Math. Phys.* **44**, 99–140.

Slepian, D. (1983), 'Some comments on Fourier analysis, uncertainty and modeling', *SIAM Rev.* **25**, 379–393.

Stinebrickner, R. and Stinebrickner, T. (2012), 'Learning about academic ability and the college dropout decision', *Journal of Labor Economics* **30**(4), 707–748.

Stinebrickner, R. and Stinebrickner, T. (2014*a*), 'Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model', *Journal of Labor Economics* **32**(3), 601–644.

Stinebrickner, R. and Stinebrickner, T. (2014*b*), 'A major in science? initial beliefs and final outcomes for college major and dropout', *The Review of Economic Studies* **81**(1), 426–472.

Stoyanov, J. (2000), 'Krein condition in probabilistic moment problems', *Bernoulli* **6**, 939–949.

Strassen, V. (1965), 'The existence of probability measures with given marginals', *The Annals of Mathematical Statistics* **36**(2), 423–439.

Tam Cho, W. K. (1998), 'Iff the assumption fits?: A comment on the king ecological inference solution', *Political Analysis* **7**, 143–163.

Tauvel, P. (2006), *Analyse complexe pour la licence 3*, Dunod.

Train, K. (2009), *Discrete choice methods with simulation*, Cambridge.

Trefethen, L. N. (2019), 'Quantifying the ill-conditioning of analytic continuation', *SIAM J. Numer. Anal* .

Tsybakov, A. (2000), 'On the best rate of adaptive estimation in some inverse problems', *Comptes Rendus de l'Academie des Sciences - Mathematics* **330**, 835–840.

Tsybakov, A. (2008), *Introduction to nonparametric estimation*, Springer.

Van den Berg, G. J. (2001), Duration models: specification, identification and multiple durations, *in* 'Handbook of econometrics', Vol. 5, Elsevier, pp. 3381–3460.

Van der Klaauw, W. (2012), 'On the use of expectations data in estimating structural dynamic choice models', *Journal of Labor Economics* **30**(3), 521–554.

Van der Klaauw, W. and Wolpin, K. I. (2008), 'Social security and the retirement and savings behavior of low-income households', *Journal of Econometrics* **145**(1-2), 21–42.

Van der Vaart, A. (2000), *Asymptotic statistics*, Cambridge University Press.

Van der Vaart, A. and Wellner, J. (1996), Weak convergence, *in* 'Weak Convergence and Empirical Processes', Springer, pp. 16–28.

Wakefield, J. (2004), 'Ecological inference for 2× 2 tables', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **167**(3), 385–425.

Waldschmidt, M. (2000), *Diophantine approximation on linear algebraic groups*, Springer.

Wand, M. P. and Jones, M. C. (1994), 'Multivariate plug-in bandwidth selection', *Computational Statistics* **9**(2), 97–116.

Widom, H. (1964), 'Asymptotic behavior of the eigenvalues of certain integral equations. II', *Arch. Ration. Mech. Anal.* **17**, 215–229.

Wiswall, M. and Zafar, B. (2015), 'Determinants of college major choice: Identification using an information experiment', *The Review of Economic Studies* **82**(2), 791–824.

Yang, Z.-H., Jiang, Y.-L., Song, Y.-Q. and Chu, Y.-M. (2014), Sharp inequalities for trigonometric functions, *in* 'Abstract and Applied Analysis', Vol. 2014, Hindawi.

Young, R. (2001), *An introduction to nonharmonic Fourier series*, Academic press.

Zafar, B. (2011*a*), 'Can subjective expectations data be used in choice models? evidence on cognitive biases', *Journal of Applied Econometrics* **26**(3), 520–544.

Zafar, B. (2011*b*), 'How do college students form expectations?', *Journal of Labor Economics* **29**(2), 301–348.

Zaller, J. R. (1992), *The nature and origins of mass opinion*, Cambridge university press.

Zettl, A. (2005), *Sturm-Liouville theory*, American Mathematical Soc.

Zhang, Y.-X., Fu, C.-L. and Yan, L. (2011), 'Approximate inverse method for stable analytic continuation in a strip domain', *J. Comput. Appl. Math.* **235**, 2979–2992.