THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Max LESELLIER

Le 22 juin 2023

Articles en microéconométrie structurelle

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Christian BONTEMPS

Jury
M. Jean-François HOUDE, Rapporteur
M. Quang VUONG, Rapporteur
Mme Cristina GUALDANI, Examinatrice
M. Mathias REYNAERT, Examinateur
M. Steven BERRY, Examinateur
M. Christian BONTEMPS, Directeur de thèse

PhD Thesis in Economics

Essays in structural microeconometrics

Max Lesellier

June 2023

Acknowledgments

First, I want to express my deepest gratitude to my advisor Christian Bontemps. I first met Christian at the end of my master year when I was still trying to figure out the research topics I wanted to explore. Christian introduced me to several different interesting research questions, which became the basis for my thesis. I am truly appreciative of his scientific guidance over the last five years and, above all, I want to thank him for always being extremely supportive and incredibly generous with his time. Christian has also been a great co-author and I am looking forward to working with him in the future. I am similarly indebted to my co-advisor Nour Meddahi. I greatly benefited from his broad and deep understanding of econometrics and his guidance was invaluable, especially in helping me structure my research ideas and writing my job market paper. As the Director of the PhD program, Nour undertook many efforts to improve the quality of the training and increase the resources available to the graduate students, and I want to thank him for that. I am also grateful to Mathias Reynaert who helped me better communicate my ideas and allowed me to appreciate the empirical implications of my job market paper. Finally, I want to thank Cristina Gualdani who gave me valuable advice at different stages of my PhD.

There are several other professors from TSE that I want to thank, particularly Michel Le Breton, Eric Gautier, Isis Durmeyer, Pierre Dubois, Vishal Kamat, Pascal Lavergne, Thierry Magnac, Olivier Degroote and François Poinas. I would like to thank Steven Berry for hosting me at Yale University and allowing me to spend one semester in a highly stimulating research environment.

I would like to give special thanks to my two PhD coauthors Hippolyte Boucher and Gökçe Gökkoca. I sincerely enjoyed our collaboration and I hope we will keep working together in the future. I also want to thank some of my fellow PhD students and friends most notably Amirreza Ahmadzadeh, Alae Baha, Léa Bignon, Lisa Botbol, Tim Ederer, Anaïs Fabre, Gyung Mo Kim, Young Kim, Moritz Loewenfeld, Paul-Henri Moisson, Jose Alfonso Munoz, Peter Neis, Tuuli Vanhapelto, Philipp Wangner, Stephan Waizmann. You made my journey through the PhD much more enjoyable.

I want to thank my non-economist friends Audrey, Florine, Marine, Raphaël, Robin, Sylvain, and Titouan for their friendship and their support over the years. I was fortunate to be able to share many great moments with all of you.

Finally, I want to thank my family for their unconditional love and support. In particular, I would like to dedicate my thesis to my parents Judith and Philippe, and my grandmother Madeleine, who have always been there for me.

Overview

In this thesis, I develop new econometric methods to test and relax statistical or equilibrium restrictions that are commonly assumed in popular industrial organization models including the random coefficient logit model, entry games, and optimal contracts. I then apply these methods to investigate how the usual assumptions affect the results obtained in several relevant empirical examples. This thesis is organized into three chapters.

The first chapter of my thesis is entitled "Testing and Relaxing Distributional Assumptions on Random Coefficients in Demand Models”. This chapter is co-authored with two fellow graduate students Hippolyte Boucher and Gökçe Gökkoca. We provide a method to test and relax the distributional assumptions on random coefficients in the differentiated products demand model initiated by Berry (1994) and Berry, Levinsohn and Pakes (1995). This model is the workhorse model for demand estimation with market-level data and it uses random coefficients to account for unobserved preference heterogeneity. In this chapter, we provide a formal moment-based specification test on the distribution of random coefficients, which allows researchers to test the chosen specification (for instance normality) without re-estimating the model under a more flexible parametrization. The moment conditions (or equivalently the instruments) chosen for the test are designed to maximize the power of the test when the distribution of Random Coefficients is misspecified. By exploiting the duality between estimation and testing, we show that these instruments can also improve the estimation of the BLP model under a flexible parametrization (here, we consider the case of the Gaussian mixture). Finally, we validate our approach with Monte Carlo simulations and an empirical application using data on car purchases in Germany.

The second chapter is entitled: "Moment Inequalities for Entry Games with Heterogeneous Types". This chapter is coauthored with my advisor Christian Bontemps and Rohit Kumar. We develop new methods to simplify the estimation of entry games when the equilibrium selection mechanism is unrestricted. In particular, we develop an algorithm that allows us to recursively select a relevant subset of inequalities that sharply characterize the set of admissible parameters. Then, we propose a way to circumvent the problem of deriving an easy-to-compute and competitive critical value by smoothing the minimum function. In our case, it allows us to obtain a pivotal test statistic that eliminates “numerically” the non-binding moments. We show that we recover a consistent confidence region by letting the smoothing parameter increase with the sample size. Interestingly, we show that our procedure can easily be adapted to the case with covariates including continuous ones. Finally, we conduct full-scale Monte Carlo simulations to assess the performance of our new estimation procedure.

The third chapter is entitled "Identification and Estimation of Incentive Contracts under Asymmetric Information: an application to the French Water Sector". This chapter has its roots in a project Christian Bontemps and David Martimort started many years ago. We develop a Principal-Agent model to represent management contracting for public-service delivery. A firm (the Agent) has private knowledge of its marginal cost of production. The local public authority (the Principal) cares about the consumers' net surplus from consuming the services and the (weighted) firm's profit. Contractual negotiation is modeled as the choice by the privately informed firm within a menu of options determining both the unit-price charged to consumers and the fixed fee. Our theoretical model characterizes optimal contracting in this environment. We then explicitly study the nonparametric identification of the model and perform a semi-parametric estimation on a dataset coming from the 2004 wave of a survey from the French Environment Institute.

Résumé

Dans cette thèse, je développe de nouvelles méthodes économétriques pour tester et relaxer les restrictions statistiques ou d'équilibre couramment supposées dans des modèles populaires d'organisation industrielle, tels que le modèle logit à coefficients aléatoires, les jeux d'entrée et les contrats optimaux. J'applique ensuite ces méthodes pour étudier comment les hypothèses habituelles affectent les résultats obtenus dans plusieurs exemples empiriques pertinents. Cette thèse contient trois chapitres.

Le premier chapitre de ma thèse s'intitule "Tester et relaxer les hypothèses de distribution sur les coefficients aléatoires dans le modèle de demande". Ce chapitre est co-écrit avec deux autres doctorants, Hippolyte Boucher et Gökçe Gökkoca. Nous proposons une méthode pour tester et relaxer les hypothèses de distribution sur les coefficients aléatoires dans le modèle de demande de produits différenciés initié par Berry (1994) et Berry, Levinsohn et Pakes (1995). Il s'agit du modèle de référence pour l'estimation des fonctions de demande avec des données agrégées de marché. Les coefficients aléatoires modélisent l'hétérogénéité non observée des préférences. Dans ce chapitre, nous proposons un test de spécification sur la distribution des coefficients aléatoires, qui permet aux chercheurs de tester la spécification choisie (par exemple la normalité) sans ré-estimer le modèle sous une paramétrisation plus flexible. Les moments sont choisis pour maximiser la puissance du test lorsque la distribution des coefficients aléatoires est mal spécifiée. En exploitant la dualité entre l'estimation et le test, nous montrons que ces instruments peuvent également améliorer l'estimation du modèle BLP sous une paramétrisation plus flexible (nous étudions le cas du mélange de normales). Enfin, nous validons notre approche avec des simulations de Monte Carlo et une application empirique sur le marché des voitures en Allemagne.

Le deuxième chapitre s'intitule "Inégalités de Moment pour les Jeux d'Entrée avec Types Hétérogènes". Ce chapitre a été co-écrit avec Christian Bontemps et Rohit Kumar. Nous développons de nouvelles méthodes pour simplifier l'estimation des jeux d'entrée en l'absence de restrictions sur le mécanisme de sélection d'équilibre. En particulier, nous développons un algorithme qui nous permet de sélectionner de manière récursive un sous-ensemble d'inégalités qui caractérisent de façon minimale l'ensemble des paramètres admissibles. Ensuite, nous proposons une procédure inférentielle compétitive en lissant la fonction minimum. Cela nous permet d'obtenir une statistique de test pivotale qui élimine "numériquement" les moments non saturés. Nous montrons que nous récupérons une région de confiance convergente en laissant le lissage diminuer avec la taille de l'échantillon. Aussi, notre procédure peut facilement être adaptée au cas avec covariables, y compris continues. Enfin, nous menons des simulations de Monte Carlo pour évaluer les performances de notre nouvelle procédure d'estimation.

Le troisième chapitre s'intitule "Identification et Estimation des Contrats d'Incitation sous Information Asymétrique : une application au secteur de l'eau en France". Nous développons un modèle principal-agent pour représenter la sous-traitance de gestion pour la prestation de services publics. Une entreprise (l'Agent) possède une connaissance privée de son coût marginal de production. L'autorité publique locale (le Principal) se préoccupe du surplus net des consommateurs et du bénéfice de l'entreprise. La négociation contractuelle est modélisée comme le choix de l'entreprise dans un menu d'options déterminant le prix unitaire facturé aux consommateurs et le montant fixe. Notre modèle théorique caractérise la sous-traitance optimale dans cet environnement. Nous étudions ensuite l'identification non paramétrique du modèle et effectuons une estimation semi-paramétrique sur des données provenant de l'enquête de l'Institut Français de l'Environnement de 2004.

Contents

Chapter 1

Testing and Relaxing Distributional Assumptions on Random Coefficients in Demand Models

co-authored with Hippolyte Boucher and Gökçe Gökkoca

Abstract

The BLP demand model for differentiated products is the workhorse model for demand estimation with market-level data. This model uses random coefficients to account for unobserved preference heterogeneity. The shape of the distribution of random coefficients matters greatly for many counterfactual quantities, such as the pass-through of cost. In this paper, we develop new econometric tools to test this distribution and improve its estimation under a flexible parametrization. First, we develop a formal moment-based specification test on the distribution of random coefficients. The moment conditions (or equivalently the instruments) chosen for the test are designed to maximize the power of the test when the RC distribution is misspecified. Second, we show that our instruments can be successfully used to estimate a flexible distribution of random coefficients. Finally, we validate our approach with Monte Carlo simulations and an empirical application using data on car purchases in Germany. We also show that these methods extend to the mixed logit demand model with individual-level data.

Keywords: demand estimation, specification test, random coefficients
JEL codes: C35, C36, L13, C52

1.1 Introduction

The differentiated product demand model initiated by Berry (1994) and Berry, Levinsohn, and Pakes (1995) has been used in a wide array of empirical studies. It enables researchers to perform demand estimation in markets with differentiated products using either macro-level (market shares) or micro-level (individual purchases) data while allowing for unobserved heterogeneity in preferences as well as price endogeneity. This unobserved heterogeneity in preferences is modeled through the use of random coefficients (RCs) in the utility function. This framework allows researchers to estimate demand functions, price elasticities and counterfactual outcomes. Applications of the BLP model have notably studied the determinants of market power, the welfare effects resulting from a merger or the introduction of a new good and the economic impact of a tax or a subsidy.[1]

The informativeness of the empirical analysis depends on how well the model can reproduce the underlying substitution patterns and approximate the shape of the demand curve, including its slope and curvature. A recent result in Miravete, Seim, and Thurk (2022) shows that the commonly used Gaussian RC on price imposes strong restrictions on the demand's curvature and thus limits the range of the implied pass-through. The degree of pass-through of taxes and costs is central to answering many questions in economics such as the impact of tariffs or a cost shock on consumer welfare. However, estimating a more flexible demand system with a non-Gaussian distribution of random coefficients is challenging. First, there is a clear trade-off between the degree of flexibility one chooses (for instance, going from a Gaussian to a Gaussian mixture) and the precision of the estimates one obtains. Therefore, it is important to be able to test the specification chosen by the researcher on the distribution of the RC (for instance, a Gaussian RC) and quantify the degree of misspecification before potentially moving to a more flexible specification. Second, to precisely estimate a more flexible distribution of RC, the researcher must choose instruments (or equivalently moment conditions) that strongly identify this distribution. The

[1]The BLP demand model has been widely applied. A non-exhaustive list of examples includes: Barahona, Otero, Otero, and Kim (2020), Berry, Levinsohn, and Pakes (1995), Crawford, Shcherbakov, and Shum (2019), Dubois, Griffith, and O'Connell (2018), Durrmeyer (2022), Grennan (2013), Grigolon, Reynaert, and Verboven (2018), Miller, Sheu, and Weinberg (2021), Miller and Weinberg (2017), Miravete, Moral, and Thurk (2018), Nevo (2000), Petrin (2002), Reynaert (2021).

instruments used by the current empirical practice work well with the standard Gaussian RC, but their performance appears to decline as the specification becomes more flexible in the simulation exercises that we perform.

In this paper, we provide novel econometric tools to address these two challenges. In particular, we provide a formal moment-based specification test on the distribution of random coefficients, which allows researchers to test the chosen specification (for instance normality) without re-estimating the model under a more flexible parametrization. The moment conditions (or equivalently the instruments) chosen for the test are designed to maximize the power of the test when the distribution of RCs is misspecified. We also show how these instruments can strengthen the identifying power of the moment conditions used for estimation, and thus be successful at estimating a flexibly parameterized distribution of RCs. As an example of a flexible parametric distribution, we consider the Gaussian mixture, which can approximate arbitrarily well any continuous distribution on the real line.

This paper consists of three main contributions. First, we construct a powerful specification test on the distribution of random coefficients. The intuition we use is the following. Any given distribution of RCs generates a structural error, which, if correctly specified, is mean-independent with respect to a set of exogenous variables. This identifying condition can be transformed into unconditional moments, which can be used to test whether the chosen distribution of RCs is correctly specified. We formally define this test and construct instruments that maximize its power against a fixed alternative. In a first step, we assume that the econometrician knows the fixed alternative and we derive an expression for the first-best instrument. We call this instrument the most powerful instrument (MPI) and show that this specific choice of instrument achieves the consistency of the test. In a second step, we provide two feasible approximations of the MPI that can be derived without the knowledge of the fixed alternative. We call these feasible MPIs the interval instruments in reference to the way they approximate the MPI.

Second, we consider the case where the researcher wants to test whether the distribution of RCs belongs to a given parametric family. For instance, the researcher may be interested in testing if the random coefficient is normally distributed. This is a composite hypothesis, and we must estimate the unknown parameters of the distribution in a first step. In a second step, we choose instruments to test if the distribution evaluated at the estimated parameters is correctly specified. Here, the interval instruments represent a natural choice of instruments as they are designed to detect deviations from the true distribution of RCs. We study the asymptotic properties of our test when the number of markets, T, goes to infinity and we prove the asymptotic validity of the test under common assumptions. In particular, we account for the statistical uncertainty stemming from the first step estimation, and we control for the magnitude of the approximations that intervene in the estimation of the BLP model. Our asymptotic results complement previous work by Freyberger (2015) on the asymptotic properties of the BLP estimator when the number of markets grows to infinity.

Third, we show that our interval instruments can be successfully used to estimate the model, and particularly so when the distribution of RCs is flexibly parameterized. We do so by exhibiting the connection between the MPI and the classical optimal instruments used for efficient estimation purposes. Specifically, we show that the MPI devoted to testing the specification of the model at the true parameter against any local alternative can be rewritten as a linear combination of the optimal instruments. This relation between the MPI and the optimal instruments helps us understand why the interval instruments, which approximate the MPI, perform so well in our simulations. So far, the literature has exclusively exploited instruments that approximate the optimal instruments (Gandhi and Houde (2019), Reynaert and Verboven (2014)). We refer to these instruments as the traditional instruments. These have been shown to work well in the usual Gaussian case. However, our simulations show that their performance declines when we depart from the Gaussian RC.

To evaluate the performance of our test and instruments, we conduct two sets of simulation experiments. First, we compare the performance of the test when using our interval instruments and when using the instruments commonly adopted by practitioners (Gandhi and Houde (2019), Reynaert and Verboven (2014)). We show that the test has the correct empirical size and that the interval instruments significantly outperform the traditional instruments in terms of power under alternative distributions. Second, we evaluate the performance of the interval instruments in estimating the model when the distribution of

RCs is flexibly parametrized, and follows a Gaussian mixture. We show that our instruments outperform the traditional instruments in terms of the mean squared error. In the case where the RC is a Gaussian mixture, the three sets of instruments perform equally well.

Finally, we apply the tools developed in this paper to estimate the demand for cars in Germany from 2012 to 2018. The objective of the empirical exercise is to see how well our instruments perform at estimating a flexible distribution of RCs using a real dataset. Given the importance of price to address most empirical questions, we increase the flexibility of the model by estimating a Gaussian mixture for the RC associated with price. Second, we use our specification test to assess how the degree of misspecification decreases when we increase the flexibility in the distribution of RCs. Third, we use our results to study how the shape of the RC on price can modify important counterfactual quantities such as the pass-through. In particular, our empirical results are consistent with the findings in Miravete, Seim, and Thurk (2022).

Related literature. Our paper contributes to several strands of the literature. First, it contributes to the literature on the flexible estimation of aggregate demand models for differentiated goods. A few recent papers have proposed non-parametric and semi-parametric methods to estimate aggregate demand functions. Compiani (2018) proposes a non-parametric estimator of the demand functions. If relaxing all the parametric assumptions makes this approach conceptually appealing, it also faces significant theoretical and practical difficulties (more stringent data requirements, large curse of dimensionality, limited scope for counterfactual analysis).[2] Lu, Shi, and Tao (2021) and Wang (2022) propose semi-parametric estimators of the distribution of RCs. These approaches are complementary to ours and the instruments we develop in this paper can be useful to implement their non-parametric IV estimation procedures, which are known to be rather sensitive to the quality of the instruments (Chetverikov and

[2] In particular, Compiani (2018) relaxes the Type 1 Extreme Value assumption on the taste shock. However, it is not clear how restrictive this assumption is. McFadden and Train (2000) shows that a mixed-logit model with flexibly distributed random coefficients can approximate any discrete choice model derived from random utility maximization. On the other hand, the Type 1 Extreme Value assumption generates massive computational gains, which allows for studying sophisticated markets with many products and many characteristics. Thus, the cost-benefit analysis seems to be largely in favor of the logit specification.

Wilhelm (2017)). Finally, Ho and Pakes (2014), Tebaldi, Torgovitsky, and Yang (2019) suggest deriving bounds directly on the counterfactual quantities.

Our paper also contributes to the literature on the non-parametric identification of the distribution of RCs in demand models (Fox and Gandhi (2011), Fox, il Kim, Ryan, and Bajari (2012), Dunker, Hoderlein, and Kaido (2022), Wang (2022), Berry and Haile (2014)). First, we slightly extend the identification result in Wang (2022) to link it directly to the primitives of the model, without assuming that the demand functions are identified. Second, we provide a practical way of constructing moments that feature high identifying power with respect to the distribution of RCs.

Third, we contribute to the literature that focuses on the practical estimation of the BLP model. First, we show that the interval instruments that we construct in this paper can be successfully used to estimate the distribution of random coefficients, and particularly so under of flexible distribution of RCs. This new set of instruments complements instruments commonly used by practitioners: Reynaert and Verboven (2014) and Gandhi and Houde (2019) (see Conlon and Gortmaker (2020) for a review). Moreover, we provide a new parametrization of the model, which facilitates the estimation when the distribution of RCs is a Gaussian mixture. This new parametrization complements previous papers that aim at improving the estimation of the model (Dubé, Fox, and Su (2012), Lee and Seo (2015), Salanié and Wolak (2019)).

Finally, our paper contributes to the literature on the asymptotic properties of the BLP estimator (Armstrong (2016), Berry, Linton, and Pakes (2004), Freyberger (2015), Ketz (2019)). In particular, we prove the asymptotic normality and the consistency of the BLP estimator in the large market framework under less stringent assumptions than the remainder of the literature.

Structure of the paper. In Section 1.2, we recall the baseline BLP model, define the structural error of the model, and provide conditions under which the distribution of RCs is non-parametrically identified. In Section 1.3, we present our specification test and derive the most powerful instrument and last we show how it relates to the classical optimal instruments. In Section 1.4, we construct two feasible approximations of the MPI. In Section 1.5, we adapt the specification test to the composite hypothesis case and show its asymptotic validity. In Section 1.6, we conduct Monte Carlo simulations to evaluate the consequences of misspecification on quantities of interest, and gauge the performance of our test and instruments. In Section 1.7, we apply our new tools to estimate the demand for cars in Germany. We conclude the paper in Section 1.8.

1.2 Model and identification

1.2.1 Indirect utility and moment restrictions

Indirect utility. We first describe the indirect utility function that induces the observed market shares. Our setting closely follows the one introduced in the seminal paper Berry, Levinsohn, and Pakes (1995). There are T markets indexed by $t = 1, ..., T$. There is a continuum of consumers indexed by i. There are J_t market-specific products in market t. Each consumer chooses a product $j \in \{0, 1, ..., J_t\}$ where $j = 0$ corresponds to the outside option. For the sake of exposition and without loss of generality, we will assume throughout our analysis that the number of products is constant across markets ($\forall t,\ J_t = J$). Product j is characterized by a vector of characteristics x_{jt}, which includes the price of the good in most empirical settings. Consumer i derives an indirect utility u_{ijt} from purchasing good $j \in \{0, 1, \ldots, J\}$ in market t:

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt}v_i + \varepsilon_{ijt}, \tag{1.2.1}$$

with the following:

- x_{1jt} is a vector of product characteristics of dimension K_1 associated with product j and for which there is no preference heterogeneity; β represents preferences for x_{1jt};
- ξ_{jt} is an unobserved demand shock on product j in market t;
- $\delta_{jt} \equiv x'_{1jt}\beta + \xi_{jt}$ denotes the mean utility for product j, the part of the utility that is common to all consumers;

- x_{2jt} is a vector of product characteristics of dimension K_2 for which there is preference heterogeneity; v_i is the associated random coefficient that follows a distribution characterized by density f and is independent of all the other variables: $v_i \perp\!\!\!\perp (x_t, \xi_t, \{\varepsilon_{ijt}\}_{j=1,...J})$;

- ε_{ijt} is a preference shock that follows an Extreme Value type I (EV1) distribution independent of all other variables and across i, j, t.

For individual i in market t, the indirect utility from purchasing the outside option is normalized to $u_{i0t} = \varepsilon_{i0t}$. From the random utility functions in (1.2.1), we can infer the demand functions for each good j in market t denoted $\rho_{jt}(f, \beta)$. Each consumer chooses the product that maximizes his or her utility. Let y_{ijt} equal 1 if individual i chooses good $j = 0, 1, \ldots, J$ in market $t = 1, \ldots, T$. We have the following:

$$\begin{aligned}
\forall j \neq 0,\;\; \rho_{jt}(f, \beta) &\equiv \mathbb{P}_{f,\beta}(y_{ijt} = 1 | x_t, \xi_t) \\
&= \mathbb{P}_{f,\beta}(\text{good } j \text{ is chosen in market } t \text{ by individual } i | x_t, \xi_t) \\
&= \mathbb{P}_{f,\beta}(u_{ijt} > u_{ikt}\; \forall k \neq j | x_t, \xi_t) \\
&= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{1jt}\beta + \xi_{jt} + x'_{2jt}v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v)dv. \qquad (1.2.2)
\end{aligned}$$

For the outside option, the demand function is written as follows:

$$\rho_{0t}(f, \beta) = \mathbb{P}_{f,\beta}(y_{i0t} = 1 | x_t, \xi_t) = \int_{\mathbb{R}^{K_2}} \frac{1}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v)dv.$$

Following the EV1 assumption on the idiosyncratic shock on utility, the demand functions take the usual logit form integrated over the distribution of preference heterogeneity. We assume in this paper that the observed market shares are equal to the shares generated by the model above at the true distribution f and the true preference parameter β:

$$\forall j,\;\; \forall t, \qquad s_{jt} = \rho_{jt}(f, \beta). \qquad (1.2.3)$$

Moment restrictions. Following the literature, we assume that the unobserved demand shock ξ_{jt} is mean independent of z_{jt}, a set of instrumental variables, namely, $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ *a.s.*. The set z_{jt} traditionally consists of the exogenous characteristics of all the products on the market as well as cost shifters, which are meant to instrument for price. Indeed, the price of a good is usually considered to be an endogenous variable since it is correlated with the unobserved demand shock ξ_{jt} through the profit maximization problem of firms.[3] To estimate the model, the researcher chooses functions of the instruments z_{jt} to construct a set of unconditional moments. We refer to these functions as estimation instruments and denote them $h_E(z_{jt})$. Likewise, in our analysis, we study the functions of the instruments that are designed to test the specification of the model. We refer to these instruments as testing instruments and we denote them $h_D(z_{jt})$, where D stands for detection.

1.2.2 Inverse demand function and structural error

Inverse demand function. For any given distribution of random coefficients $\tilde{f}$, we define the demand function $\rho \equiv (\rho_1(\cdot), ..., \rho_J(\cdot))$ as the function which maps the vector of mean utilities δ to the vector of market shares generated by the model under $\tilde{f}$:

$$
\begin{aligned}
\rho(\cdot, x_{2t}, \tilde{f}) :\ & \mathbb{R}^J \to [0,1]^J \\
& \delta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta + x_{2t} v\}}{1 + \sum_{k=1}^{J} \exp\{\delta_k + x'_{2kt} v\}} \tilde{f}(v) dv.
\end{aligned}
$$

Berry (1994) shows by applying Brouwer's fixed point that for any (s_t, x_{2t}) and for any distribution of random coefficients $\tilde{f}$ (even when $\tilde{f}$ is not the true distribution), there exists a unique $\tilde{\delta} \in \mathbb{R}^J$ such that:

$$
s_t = \rho(\tilde{\delta}, x_{2t}, \tilde{f}).
$$

[3]To deal with the endogeneity of prices, Berry, Levinsohn, and Pakes (1995) also suggests using exogenous own-product characteristics as well as exogenous characteristics from other products. The main idea behind the use of these instruments is to take advantage of the correlation between price and exogenous characteristics implied by profit-maximizing firms. To be precise, Berry, Levinsohn, and Pakes (1995) suggests using the sum of the characteristics from other products produced by the same firm and the sum of exogenous characteristics from rival firms' products as instruments.

We define the solution to the previous system of equations as the inverse demand functions: $\rho^{-1}(s_t, x_{2t}, \tilde{f}) = \tilde{\delta}$. Unfortunately, there is no closed form expression for the inverse demand function, which must be recovered numerically.

Structural error. From what precedes, we can uniquely define the structural error $\xi_{jt}(\tilde{f}, \tilde{\beta})$ generated by a distribution of random coefficient $\tilde{f}$ and a homogeneous parameter $\tilde{\beta}$:

$$\xi_{jt}(\tilde{f}, \tilde{\beta}) = \rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x'_{1jt}\tilde{\beta}. \tag{1.2.4}$$

The non-linear nature of the model is captured by the inverse demand function which enters the expression of the structural error. The absence of an analytical formula for the inverse demand implies that there is no closed form expression for the structural error, which complicates the estimation of the BLP demand model. If we consider a parametric family of distributions $\tilde{\mathcal{F}} = \{\tilde{f}(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \tilde{\Lambda}\}$, then the structural error generated by a specific element in $\tilde{f}(\cdot|\tilde{\lambda}) \in \tilde{\mathcal{F}}$ and $\tilde{\beta}$ is defined as follows:

$$\xi_{jt}(\tilde{f}(\cdot|\tilde{\lambda}), \tilde{\beta}) = \rho_j^{-1}(s_t, x_{2t}, \tilde{f}(\cdot|\tilde{\lambda})) - x'_{1jt}\tilde{\beta}.$$

1.2.3 Non-parametric identification

The main objective of this paper is to provide tools to test the specification on the distribution of random coefficients and to improve its estimation under a flexible specification. A natural first step is to study the conditions under which this distribution is non-parametrically identified. The identification of random coefficients in multinomial choice models has been studied extensively in the literature (Allen and Rehbeck (2020), Berry and Haile (2014), Dunker, Hoderlein, and Kaido (2022), Fox and Gandhi (2011), Fox, il Kim, Ryan, and Bajari (2012), Wang (2022)). We summarize some of these findings in Appendix 1.C.1. In this Section, we build on an important identification result in Wang (2022) to recover a set of sufficient identifying conditions directly on the primitives of the model. We also show that the identification result holds with a less stringent exogeneity assumption than in Wang (2022).

In contrast to the rest of the literature, Wang (2022) adopts all the parametric assumptions in the standard BLP model and looks for a set of sufficient restrictions under which the identification of the demand functions implies the identification of the distribution of random coefficients. This approach allows him to obtain conditions that are less stringent than the rest of the literature. In particular, Wang (2022) makes no special regressor assumption, no full support assumption, and no continuity assumption on the covariates. Specifically, he shows that if the demand functions $\rho = (\rho_1, ..., \rho_J)$ are identified on an open set of $\mathbb{R}^J$, then the distribution of random coefficients is identified.[4] His proof exploits the real analytic property of the demand functions.[5] In this paper, we build on this injectivity result to find sufficient identifying conditions directly on the primitives of the model (without assuming identification of the demand functions). We also show using a random permutation of the indices that we only require the demand shock ξ_{jt} to be mean independent of the instrumental variables z_{jt} across products, but we do not require this to hold for each product j taken separately. Formally, we only require $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ *a.s.* and not $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ *a.s.* for all product j as previously. This is less restrictive, as demand shocks can now be on average non-zero for certain products and account for unobserved quality inherent to each product.

Let us formally state the assumptions that we impose to recover the point identification of (f, β).

Assumption A

(i) Strict exogeneity: $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ *a.s.;*

(ii) Completeness: for any measurable function g such that $\mathbb{E}[|g(s_t, x_t)|] < \infty$, if $\mathbb{E}[g(s_t, x_t)|z_{jt}] = 0$ a.s., then $g(s_t, x_t) = 0$ a.s.;

(iii) The distribution of the data $(s_t, x_{2t}, x_{1t}, z_t)$ is fully observed by the econometrician and market shares s_t are generated by the demand model defined in Section 1.2.1 by equations (1.2.1) and (1.2.3);

(iv) Detectable difference in distributions: we say f and $\tilde{f}$ differ (and write $f \neq \tilde{f}$) if there exists $\bar{v} \in \mathbb{R}^{K_2}$ such that $F(\bar{v}) \neq \tilde{F}(\bar{v})$;

(v) Let $x_t = (x_{1t}, x_{2t})$ then x_t is such that $\mathbb{P}(x_t' x_t \text{ is positive definite}) > 0 \ \forall t$;

(vi) There exists $\bar{x}_t \in \mathcal{X}$ and an open set $\mathcal{D} \subset \mathbb{R}^J$ such that $\delta_t = \bar{x}_{1t}\beta_0 + \xi_t$ varies on $\mathcal{D}$ a.s..

[4] Identification of demand functions can be achieved using Theorem 1 in Berry and Haile (2014).

[5] In particular, the real analytic property yields that the local identification of ρ on $\mathcal{D} \subset \mathbb{R}^J$ implies the identification of ρ on $\mathbb{R}^J$ From the global identification of ρ, he is then able to show that the random coefficients' distribution is identified under a simple rank condition on x_{2t}.

In A(i), we assume that the instruments are strictly exogenous. Assumption A(ii) is a completeness assumption that states that the instruments are strongly relevant with respect to (s_t, x_t). This assumption is typical of semiparametric or nonparametric IV models and is equivalent to a full rank assumption in a linear IV model. Intuitively, it means that if the inverse demands are different almost surely, then the instruments will be able to detect the difference. The completeness assumption is a strong assumption that has been widely used in this literature (Berry and Haile (2014), Dunker, Hoderlein, and Kaido (2022), Wang (2022)). Assumption A(v) is a standard rank condition. Assumption A(vi) is meant to ensure that there is enough variation in δ_t to apply the injectivity result in Wang (2022). This assumption indicates that there needs to be sufficient variation in product characteristics across markets in the data to identify f. In practice, product characteristics are very similar from one market to the other and may not yield sufficient variation. A judicious solution is to create inter-market variation by interacting product characteristics with demographic variables characterizing each market. Let us now state our formal identification result.

Proposition 2.1 *Under Assumption A, the distribution of random coefficients f and the homogeneous preference parameters β are non-parametrically identified:*

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \iff \quad \mathbb{E}[\xi_{jt}(\tilde{f}, \tilde{\beta})|z_{jt}] = \mathbb{E}\left[\rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x'_{1jt}\tilde{\beta}\middle| z_{jt}\right] = 0 \ \ a.s..$$

The proof is in Appendix 1.B.1. The identification result above entails that under some fairly weak conditions and in the presence of instruments that generate sufficient variation in the product characteristics, the observed data identifies the distribution of random coefficients non-parametrically. Formally, the model is at the true pair (f, β) if and only if the associated structural error is mean independent of the instrumental variables z_{jt}. We use this identification result to show the consistency of our test under a specific choice of instruments that we will characterize thereafter.

1.3 Detecting misspecification: the most powerful instrument

The aim of this section is to recover the instrument with the greatest ability to detect misspecification in the distribution of RCs. To do so, we consider a setting in which the econometrician wants to test a simple hypothesis of the form $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$. The upper bar is used to stress the fact that $\overline{H}_0$ is a simple hypothesis, in contrast to the composite hypothesis $H_0 : f \in \mathcal{F}_0$ that we study in Section 1.5. Our approach builds on a simple intuition: if the model under $\overline{H}_0$ is misspecified, then the structural error will depart from the true demand shock ξ_{jt}, and our goal is to find the best instrument to pin down this deviation. We proceed as follows. First, we introduce a moment-based test for $\overline{H}_0$ and we show its asymptotic validity. Next, we derive an analytical expression for the instrument that maximizes the power of our test against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$. We call this instrument the most powerful instrument (MPI) and we show how it relates to the classical optimal instruments, derived for efficient estimation purposes. In Section 1.4, we provide two feasible approximations of the MPI, which have the critical property of being invariant with respect to the alternative $\overline{H}_a$.

1.3.1 A moment-based test

We want to test $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against $H_a : (f, \beta) \neq (f_0, \beta_0)$. For any set of testing instruments $h_D(z_{jt})$, we have the following implication:

$$\overline{H}_0 : (f, \beta) = (f_0, \beta_0) \implies \overline{H}_0' : \quad \mathbb{E}[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)] = 0.$$

We propose to test $\overline{H}_0$ indirectly through its implication $\overline{H}_0'$, which is a set of unconditional moment conditions. We test $\overline{H}_0'$ with a moment-based test. Our test statistic writes as follows:

$$S_T(h_D, f_0, \beta_0) = TJ \left(\frac{1}{TJ} \sum_{j,t} \xi_{jt}(f_0, \beta_0) h_D(z_{jt}) \right)' \hat{\Omega}_0^{-1} \left(\frac{1}{TJ} \sum_{j,t} \xi_{jt}(f_0, \beta_0) h_D(z_{jt}) \right), \qquad (1.3.5)$$

with $\hat{\Omega}_0$ a consistent estimator of Ω_0 the asymptotic variance-covariance matrix of $\frac{1}{\sqrt{TJ}} \sum_{j,t} h_D(z_{jt})\xi_{jt}(f_0, \beta_0)$, that is $\Omega_0 = \mathbb{E}[\xi_{jt}^2(f_0, \beta_0) h_D(z_{jt}) h_D(z_{jt})']$. We study the asymptotic properties of our test as the number of markets, T, goes to infinity. As the focus of this section is on the construction of the most powerful instrument, we postpone the treatment of the specific challenges implied by parameter uncertainty (i.e. when β_0 and f_0 must be estimated beforehand) and by the numerical approximations involved in the derivation of the structural error (in practice, the researcher derives a numerical approximation of $\xi_{jt}(f_0, \beta_0)$) to Section 1.5. Additionally, to keep the results as simple as possible while retaining the key intuitions, we assume independence of the demand shocks in a given market conditional on z_{jt}. This last assumption is relaxed in the proofs in Appendix 1.B.2 and in Section 1.5.

Proposition 3.1 *Assume that (s_t, x_t, z_t) are i.i.d. across markets and consistent with the probability model defined by equations (1.2.1), (1.2.2) and (1.2.3) evaluated at (f, β), $\mathbb{E}[\|\xi_{jt}(f_0, \beta_0) h_D(z_{jt})\|^2] < +\infty$, Ω_0 has full rank, and, for $k \neq j$, $\xi_{jt} \perp\!\!\!\perp \xi_{kt} | z_t$. We have the following:*

- *under $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$,* $\quad S_T(h_D, f_0, \beta_0) \xrightarrow[T \to +\infty]{d} \chi^2_{|h_D|_0}$,
- *under $H'_a :$ $\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right] \neq 0$,* $\quad \forall q \in \mathbb{R}^+,\ \mathbb{P}(S_T(h_D, f_0, \beta_0) > q) \xrightarrow[T \to +\infty]{} 1$,

with $|\cdot|_0$ being the counting norm.

The previous proposition indicates that as long as the testing instruments are functions of z_{jt}, our test procedure is asymptotically valid for $\overline{H}_0$. We are testing $\overline{H}_0$ by virtue of its implication $\overline{H}'_0$: $\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right] = 0$ and, as a consequence, the power properties of our test hinge critically on the choice of the testing instruments $h_D(z_{jt})$. This is the focus of the next subsection.

1.3.2 The most powerful instrument (MPI)

The choice of testing instruments $h_D(z_{jt})$ is key to maximize the rejection rate of $\overline{H}_0$ under any alternative $H_a : (f, \beta) \neq (f_0, \beta_0)$. To guide our choice of instruments, we first derive the instrument that maximizes the power of the moment-based test introduced previously when the econometrician tests $\overline{H}_0$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0)$. We refer to this instrument as the most powerful instrument (MPI). In practice, the researcher is often reluctant to fix the alternative. However, the MPI represents a useful first-best solution for which we provide feasible approximations in Section 1.4.

Power criterion. We now introduce the criterion that we use to define the most powerful instrument. The literature offers many ways to compare the power of competing tests (see Gourieroux and Monfort (1995) for a comprehensive review). In this paper, we favor the non-local approach developed in Bahadur (1960). In Bahadur's perspective, the econometrician chooses the test with the smallest level α needed to attain a given power against a fixed alternative and for a given number of observations. In other words, the econometrician chooses the test that minimizes the risk of type I error ceteris paribus. The main alternative to this approach is to discriminate between two tests based on their power against local alternatives.[6] In a parametric framework, local strategies are based on the analysis of the power properties of competing tests under a sequence of local alternatives θ_T which converges to θ_0 at a given rate (usually $\frac{1}{\sqrt{T}}$). The econometrician can compare two competing tests by means of their power functions (or more precisely, the limits of these power functions when sample sizes go to $+\infty$). This is called the direct approach. The dual approach, which is known as Pitman's relative efficiency, consists of comparing the rates at which the minimal number of observations must increase to ensure a given level of power.

We believe that Bahadur's non-local approach is better suited for the testing problem we study in this paper. The comparison criterion, known as the asymptotic slope of the test, is in our case straightforward to derive, whereas it is not clear how one should derive Pitman's efficiency criterion when the test concerns non-parametric objects such as distributions. Moreover, we study the properties of our test against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$ as in Bahadur's case, which is not necessarily local. Finally, the literature has highlighted many limitations of the local approach. Local criteria are often unable to discriminate between tests even when these tests lead to different decisions (see Silvey (1959)). In addition, as shown in Dufour and King (1991), a locally optimal test in a neighborhood of H_0 may perform very poorly away from H_0.

[6] In the interest of conciseness, we voluntarily omit the class of tests that rely on the exact distributions of the test statistic as, in our case, the exact distribution of our test is unknown. Thus, we rely on asymptotic methods, which is the most common case in the literature.

Let us now present the intuition for Bahadur's comparison approach. From Section 1.3.1, we have:

$$\text{Under } \overline{H}_0: \qquad S_T \equiv S_T(h_D, f_0, \beta_0) \xrightarrow{d} S \qquad \text{with } S = \chi^2_{|h_D|_0}.$$

Following the same notations as in Gourieroux and Monfort (1995), we denote:

$$\Lambda(s) = \mathbb{P}_{\overline{H}_0}(S \geq s).$$

The critical value is usually derived using the asymptotic distribution of the test statistic under $\overline{H}_0$. The approximate critical region at a given level α is then given by:

$$CR_\alpha = \{S_T \geq \Lambda^{-1}(\alpha)\} = \{\Lambda(S_T) \leq \alpha\}.$$

The main idea in Bahadur's approach entails deriving the level of the test if one takes the value of the test statistic as the critical value (this is also known as the p-value). Namely:

$$\alpha_T = \Lambda(S_T).$$

Bahadur suggests preferring the test that displays the lowest level α_T at least asymptotically. A formal analysis of the asymptotic behavior of α_T shows that it is better to consider the limit of a transformation of α_T than the limit of α_T directly. This gives rise to the concept of the approximate slope of the test.

Definition 1 (Asymptotic slope of the test)

(i) $K_T = -\frac{2}{T}\log(\Lambda(S_T))$ *is the approximate slope of the test,*

(ii) Under $\overline{H}_a$*:* $plim\ K_T = c(f_a, \beta_a)$ *is the asymptotic slope of the test,*

with $plim$*, the limit in probability when* $T \to +\infty$*.*

Under the alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$, consider two sequences of tests based on S_T^1 and S_T^2 with asymptotic slopes $c^1(f_a, \beta_a)$ and $c^2(f_a, \beta_a)$ respectively. The test based on S_T^1 is asymptotically preferred to the test based on S_T^2 in Bahadur's sense if and only if $c^1(f_a, \beta_a) > c^2(f_a, \beta_a)$. To derive the asymptotic slopes of our test, we apply an important result in Geweke (1981), which states that if under H_0: $S_T \xrightarrow[T\to+\infty]{d} \chi^2_q$ (with any $q \in \mathbb{N}^*$), then $\frac{1}{T}S_T \xrightarrow{a.s.} c(f_a, \beta_a)$ (when the limit exists). In our test, the limiting distribution is chi-squared. Thus, the asymptotic slope of our test with instrument $h_D(z_{jt})$ writes:

$$c_{h_D}(f_a, \beta_a) = plim\ \frac{1}{T} S_T(h_D, f_0, \beta_0) = J\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right]' \Omega_0^{-1} \mathbb{E}\left[\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right].$$

Let us note that the asymptotic slope can also be interpreted as a measure of the speed of divergence of the test statistic in terms of population moments, i.e. speed of divergence $\approx T \times c_{h_D}(f_a, \beta_a)$. An important remark is that the asymptotic slope allows us to define an envelope on the power that can be attained by our moment-based test. In the next Proposition, we derive an analytical expression for the instrument that maximizes the slope of the test against a fixed alternative $\overline{H}_a$

Derivation of the most powerful instrument. To construct the MPI, we use the following decomposition of the structural error generated under $\overline{H}_a$:

$$\xi_{jt}(f_0, \beta_0) = \underbrace{\xi_{jt}(f_a, \beta_a)}_{\text{true error under } \overline{H}_a} + \underbrace{\xi_{jt}(f_0, \beta_0) - \xi_{jt}(f_a, \beta_a)}_{\Delta_{0,a}^{\xi_{jt}}},$$

with $\Delta_{0,a}^{\xi_{jt}}$ being the correction term due to misspecification under the alternative $\overline{H}_a$.

Proposition 3.2 (Most powerful instrument)

Let $\mathcal{H}$ be the set of measurable vectorial functions of z_{jt}. Under any fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$, we have the following:

$$\left(\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)^2 | z_t\right]\right)^{-1} \mathbb{E}[\Delta_{0,a}^{\xi_{jt}} | z_{jt}] \in \underset{h_D \in \mathcal{H}}{argmax}\ c_{h_D}(f_a, \beta_a).$$

The proof is given in Appendix 1.B.2. The MPI equals the conditional expectation of the correction term $\Delta_{0,a}^{\xi_{jt}}$ divided by a conditional variance term $\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)^2 | z_{jt}\right]$. For exposition purposes, we drop the conditional variance term in the subsequent analysis and take the homoskedastic MPI

$h^*_D(z_{jt}) = \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]$ as the reference MPI.[7] Methods have been proposed to estimate the conditional variance term non-parametrically and could be adapted to our case. However, it is well known that the conditional variance, which also appears in the formulation of the optimal instruments, is difficult to model and estimate in practice. In the BLP framework, the large dimension of z_{jt} makes the exercise even more difficult. Hence, researchers typically ignore this term or impose a restrictive and ad-hoc structure on the form that it can take (for instance, Reynaert and Verboven (2014)'s approximation of the optimal instruments in the BLP model ignores the variance term). The homoskedastic MPI, $h^*_D(z_{jt})$, features other appealing properties including (i) consistency of the associated test and (ii) maximizing correlation with the structural error under the alternative.[8] For simplicity, in what follows, we refer to the homoskedastic MPI as the MPI.

(i) Consistency. By setting h_D equal to h^*_D, our moment-based test becomes consistent against any fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0)$. Namely, we have the following result:

Proposition 3.3 (Consistency of the test with the MPI) *Under Assumption A and the same assumptions as in Proposition 3.1, we have:*

$$\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0) \implies \forall q \in \mathbb{R}^+, \; \mathbb{P}(S_T(h^*_D, f_0, \beta_0) > q) \underset{T \to +\infty}{\longrightarrow} 1.$$

The proof of this result is given in Appendix 1.B.2.

(ii) Correlation with the structural error. Another interesting property of the MPI is to be the function of z_{jt} that maximizes the correlation with the structural error.

Proposition 3.4 (Correlation between the MPI and the structural error)
Let $\mathcal{H}$ be the set of measurable functions of z_{jt}, we have under $\overline{H}_a$:

$$\forall \alpha \in \mathbb{R}^*, \quad \alpha \, \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] \in \underset{h \in \mathcal{H}}{\arg\max} \; \left| corr(\xi_{jt}(f_0, \beta_0), h(z_{jt})) \right| .$$

[7] This last expression corresponds to the exact formulation of the MPI under homoskedasticity.

[8] The consistency of the test also holds when we keep the conditional variance term.

The proof is given in Appendix 1.B.2. Intuitively, the MPI $h_D^*(z_{jt})$ is designed to fully capture the exogenous variation contained in the correction term $\Delta_{0,a}^{\xi_{jt}}$ implied by the misspecification, which yields the result above.

1.3.3 Connection with the optimal instruments

The optimal instruments from Chamberlain (1987) minimize the asymptotic variance-covariance of the GMM estimator when the parameter of interest is identified by conditional moment restrictions. We show that the MPI devoted to testing the specification of the model at the true parameter against any fixed local alternative can be rewritten as a linear combination of the optimal instruments. This connection between the MPI and the optimal instruments helps us understand why the feasible approximations of the MPI we construct in Section 1.4 improve the performance of the BLP estimator in our Monte Carlo simulations when the distribution of RCs is flexible. In this subsection, we first derive the optimal instruments. Then, we exhibit the relation between the optimal instruments and the MPI.

The estimation of the model works as follows. The researcher assumes that f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ and the objective is to estimate the true parameter $\theta_0 = (\beta_0', \lambda_0')'$ under this parametric restriction. In the estimation context that we study here, θ_0 refers to the true parameter. For now, let us assume that the model is correctly specified: $f \in \mathcal{F}_0$ and we shorten the notations by removing the dependence of the structural error in $f_0(\cdot|\tilde{\lambda})$, which becomes implicit in this context. Namely, $\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$ becomes $\xi_{jt}(\tilde{\theta})$. We further assume that θ_0 is point identified by the following moment restriction: $\mathbb{E}[\xi_{jt}(\theta_0)|z_{jt}] = 0$ *a.s.*.[9] The researcher must choose the set of instruments $h_E(z_{jt})$ (or equivalently, the unconditional moments) to include in the GMM objective function:

$$\hat{\theta} = \underset{\tilde{\theta}}{\text{Argmin}}\, TJ \left(\frac{1}{TJ} \sum_{j,t} \hat{\xi}_{jt}(\tilde{\theta}) h_E(z_{jt}) \right)' \hat{W} \left(\frac{1}{TJ} \sum_{j,t} \hat{\xi}_{jt}(\tilde{\theta}) h_E(z_{jt}) \right).$$

[9] The identification conditions in the parametric case are less stringent than the conditions for the non-parametric identification in Assumption A.

Optimal instruments in the BLP demand model. Traditionally, the instruments $h_E(z_{jt})$ are chosen to minimize the asymptotic variance-covariance of the estimator $\hat{\theta}$. The instruments that reach this objective are called the optimal instruments. The resulting estimator is said to be efficient in the sense that its asymptotic variance cannot be reduced by using additional moment conditions. There is a large body of literature on the derivation of optimal instruments in econometric models (Amemiya (1974), Chamberlain (1987), Newey (1990, 2004)). The BLP estimator $\hat{\theta}$ is a non-linear GMM estimator and classical results in Chamberlain (1987) and Amemiya (1974) show that the optimal instruments in this case write:

$$h_E^*(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]^{-1}\mathbb{E}\left[\frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}\middle|z_{jt}\right],$$

The corresponding efficiency bound (obtained by setting $h_E = h_E^*$) writes:

$$V^* = \mathbb{E}\left[\mathbb{E}\left[\frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}\middle|z_{jt}\right]\mathbb{E}\left[\frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}\middle|z_{jt}\right]'\mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]^{-1}\right]^{-1}.$$

For the sake of exhaustivity, we show this result in Appendix 1.B.2. As for the MPI, the formulation of the optimal instruments above is obtained under the assumption of conditional independence of demand shocks ξ_{jt} in the same market: $k \neq j$, $\xi_{jt} \perp\!\!\!\perp \xi_{kt}|z_t$. In Appendix 1.B.2, we derive the expression for the optimal instruments under weaker assumptions on the demand shock.[10] Consistent with what we did in the case of the MPI, we drop the conditional variance term $\mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]^{-1}$.

Connection between the MPI and the optimal instruments. Let θ_0 the true parameter. Under the parametric assumption $f \in \mathcal{F}_0$, the simple hypothesis $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ we studied previously becomes $\overline{H}_0 : \theta = \theta_0$. It is straightforward to show that, in the parametric case, the associated MPI against a fixed alternative $\overline{H}_a : \theta = \theta_a$ writes: $h_D^*(z_{jt}) = \mathbb{E}\left[\Delta_{\theta_0,\theta_a}^{\xi_{jt}}|z_{jt}\right]$ with $\Delta_{\theta_0,\theta_a}^{\xi_{jt}} = \xi_{jt}(\theta_0) - \xi_{jt}(\theta_a)$. By taking a Taylor expansion of $\xi_{jt}(\theta_a)$ around θ_0, we obtain the following:

$$\Delta_{\theta_0,\theta_a}^{\xi_{jt}} = \frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}(\theta_0 - \theta_a) + o(||\theta_0 - \theta_a||_2).$$

[10] We allow for unrestricted forms of correlation between demand shocks within a given market.

We see that when θ_a is in a neighborhood of θ_0, the MPI, $h_D^*(z_{jt})$, against this fixed alternative is a linear combination of the optimal instruments $h_E^*(z_{jt})$:

$$h_D^*(z_{jt}) = \mathbb{E}\left[\Delta_{\theta_0,\theta_a}^{\xi_{jt}} | z_{jt}\right] \approx \underbrace{\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}\bigg| z_{jt}\right]'}_{h_E^*(z_{jt})}(\theta_0 - \theta_a).$$

It follows that classical optimal instruments can be interpreted as an approximation of the MPI devoted to testing $H_0 : \theta = \theta_0$ against any fixed local alternative.[11] Moreover, let us note that the connection between the MPI and the optimal instruments holds if we keep the conditional variance term in both cases.

1.4 A feasible most powerful instrument

The MPI is the most powerful instrument to reject $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$. Its derivation requires the knowledge of the alternative while in practice the econometrician typically wants to remain agnostic about the alternative. Moreover, the MPI is defined as a conditional expectation of a non-linear function with respect to a large dimension vector z_{jt}, and thus, even if the alternative $\overline{H}_a$ is known, the MPI can be difficult to compute. In this section, we remain in the same configuration, where the econometrician wants to test $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$. However now, we assume that this alternative is unknown to the econometrician. We provide two feasible approximations of the MPI, which do not depend on $\overline{H}_a$, and that, unlike the MPI, can be computed in practice. To do so, we show that the MPI can be approximated by a linear combination of known functions of z_{jt}. We call these interval instruments in reference to the way these functions are derived. Our feasible MPI is simply the vector of the interval instruments. The cost to incur for feasibility is that the properties we established for the MPI do not carry over to the feasible MPI. Nevertheless, our Monte Carlo simulations in Section 1.6 show that the interval instruments perform very well in practice.

[11] This interpretation of the optimal instruments only holds when the model is well specified i.e. $f \in \mathcal{F}_0$, and thus, in general, the optimal instruments shouldn't be used to test the specification of the model.

By construction, in the BLP demand model, the correction term writes:

$$\begin{aligned}\Delta_{0,a}^{\xi_{jt}} &= x_{1jt}'(\beta_a - \beta_0) + \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a) \\ &= x_{1jt}'(\beta_a - \beta_0) + \Delta_j(s_t, x_{2t}, f_0, f_a).\end{aligned} \tag{1.4.6}$$

The previous equation shows that the correction term is the sum of a linear part, which is standard, and a non-linear part which is specific to the BLP demand model.

Linear part. The linear part of the MPI writes: $\mathbb{E}[x_{1jt}|z_{jt}]'(\beta_a - \beta_0) = \mathbb{E}[x_{1jt}|z_{jt}]'\gamma$. Thus, for its linear part, the MPI is a linear combination of the conditional expectation of x_{1jt} with respect to the exogenous variables with unknown weights. If one is interested in specifically testing that $\beta = \beta_0$, informative instruments simply consist of the variables in $\mathbb{E}[x_{1jt}|z_{jt}]$.

Non-linear part. The non-linear part, $\Delta_j(s_t, x_{2t}, f_0, f_a)$, is the part that is implied by the misspecification on the distribution of RCs and for which we need to recover a feasible approximation. Equation (1.4.6) indicates that the non-linear part is the difference between the inverse demand functions generated by f_0 and f_a. We now go one step further and derive two analytical approximations of $\Delta_j(s_t, x_{2t}, f_0, f_a)$ which we then use as building blocks to construct our feasible approximations of the MPIs. The first approximation is based on a local expansion around f_0. The second approximation is based on an identity that is valid everywhere. The first approximation is more precise locally whereas the second one is more robust to large deviations from f_0.

1.4.1 Local approximation

First, we consider a local approximation of $\Delta_j(s_t, x_{2t}, f_0, f_a)$. This approximation corresponds to the first order term in the expansion of $\Delta(s_t, x_{2t}, f_0, f_a)$ "around f_0", which is recovered by exploiting the properties of the inverse demand function, which is both $\mathcal{C}^\infty$ and bijective in s_t.

Proposition 4.1
A first order expansion of $\Delta(s_t, x_{2t}, f_0, f_a)$ around f_0 writes:

$$\Delta(s_t, x_{2t}, f_0, f_a) = \left(\frac{\partial\rho(\delta_t^0, x_{2t}, f_0)}{\partial\delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \left[\frac{\exp\{\delta_t^0 + x_{2t}v\}}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2kt}v\right\}} - \rho(\delta_t^0, x_{2t}, f_0)\right] f_a(v) + \mathcal{R}_0,$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0)$ and $\mathcal{R}_0 = o\big(\int_{\mathbb{R}^{K_2}} |f_a(v) - f_0(v)|dv\big)$.

The proof is in Appendix 1.B.3. We first observe that for any density f_0, we can construct artificial market shares s_t^0 such that $\rho^{-1}(s_t, x_{2t}, f_a) = \rho^{-1}(s_t^0, x_{2t}, f_0)$. Then, we recover the final result by taking a Taylor expansion of $\rho^{-1}(s_t^0, x_{2t}, f_0)$ around s_t and showing that the remainder is bounded.[12] This approximation is local by design: it works best when f_a is a local deviation from f_0, even if it can be used more generally. To make this expression useful in practice, we must still overcome two difficulties. The distribution f_a is unknown to the econometrician. In addition, some variables such as δ_{jt}^0 are endogenous. However, notice that the previous expression may be particularly useful if the econometrician is interested in testing $\overline{H}_0$ against a fixed and known alternative as we did in the previous section.

Discretizing the integral. To solve for the fact f_a is unknown to the econometrician, we replace the integral in which f_a appears by a finite Riemann approximation. Namely,

$$\int_{\mathbb{R}} \frac{\exp\left\{x'_{2jt}v\right\}}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2kt}v\right\}} f_a(v)dv \approx \sum_{l=1}^L \omega_l(f_a) \frac{\exp\{x'_{2jt}v_l\}}{1 + \sum_{k=1}^J \exp\{\delta_{kt}^0 + x'_{2kt}v_l\}},$$

with $\{v_l\}_{l=1,\ldots,L}$ the points chosen in the domain of definition of f_a, and $\{\omega_l(f_a)\}_{l=1,\ldots,L}$ the associated weights.[13] We provide more details on how to choose the points in Appendix 1.C.4. It is important to observe that in the Riemann approximation, only the weights depend on the alternative f_a. This approximation can also be interpreted as approaching a continuous distribution with a discrete one, where

[12] The expansion is taken around s_t because s_t^0 depends on f_a and is thus unknown to the researcher.
[13] In the usual Riemann sum, the weights correspond to density evaluated at point $v_l : f_a(v_l)$ times the width of the interval around v_l.

each point in $\{v_l\}_{l=1,...,L}$ represents a specific consumer type with an associated probability $w_l(f_a)$. The non-linear part of the MPI can thus be approximated as follows:

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l(f_a)\, \mathbb{E}[\pi_{j,l}(s_t, x_t)|z_{jt}],$$
$$\text{with } \pi_{j,l}(s_t, x_t) = \left(\frac{\partial\rho(\delta_t^0, x_{2t}, f_0)}{\partial\delta}\right)^{-1} \left[\frac{\exp\{\delta_t^0 + x_{2t}v_l\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}^0 + x_{2kt}' v_l\right\}} - \rho(\delta_t^0, x_{2t}, f_0)\right]_j.$$

Approximating the conditional expectation. Ideally, we would like to estimate the conditional expectation of $\pi_{j,l}(s_t, x_t)$ with respect to z_{jt}. The endogenous variables are $\{\delta_{jt}^0\}_{j=1,....,J}$, and the potential endogenous variables in $\{x_{2jt}\}_{j=1,...,J}$, which often include prices. In practice, computing the conditional expectation is challenging because the dimension of z_{jt} can be very large and the functions $\pi_{j,l}(\cdot)$ are highly non-linear and non-separable in the endogenous variables. This makes it unappealing to use standard non-parametric estimation methods.[14] In the same spirit as Reynaert and Verboven (2014), we first project the endogenous variables on the space spanned by a relevant subset of z_{jt}. We mark the projected endogenous variables with a hat and we plug them into our functions $\pi_{j,l}(\cdot)$. Namely, we have the following approximation for every interval instrument l:

$$\mathbb{E}[\pi_{j,l}(s_t, x_t)|z_{jt}] \approx \hat{\pi}_{j,l}(z_{jt}) = \left(\frac{\partial\rho(\hat{\delta}_t^0, \hat{x}_{2t}, f_0)}{\partial\delta}\right)^{-1} \left[\frac{\exp\{\hat{\delta}_t^0 + \hat{x}_{2t}v_l\}}{1 + \sum_{k=1}^{J} \exp\left\{\hat{\delta}_{kt}^0 + \hat{x}_{2kt}' v_l\right\}} - \rho(\hat{\delta}_t^0, \hat{x}_{2t}, f_0)\right]_j.$$

We show in Appendix 1.C.2 that this strategy yields an estimator of the conditional expectation that converges faster to a first order approximation of the conditional expectation.

Test procedure. From what precedes, the MPI (for its non-linear part) can be approximated as follows: $h_D^*(z_{jt}) \approx \sum_{l=1}^{L} \omega_l(f_a)\ \hat{\pi}_{j,l}(z_{jt})$. As we don't know the weights $\omega_l(f_a)$, we propose to take

[14] For instance, a Sieve nonparametric estimator of the conditional mean. The dimension of z_{jt} makes this approach of little relevance in practice.

the vector $\hat{\pi}_j(z_{jt}) = (\hat{\pi}_{j,1}(z_{jt}), ..., \hat{\pi}_{j,L}(z_{jt}))'$ as our testing instruments. We call them interval instruments in reference to the way we divide the support into several intervals to construct this approximation. Following the test procedure presented in Section 1.3.1, we perform a moment based test for $\overline{H}_0 : \mathbb{E}\left[\hat{\pi}_j(z_{jt})\xi_{jt}(f_0, \beta_0)\right] = 0$. Under the same assumptions as in Proposition 3.1 and setting $h_D(z_{jt}) = \hat{\pi}_j(z_{jt})$, we have the following:

$$\text{Under } H_0 : S_T(h_D, f_0, \beta_0) \xrightarrow[T\to+\infty]{d} \chi^2_L.$$

This approach has the advantage of being feasible since we can construct the vector of interval instruments $\hat{\pi}_j(z_{jt})$, while remaining completely agnostic about f_a. The price to pay is that we lose the optimality properties of the MPI. We further discuss the properties of the feasible MPI in Appendix 1.C.7. Moreover, the infeasible MPI, $h^*_D(z_{jt})$, is of dimension one and its test statistic is distributed as χ^2_L asymptotically. In contrast, the feasible MPI is of dimension L and its asymptotic distribution is a χ^2_L. This increase in the number of degrees of freedom may lead to some loss of power. An alternative approach would consist in letting the researcher choose the weights $\{\hat{\omega}_l\}_{l=1,...,L}$ and recover an instrument of dimension one. However, for this approach to work well and retain good power properties, the econometrician must choose the weights so that they approximately match the real weights $\{w_l(f_a)\}_{l=1,...,L}$. This requires a good prior knowledge of the cumulative distribution function of the alternative distribution f_a. Nevertheless, our Monte Carlo simulations in Section 1.6 show that the feasible MPIs that we propose perform very well in practice.

1.4.2 Global approximation

Second, we consider a global approximation that is based on an identity which is valid everywhere and not only when f is close to f_a. Simple algebraic operations (see Appendix 1.B.3) allow us to derive the following expression for $\Delta_j(s_t, x_{2t}, f_0, f_a)$. Let $\delta^0_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_0)$ and $\delta^a_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_a)$. We have:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^a_{kt}+x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2kt}v\right\}} f_0(v)dv}\right).$$

As for the local approximation, we cannot directly exploit this formula as some quantities such as f_a and δ^a_{jt} are unknown and some variables such as δ^0_{jt} are endogenous. To remedy these two difficulties, we apply the same methods as previously described: we discretize the integral, and we project the endogenous variables onto the space spanned by a relevant subset of z_{jt}. To solve for the fact that the mean utility δ^a_{jt} under the alternative is unknown, we replace it with the mean utility under the null δ^0_{jt}. This should not alter the approximation too much given that δ^a_{jt} only enters the expression at the denominator within a sum, which averages out the differences between δ^a_{jt} and δ^0_{jt} across products. In the end, we are able to provide the following approximation for the non-linear part of the MPI:

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \log\left(\sum_{l=1}^{L} \bar{\omega}_l(f_a)\ \hat{\bar{\pi}}_{j,l}(z_{jt})\right) \text{ with } \hat{\bar{\pi}}_{j,l}(z_{jt}) = \frac{\frac{\exp\{x'_{2jt}v_l\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{kt}+x'_{2kt}v_l\right\}}}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{jt}+x'_{2kt}v\right\}} f_0(v)dv},$$

where $\{\bar{\omega}_l(f_a)\}_{l=1,\dots,L}$ correspond to the unknown weights and the $\hat{\bar{\pi}}_{j,l}(z_{jt})$ are set of global interval instruments. The MPI can thus be approximated by the logarithm of a weighted sum of known functions of z_{jt}. As we did previously, we use $\hat{\bar{\pi}}_j(z_{jt}) = (\hat{\bar{\pi}}_{j,1}(z_{jt}), \dots, \hat{\bar{\pi}}_{j,L}(z_{jt}))'$ as instruments to test $\overline{H}_0$. All the weights are positive and sum to one, which entails that the non-linear part of the correction term is an increasing function of our instruments. This approximation is said to be global because contrary to the first approximation we study, it does not require f_0 to be close to f_a. Nevertheless, if f_a is close to f_0, then the fraction κ inside the logarithm is close to 1 and the well-known approximation $\log(\kappa) \approx \kappa - 1$ allows us to directly rewrite the MPI as a linear combination of our instruments.

Overall, the feasible MPIs that we derive in this section allows us to approximate the most powerful instrument against a fixed alternative while remaining agnostic about this alternative.

1.4.3 Feasible MPIs for estimation

In the estimation framework, the researcher stipulates that f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ and wants to estimate the true parameter $\theta_0 = (\beta_0', \lambda_0')'$ under this parametric restriction. From the connection between the MPI and the local instruments that we present in Section 1.3.3, we can infer that good estimation instruments $h_E(z_{jt})$ ought to approximate the MPI devoted to testing $H_0 : \theta = \theta_0$ against any local alternative. If we have an initial estimator of θ_0, we can directly use the interval instruments presented previously to approximate the MPI devoted to testing $H_0 : \theta = \theta_0$ against an unknown alternative. The fact that the feasible MPIs do not depend on the alternative is key for estimation. Moreover, the transformation of the MPI into a vector of instruments of dimension $L \geq |\lambda_0|$ is necessary for estimation as the number of instruments must be greater than the dimension of the parameter to estimate.[15] In Appendix 1.C.5, we propose a version of the interval instruments that does not require a first step estimate of θ_0 and that can be computed directly from the logit specification.

1.5 Composite hypothesis

In the traditional estimation procedure, which encompasses almost all the applications of the BLP model, the econometrician must make a parametric assumption on the distribution of random coefficients to estimate the model. Formally, the econometrician assumes f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$, where $\tilde{\lambda}$ is a parameter that must be estimated. In applied work, researchers typically assume that f is normally distributed. This parametric choice is rarely grounded in economic theory and, if too restrictive, is likely to impose arbitrary restrictions on some key counterfactual quantities such as the pass-through. In this section, we develop a formal specification test for $H_0 : f \in \mathcal{F}_0$. In comparison to the test in Section 1.3.1, we must now estimate the parameters of the distribution $\theta_0 = (\beta_0', \lambda_0')'$ in a first step, which generates parameter uncertainty. Moreover, we propose a rigorous treatment of the numerical approximations involved in the derivation of the structural error $\xi_{jt}(\tilde{\theta})$. We organize this section as follows. First, we define the pseudo-true value associated with a given spec-

[15] The linear parameter β_0 has its own instruments, which are simply the variables in x_{1jt}.

ification and the first stage estimator. Second, we define our test procedure and its implementation in practice. Finally, we study the asymptotic properties of our test.

1.5.1 Pseudo-true value and first stage estimator

To estimate the BLP model, researchers must make three choices. They must choose the parametric family $\mathcal{F}_0$, the instruments $h_E(z_{jt})$ to estimate the model, and a weighting matrix W, which weights the different moments included in the objective function. Given these three choices, we can define the BLP pseudo-true value $\theta(\mathcal{F}_0, h_E, W) \equiv \theta_0 = (\beta_0', \lambda_0')'$ as follows:[16]

$$\theta(\mathcal{F}_0, h_E, W) \in \underset{\tilde{\theta}}{\text{Argmin}} \, \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})h_E(z_{jt})\right]' W \mathbb{E}\left[h_E(z_{jt})\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})\right].$$

If the model is well-specified ($f \in \mathcal{F}_0$) and the pseudo-true value is unique, then the pseudo-true value is the true value: $\theta_0 = \theta$. Under misspecification, θ_0 is a parameter whose value depends on $(\mathcal{F}_0, h_E, W)$. For exposition purposes, we omit this dependence in the subsequent analysis. Moreover, here we remain general and do not impose that W must be equal to the usual optimal weighting matrix. It is often the case in practice, that the researchers choose the identity matrix or regularize the weighting matrix.

First stage estimator $\hat{\theta}$. The first stage estimator is an empirical counterpart of the BLP pseudo-true value defined previously. The minimization is done with respect to sample analogs. Additionally, we know that there is no closed form expressions for the structural error $\xi_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$, and thus, we must use a feasible counterpart $\hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$ instead.

$$\hat{\theta}(\mathcal{F}_0, h_E, \hat{W}) \equiv \hat{\theta} = \underset{\tilde{\theta}}{\text{Argmin}} \left(\sum_{j,t} \hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})h_E(z_{jt})\right)' \hat{W} \left(\sum_{j,t} \hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})h_E(z_{jt})\right). \quad (1.5.7)$$

The construction of the feasible structural error $\hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$ requires the following 3 numerical approximations:

[16] Our definition of a pseudo-true value is closely related to the approach in White (1982) in the context of maximum likelihood. In his case, the pseudo true value minimizes the Kullback-Leibler distance between the assumed likelihood and the true likelihood, whereas in our case, the pseudo-true value minimizes a weighted sum of population moments.

1. The econometrician does not observe a continuum of consumers as in the theoretical model but only empirical averages $\hat{s}_{jt}$ over the n_t individuals in market t.

$$\hat{s}_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_{ijt}, \tag{1.5.8}$$

where $y_{ijt} \in \{0;1\}$ are i.i.d. choices over the $i = 1, \ldots, n_t$.

2. There is no closed form for $\rho_j(., x_{2t}, f_0(\cdot|\tilde{\lambda}))$, the integral has to be computed through numerical integration. A prominent example is Monte Carlo integration:

$$\hat{\rho}_j(\delta, x_{2t}, f_0(|\tilde{\lambda})) = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\{\delta_j + x'_{2jt} v_r\}}{1 + \sum_{k=1}^{J_t} \exp\{\delta_k + x'_{2kt} v_r\}}, \tag{1.5.9}$$

with v_r iid draws from $f_0(\cdot|\tilde{\lambda})$.

3. There is no analytical way to recover the inverse of the demand functions $\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$. The most popular way to derive the inverse demand is by solving the following contraction:

$$C : (\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) : \delta \mapsto \delta + \log(s_t) - \log(\rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda}))).$$

This solution has given rise to the popular nested fixed point GMM procedure.[17]

In Section 1.5.3, we explicitly state the assumptions that allow us to neglect these approximations asymptotically.

1.5.2 Test procedure

Under Assumption A, and assuming $h_E(z_{jt})$ and W are such that the pseudo-true value θ_0 is unique, the following equivalence holds:

$$\begin{aligned} H_0 : f \in \mathcal{F}_0 &\iff \overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0) \\ &\iff \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)|z_{jt}] = 0 \;\; a.s.. \end{aligned}$$

[17] Another solution that has gained traction in the literature is the MPEC procedure (Dubé et al. (2012)) that replaces the BLP inversion at each step of the minimization by imposing equilibrium constraints on the minimization program.

The pseudo true value reduces the dimensionality of the problem by allowing us to move from a composite hypothesis $H_0 : f \in \mathcal{F}_0$ to the simple hypothesis $\overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0)$ studied previously. As we did in Section 1.2, we propose a moment-based test of H_0.[18] Under H_0, for every set of testing instruments $h_D(z_{jt})$, the following moment conditions must hold:

$$H_0 : f \in \mathcal{F}_0 \iff \overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0) \implies \overline{H}'_0 : \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt})\right] = 0.$$

We now develop a procedure to test $\overline{H}'_0$. In comparison to the test in Section 1.3.1, we must now account for the fact that the pseudo-true value needs to be estimated to derive the test statistic, which generates parameter uncertainty. Moreover, we propose a rigorous treatment of the numerical approximations involved in the derivation of the structural error.

Test statistic. For any choice of testing instruments $h_D(z_{jt})$, our objective is to test $\overline{H}'_0 :\ \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt}$ 0 where $\theta_0 = (\beta'_0, \lambda'_0)'$ is the pseudo-true value associated with the parametric family $\mathcal{F}_0$.[19] In order to test $\overline{H}_0$, we consider the following Wald test statistic:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) = TJ \left(\frac{1}{TJ} \sum_{j,t} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \right)' \hat{\Sigma} \left(\frac{1}{TJ} \sum_{j,t} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \right).$$

where $\hat{\Sigma}$ is a weighting matrix chosen by the econometrician and $\hat{\theta} = (\hat{\beta}, \hat{\lambda})$ is a consistent estimator of θ_0. The number of markets T is the dimension that we let grow to infinity to the asymptotic properties

[18] Other testing approaches could have been considered. First, one could use the previous equivalence to directly test H_0 via an integrated conditional moment test. We do not follow this route for at least two reasons. First, this test will contain no information on the nature of the misspecification (it could be completely unrelated to the distribution of RC). Second, in practice the dimension of z_{jt} is often very large, which substantially reduces the power of this kind of test. Another testing approach would have entailed testing $H_0 : f \in \mathcal{F}_0$ against a larger class of densities that encompasses $\mathcal{F}_0$. For instance, if $\mathcal{F}_0$ is the family of normal distributions, encompassing families are mixtures of Gaussians with a larger number of components. We do not follow this route for two reasons. First, it is not desirable to restrict the alternative to a class of distributions that encompass the null as the econometrician does not know a priori the misspecification. Second, estimating the BLP model with a more flexible parametrization is challenging. An advantage of our test procedure is that it doesn't require estimating the model with a more flexible parametrization.

[19] Remember that under an alternative specification, the pseudo true value also depends on the estimation instruments $h_E(z_{jt})$ and the weighting matrix.

of our test. We motivate this choice in Appendix 1.C.3. Under some regularity conditions that we make explicit in the following section, the asymptotic distribution of the test statistic under $\overline{H}_0'$ is as follows:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z'\Sigma Z, \tag{1.5.10}$$

$$\text{with } \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{J} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \xrightarrow{d} Z \sim \mathcal{N}(0, \tilde{\Omega}_0). \tag{1.5.11}$$

Σ is the probability limit of $\hat{\Sigma}$. We make $\tilde{\Omega}_0$ explicit in the next subsection (in particular, the derivation of $\tilde{\Omega}_0$ takes into account parameter uncertainty). Given that $\hat{\Sigma}$ is chosen by the econometrician and it is possible to derive a consistent estimator of $\tilde{\Omega}_0$, the econometrician can always simulate the asymptotic distribution of the test statistic. In some polar cases, which we present hereafter, the asymptotic distribution of our test statistic is pivotal chi-square distribution that does not require to be simulated.

Two polar cases. For the sake of exposition, let us now describe two polar cases where the asymptotic distributions are pivotal chi-square distributions, which do not require to be simulated. Denote by $|\cdot|_0$ the counting norm.

1. **Sargan-Hansen J test:** If the set of estimation instruments and the set of testing instruments are the same ($h_E = h_D$), if $\hat{W}$ is the 2-step GMM optimal weighting matrix and if $\hat{\Sigma} = \hat{W}^{-1}$, then our test boils down to the usual Sargan-Hansen J test and we have under $\overline{H}_0'$:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_E|_0 - |\theta|_0}.$$

2. **Non-redundant h_D and h_E:** if $\tilde{\Omega}_0$ has full rank and if the econometrician sets $\hat{\Sigma} = \hat{\tilde{\Omega}}_0^{-1}$, then our test statistic has the following asymptotic distribution under $\overline{H}_0'$:[20]

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_D|_0}.$$

[20] If Ω_0 is singular, one can always use directly the asymptotic distribution in 1.5.10 or apply the singularity-robust procedure proposed in Andrews and Guggenberger (2019).

Choice of the testing instruments. As previously indicated, the power properties of our test hinge critically on the choice of testing instruments $h_D(z_{jt})$. We established that the MPI and its feasible counterparts, the interval instruments, feature attractive properties in testing $\overline{H}_0 : (f,\beta) = (f_0(\cdot|\lambda_0), \beta_0)$ against any fixed alternative. Thus, it is natural to use these instruments for the specification test above. In particular, we show that the consistency of the test with the MPI carries over to the general specification test above in Appendix 1.B.5.

1.5.3 Asymptotic validity

We now study the asymptotic properties of our test when the number of markets T goes to infinity. To establish the asymptotic validity and consistency of our test, we exploit classical results on the asymptotic normality of the non-linear GMM estimator (Hansen (1982), Newey (1990)) as well on the large-T asymptotics of the BLP estimator (Freyberger (2015)). The main challenge here is to control the magnitude of the approximations that intervene in the derivation of the structural error so that they can be neglected asymptotically. Contrary to Freyberger (2015), we do not assume the convergence of any moments ex-ante and we allow for the approximation error between demand and observed market shares to be non-zero.

Assumption B
(i) $(s_t, x_t, z_t)_{t=1}^T$ *are i.i.d. across markets and are consistent with the probability model defined by equations (1.2.1), (1.2.2) and (1.2.3) evaluated at* (f,β)*;*
(ii) Strong Exogeneity: $\mathbb{E}[\xi_{jt}(f,\beta)|z_{jt}] = 0$ *a.s.;*
(iii) Finite moment conditions: x_{2t} *has bounded support and* x_{1t} *has finite 4th moments.*

In B(i), we assume that the data are i.i.d. across markets, an assumption which we could relax slightly (technically, only certain moments need to be identical across markets), and that the data are generated by the BLP demand model at a given pair (f,β). In B(ii), we assume exogeneity of our instrumental variables. Let us stress that to show the asymptotic validity of our specification test, we do not require (f, β) to be non-parametrically identified, as we just need parametric identification under H_0. In particular, we do not need all the assumptions in A. B(iii) is a necessary condition to recover the asymptotic normality of the BLP estimator. x_{1t} having finite 4th moments is standard. x_{2t} having bounded support has two purposes. First, it implies that the structural error has a finite 4th moment, Compiani (2018) makes the same assumption on price for this purpose. Second, it ensures that the mapping used in the nested fixed point algorithm is a proper smooth contraction, which allows us to prove that the NFP algorithm converges (without truncating the contraction mapping as in Berry (1994) and Berry et al. (1995)) and control for the NFP approximation bias.

Assumption C

$\mathcal{F}_0$ is such that :

(i) λ_0 belongs to the interior of Λ_0 with Λ_0 compact;

(ii) $\tilde{\lambda} \mapsto \rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ is well defined and continuously differentiable on Λ_0.

In C(i), we assume that, for any given DGP, the associated pseudo-true-value λ_0 associated with the family $\mathcal{F}_0$ lies in a compact space Λ_0. This condition is standard in establishing the consistency and asymptotic normality of M-estimators. Second in C(ii), we impose that the demand function and its derivative with respect to λ should both be well defined and continuous.

Next, we impose conditions on the instruments that are used for estimation $h_E(z_{jt})$ and for testing $h_D(z_{jt})$ and on the BLP estimator itself.

Assumption D

For a given $\mathcal{F}_0$ that satisfies Assumption C and for some weighting matrix W and Σ, the following conditions must hold:

(i) Finite moments for instruments: $h_E(z_{jt})$ and $h_D(z_{jt})$ are not perfectly colinear and have finite 4th moments;

(ii) Global identification of θ_0: $\exists! \theta_0$ such that $\forall \tilde{\theta} \neq \theta_0$:

$$\mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})h_E(z_{jt})'\right] W\mathbb{E}\left[\sum_j h_E(z_{jt})\xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})\right] > \mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)h_E(z_{jt})'\right] W\mathbb{E}\left[\sum_j h_E(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right];$$

(iii) Local identification: $\Gamma(\mathcal{F}_0,\theta_0,h_E) = \mathbb{E}\left[\sum_j h_E(z_{jt})\frac{\partial\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)}{\partial\theta'}\right]$ *and* $\Gamma(\mathcal{F}_0,\theta_0,h_D)$ *have full column rank;*

(iv) W *and* Σ *are symmetric positive definite and* $\hat{W} \xrightarrow{\mathbb{P}} W$, $\hat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$*;*

(v) $\hat{\theta}$ *minimizes objective function (1.5.7) and satisfies the FOC of the minimization problem:*

$$\left(\sum_{j,t}\frac{\partial\hat{\xi}_{jt}(f(\cdot|\hat{\lambda}),\hat{\beta})}{\partial\theta}h_E(z_{jt})\right)'\hat{W}\left(\sum_{j,t}\hat{\xi}_{jt}(f(\cdot|\hat{\lambda}),\hat{\beta})h_E(z_{jt})\right) = 0.$$

Assumption D restricts the class of instruments which can be used for estimation and for testing. More specifically, D(i) and D(iii) are common regularity conditions necessary to establish asymptotic results whereas D(ii) is an identification condition which ensures that the pseudo true value θ_0 is uniquely defined, which is critical to show the consistency of the BLP estimator. Finally, Assumptions D(iv) and D(v) impose regularity conditions on the weighting matrix as well as on the BLP estimator itself.

The next assumptions ensure that the numerical approximations involved in the derivation of the structural error do not interfere with the asymptotic theory.

Assumption E

(i) Let n_t *be the number of individuals in market* t, $(n_t)_{t=1}^T$ *is i.i.d. and independent from all other variables. First it must be that* $\forall t$ $\sqrt{T}\mathbb{E}(n_t^{-1/2}) \underset{T\to+\infty}{\to} 0$*. Second observed market share* $\hat{s}_t$ *in market* t *must write:*

$$\hat{s}_{jt} = \frac{1}{n_t}\sum_{i=1}^{n_t} y_{ijt},$$

with $(y_{ijt})_{i=1}^{n_t}$ *i.i.d. draws generated by the BLP demand model at a given pair* (f,β) *conditional on* (x_t,ξ_t)*.*

(ii) Let R *be the number of simulations, then the simulated demand for product* j *writes:*

$$\hat{\rho}_{jt}(\delta,x_{2t},f_0(\cdot|\tilde{\lambda})) = \frac{1}{R}\sum_r \frac{\exp\{\delta_j + x'_{2jt}v_r\}}{1+\sum_k \exp\{\delta_k + x'_{2kt}v_r\}},$$

where $v_r \overset{iid}{\sim} f_0(\cdot|\tilde{\lambda})$, *and* $\frac{T}{R} \underset{T\to+\infty}{\to} 0$.

(iii) Let H *be the stopping time for the contraction (which depends on* T*) and* ϵ *the fixed Lipschitz constant of the contraction mapping used to invert the demand function, then it must be that* $\sqrt{T}\epsilon^H \underset{T\to+\infty}{\to} 0$.

A sufficient condition for E(i) to hold is that the minimum number of individuals observed in any market is of higher order than the total number of markets. This condition can be checked in practice.[21] Assumptions E(ii) and E(iii) can also be checked in practice and are more manageable because R and H are chosen by the researcher and can always be increased so that these assumptions hold.

Given our assumptions, we derive the asymptotic distribution of our test statistic under the null, and show that the test is consistent.

Theorem 5.1 *Let* $\hat{\theta} = \hat{\theta}(\mathcal{F}_0, \hat{W}, h_E)$ *be the BLP estimator associated with distributional assumption* $\mathcal{F}_0$*, weighting matrix* $\hat{W}$*, estimating instruments* h_E*. Under assumptions B-E,*

- *Under* $\overline{H}_0' : \ \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt})\right] = 0,$

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \underset{T\to+\infty}{\overset{d}{\to}} Z'\Sigma Z, \qquad Z \sim \mathcal{N}(0, \tilde{\Omega}_0),$$

$$\textit{where} \quad \tilde{\Omega}_0 = \begin{pmatrix} I_{|h_D|_0} & G \end{pmatrix} \begin{pmatrix} \Omega(\mathcal{F}_0, h_D) & \Omega(\mathcal{F}_0, h_D, h_E) \\ \Omega(\mathcal{F}_0, h_D, h_E)' & \Omega(\mathcal{F}_0, h_E) \end{pmatrix} \begin{pmatrix} I_{|h_D|_0} \\ G' \end{pmatrix},$$

$$\Omega(\mathcal{F}_0, h_D, h_E) = cov\left(\sum_j \xi_{jt}(f(.|\lambda_0), \beta_0) h_D(z_{jt}), \sum_j \xi_{jt}(f(.|\lambda_0), \beta_0) h_E(z_{jt}) \right),$$

$$G = -\Gamma(\mathcal{F}_0, \theta_0, h_D) \left[\Gamma(\mathcal{F}_0, \theta_0, h_E)' W \Gamma(\mathcal{F}_0, \theta_0, h_E)\right]^{-1} \Gamma(\mathcal{F}_0, \theta_0, h_E)' W.$$

- *Under* $H_a' : \ \mathbb{E}\left[h_D(z_{jt}) \xi_{jt}(f_0(.|\lambda_0), \beta_0)\right] \neq 0,$

$$\forall q \in \mathbb{R}^+, \ \mathbb{P}(S_T(h_D, \mathcal{F}_0, \hat{\theta}) > q) \underset{T\to+\infty}{\to} 1.$$

[21] Note that by making stronger assumptions on the higher moments and the support of the observed characteristics, it is possible to find milder conditions on the number of individuals relative to the number of markets.

The proof of Theorem 5.1 is in Appendix 1.B.4 and comprises three main steps. First, we show that under the assumptions in E, the numerical approximation becomes asymptotically negligible. Second, we show the consistency and asymptotic normality of the BLP estimator. Finally, we derive the asymptotic distribution of the test statistic, taking into account parameter uncertainty (θ_0 is estimated and not observed). The apparent complexity of the asymptotic variance-covariance matrix Ω_0 is a direct consequence of parameter uncertainty.

1.6 Monte Carlo experiments

In this section, we conduct three distinct sets of Monte Carlo experiments. First, we implement a simple simulation exercise to assess the effects of incorrectly specifying the distribution of random coefficients on quantities of interest such as price elasticities or cross-price elasticities, which are known to play a key role in shaping the counterfactuals. In a second set of Monte Carlo experiments, we study the finite sample performances of the specification test developed in Section 1.5 with different sets of testing instruments. We first examine the size of our test in finite sample. Then, we investigate the power properties of our test under alternative specifications (with alternatives including Gaussian mixtures, gamma distributions and local alternatives). We show that our test with the interval instruments significantly outperforms the traditional J-test with the usual instruments. Finally, in the last Monte Carlo exercise, we study the performance of the interval instruments to estimate the parameters of the model by means of comparison with the commonly used instruments in the literature.

1.6.1 Simulation design

For the sake of exposition, we will keep the same simulation design for all the simulation experiments. The simulation design closely follows the simulation design used in Dubé et al. (2012), Reynaert and Verboven (2014). The market includes $J = 12$ products, which are characterized by 3 exogenous product attributes x_a, x_b and x_c that follow a joint normal distribution. The price p is endogenous and depends on the observed and unobserved characteristics and on some cost shifters c_1 and c_2. Consumer heterogeneity is present only in x_c, and the random coefficient v_i associated with x_c follows various distributions depending on the simulation exercise. The sample size T varies between 50, 100 and 200 markets. We can summarize the DGP as follows:

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt} \qquad \xi_{jt} \sim \mathcal{N}(0,1),\ \varepsilon_{ijt} \sim EV1,$$

$$\text{and} \begin{bmatrix} x_{a,j} \\ x_{b,j} \\ x_{c,j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix} \right),$$

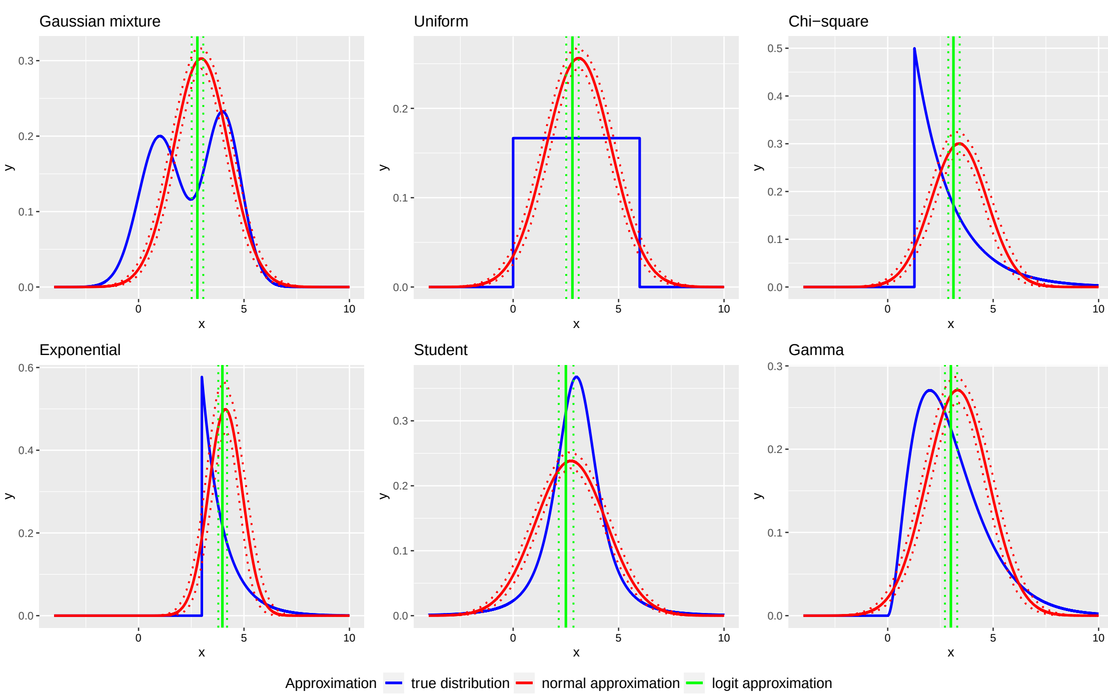
$$p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt} \quad \text{with } u_{j,t} \sim U[-4,-2],\ c_{1jt} \sim U[2,4] \ and\ c_{2jt} \sim U[3,5].$$

Market shares are generated by integrating over $20,000$ consumers. This allows us to essentially remove the approximation error between the observed and theoretical market shares.

1.6.2 Counterfactuals under an alternative distribution

We now present a simple exercise to illustrate how the misspecification of random coefficients can affect the estimation of quantities of interest such as price elasticities and cross-price elasticities. To do so, we simulate data using the simulation design introduced above and we take various distributions for the random coefficient v_i (respectively: Gaussian mixture, Uniform, Chi-square, Exponential, Student, Gamma). We ensure that all the distributions have the same mean and variance (3 and 3, respectively). For each distribution, we simulate $T = 100$ markets of data and we estimate the model either assuming no heterogeneity (simple logit) or assuming that v_i is normally distributed. We replicate the same exercise 500 times for each distribution. This allows us to recover the mean estimate for the parameters as well as to construct 95% "confidence bins" (by trimming the observations below the 2.5% quantile and above the 97.5% quantile). We plot the true densities and their estimated counterparts under the normal and logit assumptions in Figure 1.1. We observe that the estimated logit parameters and the estimated means of the normal distributions always coincide and are close to 3 for all the distributions. However, there is some variation between the different specifications. For instance, the estimated means are larger with the exponential distribution. The estimated variances also vary from one specification to the other. The estimated variances for the exponential distribution are smaller, while they are larger for the student distribution.

Figure 1.1: True densities and estimated densities under normal and logit specifications



In a second stage, we simulate $N = 5,000$ draws from the true distributions as well as from the estimated logit and normal approximations to compute the demand, the price-elasticity and the cross-price elasticity for the product j^* with the highest value for x_c.[22] The cross-price elasticity is arbitrarily taken for product $j = 1$ with respect to p_{j^*}. We derive the quantities of interest for 100 equally spaced values of p_{j^*} ranging in $]0, 10[$. We plot the elasticities in Figure 1.2 and cross-price elasticities in Figure

[22]The expressions for both price-elasticities and the cross-price elasticities are in Appendix 1.D.1.

1.3 generated by the true distribution as well as those generated by the logit and normal approximations, respectively. We proceed similarly with the demand functions. We see in Figure 1.9 in Appendix).

One can observe that, as expected, the logit specification poorly replicates the substitution patterns. In particular, it consistently overstates the magnitude of the elasticities and cross-elasticities with respect to the true ones. The absence of consumer heterogeneity on characteristic c implies that consumers can "renounce' more easily to product j^* when its price increases. By introducing some heterogeneity, the normal approximation somewhat attenuates this issue. However, significant discrepancies in the shape of elasticities and cross-price elasticities remain. As most counterfactual analyzes rely on the substitution patterns generated by the model, these differences will inevitably create significant biases.

Figure 1.2: Price elasticities

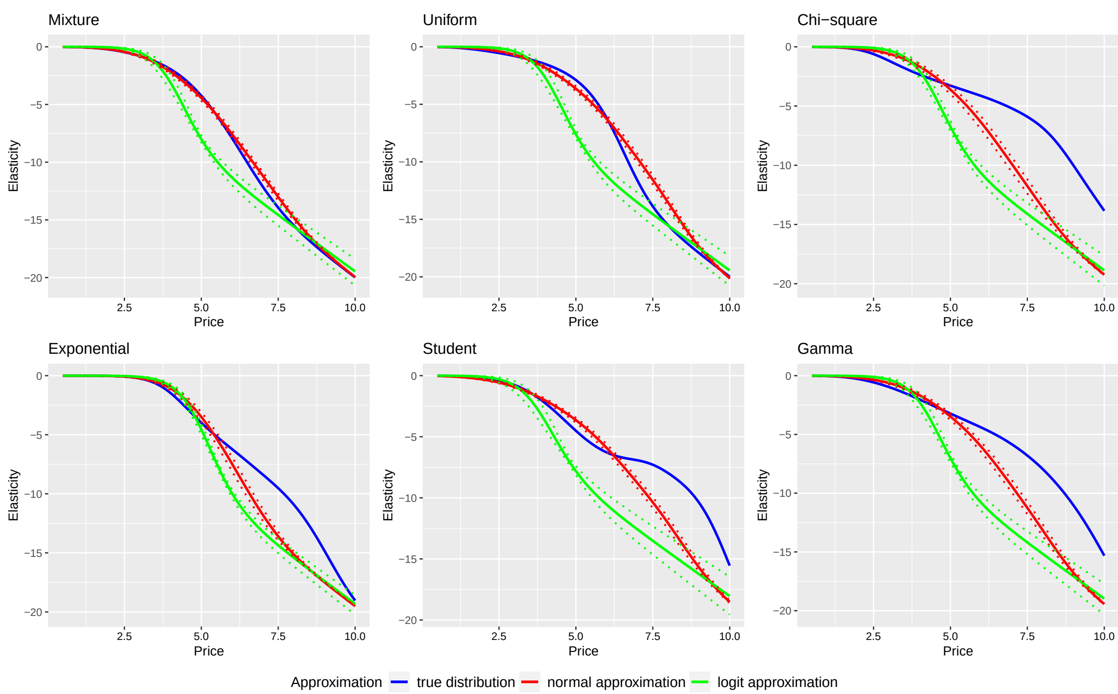
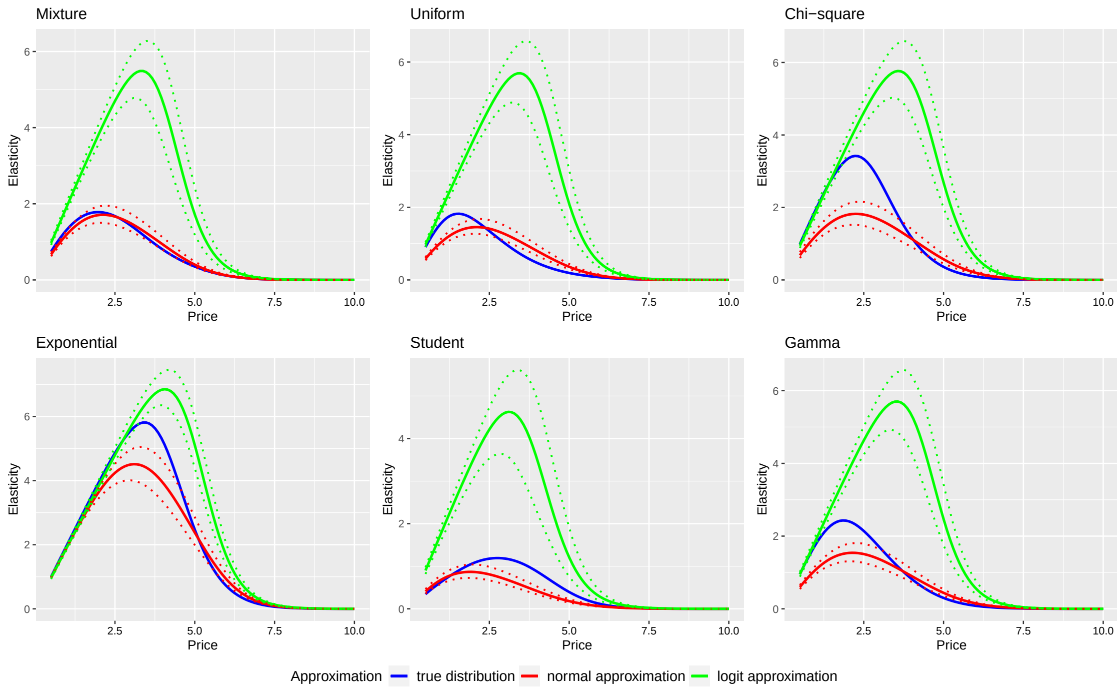


Figure 1.3: Cross-price elasticities



1.6.3 Finite sample performance of the specification test

We now study the empirical size and power of our test under different sample sizes and for different sets of testing and estimating instruments. Once again, the data are generated according to the simulation design exhibited previously for various distributions of v_i. The assumption made throughout the simulations is $H_0 : f \in \mathcal{F}_0$, where $\mathcal{F}_0$ is the family of normal distributions. In other words, we always assume that the random coefficient is normally distributed and we test this hypothesis. We set the nominal size to 5%. We study the finite sample performances of the specification test that we presented in Section 1.5 using different sets of estimation and testing instruments. For estimation, we take the instruments commonly adopted by practitioners: the differentiation instruments of Gandhi and Houde (2019) and the ”optimal” instruments of Reynaert and Verboven (2014). Both of these sets are approximations of the classical optimal instruments. Second, we compare the performance of the test when performing the standard Sargan-Hansen J test (i.e. when we use the same instruments for testing and estimation) and when we use the global and local approximations of the MPI that we constructed in Sections 1.4.2 and 1.4.1. We denote the latter tests as I Local and I Global respectively. The BLP estimator is computed following the NFP GMM procedure described in Section 1.5.1. For the optimization, only an analytic Jacobian is provided. We ensure that the number of tested restrictions is of the same magnitude across the different sets of instruments. More details on the exact sets of instruments and on the estimation procedure for this specific set of simulations are given in Appendix 1.D.2.

Empirical size

The size is the probability of rejecting the null hypothesis when the null is true, so we compute the empirical size by counting and averaging the number of times we reject the null for nominal size 5% over the $1,000$ simulations when the random coefficient v_i is normally distributed. Below in Table 1.1, we report the empirical sizes of the test with the different sets of instruments described above for the different sample sizes $T \in \{50, 100, 200\}$ and for different distributions of the RC such that $v_i \sim f \in \mathcal{F}_0$.

Table 1.1: Empirical size for nominal size 5% (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local
$v_i \sim \mathcal{N}(-1, 0.5^2)$	0.294	0.083	0.091	0.145	0.078	0.063	0.138	0.078	0.058	0.094	0.084	0.047	0.08	0.052	0.053	0.064	0.05	0.04
$v_i \sim \mathcal{N}(0, 0.75^2)$	0.293	0.084	0.085	0.148	0.081	0.071	0.137	0.061	0.06	0.1	0.059	0.05	0.074	0.053	0.045	0.062	0.048	0.036
$v_i \sim \mathcal{N}(1, 1^2)$	0.287	0.084	0.083	0.142	0.084	0.073	0.142	0.055	0.054	0.098	0.053	0.047	0.079	0.042	0.03	0.058	0.035	0.025
$v_i \sim \mathcal{N}(2, 2^2)$	0.288	0.087	0.077	0.145	0.071	0.072	0.138	0.069	0.051	0.099	0.053	0.056	0.077	0.044	0.041	0.069	0.037	0.044
$v_i \sim \mathcal{N}(3, 3^2)$	0.287	0.089	0.071	0.137	0.075	0.066	0.145	0.074	0.06	0.098	0.06	0.061	0.076	0.044	0.037	0.061	0.046	0.046

We observe that with a moderate sample size ($T = 50, J = 12$), all the tests are over-sized. This is within expectations and due to the approximations inherent to the estimation of the BLP models as described in Section 1.5 and the relatively large number of instruments used for estimation and testing purposes.[23] However, we notice that the Sargan-Hansen J tests are much more over-sized than the

[23] The number of over-identifying restrictions lies between 6 and 8. The Sargan-Hansen J tests are known to suffer from size distortions as the number of instruments increases.

tests with the interval instruments: the rejection rate is above 25% for the Sargan-Hansen J test with differentiation instruments vs around 8% for the I test. Increasing the sample size improves the tests' empirical levels and shifts them towards the nominal level, which is a good indication of the asymptotic validity of our test. Even with a relatively large number of markets ($T = 200$), the Sargan-Hansen J tests remain slightly oversized (rejection rate is still slightly above 5%). On the other hand, for the test with interval instruments, the empirical size appears to match the nominal level for all but two configurations, where it seems to be slightly undersized.

Empirical power

Power is the probability of rejecting the null hypothesis under an alternative. We compute the empirical power by counting and averaging the number of times we reject the null for the test of nominal size 5% over the 1000 simulations when the distribution of random coefficients is misspecified. The simulation setup remains the same as previously with the only modification being that the true distribution of v_i is now either a mixture of normals or a Gamma. We report the power against the different alternatives in the subsequent tables. The main takeaway from our results is that the test with the interval instruments as testing instruments (I global and I local) largely outperforms the traditional Sargan-Hansen J-test against all the alternative distributions considered in our simulations.

Power against Gaussian mixture alternatives. We simulate data with the random coefficients distributed according to the Gaussian mixtures described below. We plot the true distributions in Figure 1.4. We report the results in Table 1.2. We observe that the test with the interval instruments has great power against all the mixtures tested. The rejection rates go to 1 very quickly in comparison to the Sargan-Hansen J tests.

$$
\begin{aligned}
&v = Dv_1 + (1 - D)v_2, \quad \mathbb{P}(D = 1) = p, \quad \mathbb{P}(D = 0) = 1 - p, \\
&v_1 \sim \mathcal{N}\left(-\sqrt{\frac{3p}{1-p}} + 2, 1\right) \quad v_2 \sim \mathcal{N}\left(\sqrt{\frac{3(1-p)}{p}} + 2, 1\right),
\end{aligned}
$$

with $p \in \{0.1; 0.2; 0.3; 0.4; 0.5\}$.

Figure 1.4: Densities, Gaussian mixture alternatives

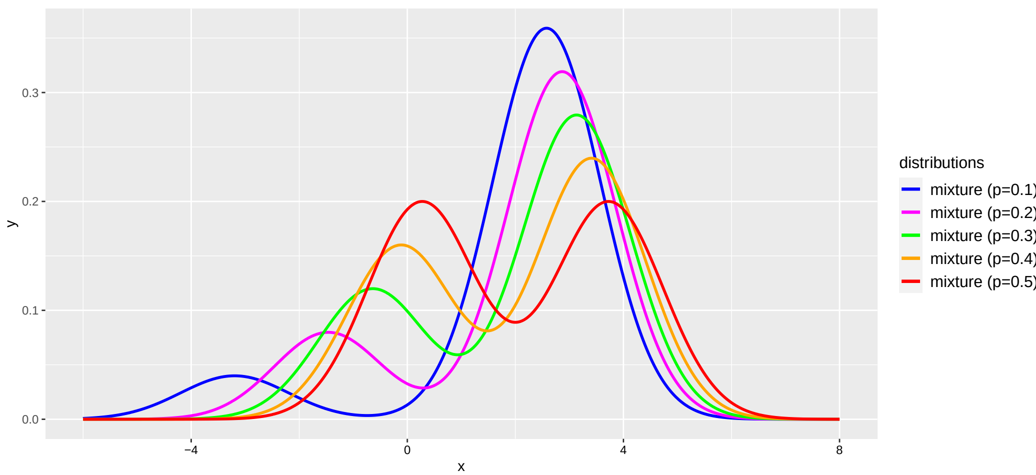


Table 1.2: Empirical power, Gaussian mixture alternatives (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local
Mixture 1	0.533	0.991	0.987	0.719	0.989	0.989	0.604	1	1	0.967	1	1	0.829	1	1	1	1	1
Mixture 2	0.626	0.996	0.998	0.613	0.997	0.998	0.723	1	1	0.905	1	1	0.933	1	1	1	1	1
Mixture 3	0.629	0.992	0.995	0.43	0.996	0.997	0.741	1	1	0.7	1	1	0.941	1	1	0.977	1	1
Mixture 4	0.601	0.983	0.982	0.275	0.981	0.981	0.713	1	0.999	0.368	1	1	0.921	1	1	0.672	1	1
Mixture 5	0.56	0.907	0.904	0.157	0.9	0.906	0.635	0.993	0.995	0.124	0.995	0.996	0.855	1	1	0.146	1	1

Power against Gamma alternatives. We simulate data with the random coefficients distributed according to the Gamma distribution described below. We plot the true distributions in Figure 1.5. We report the results in table 1.3. We observe that the test with interval instruments has great power against all the Gamma distributions tested except for the first one, which we can see on the plot has a distribution that is relatively close to a normal distribution. Even for the first Gamma distribution, it still outperforms the traditional sets of instruments. For all the other Gamma distributions, the rejection rates go to 1 very quickly in comparison to the Sargan-Hansen J-tests. This confirms the superiority of the interval instruments in detecting misspecification in the distribution of random coefficients. In Appendix 1.D.2, we also study the power properties of our test against local alternatives.

$$v \sim \Gamma(2, k) \quad with\ k \in \{0.25, 0.5, 0.75, 1, 1.5\}$$

Figure 1.5: Densities, Gamma alternatives

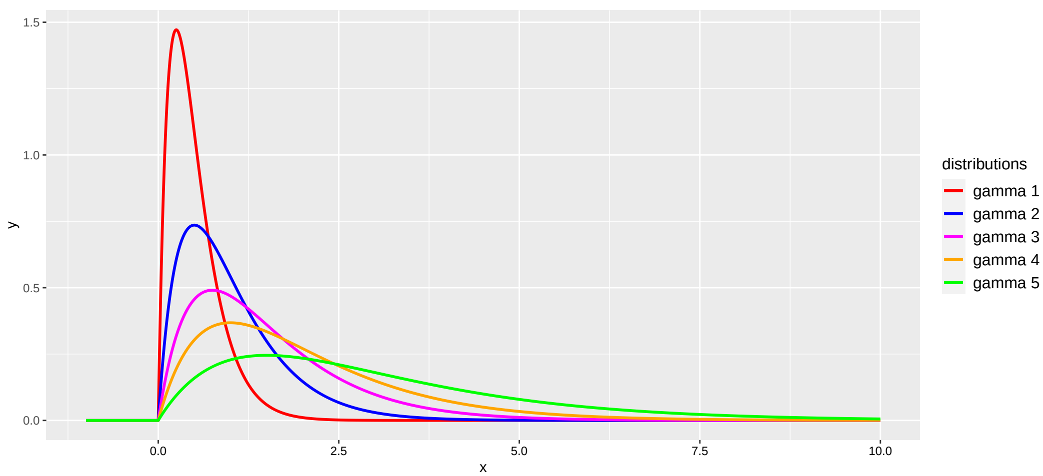


Table 1.3: Empirical power, Gamma alternatives (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local
Gamma 1	0.293	0.106	0.093	0.142	0.082	0.074	0.154	0.083	0.073	0.094	0.092	0.08	0.118	0.155	0.139	0.066	0.156	0.138
Gamma 2	0.516	0.747	0.752	0.14	0.781	0.77	0.562	0.983	0.978	0.095	0.982	0.98	0.492	1	1	0.08	1	1
Gamma 3	0.607	0.96	0.962	0.157	0.963	0.969	0.693	0.998	1	0.156	1	1	0.922	1	1	0.161	1	1
Gamma 4	0.622	0.97	0.99	0.207	0.962	0.995	0.748	0.999	1	0.263	1	1	0.933	1	1	0.412	1	1
Gamma 5	0.687	0.991	0.999	0.371	0.988	0.999	0.812	1	1	0.585	1	1	0.976	1	1	0.865	1	1

1.6.4 Finite sample performance of interval instruments for estimation

In our last simulation exercise, we evaluate the performance of our interval instruments in estimating the parameters associated with the RC when the distribution of random coefficients is flexibly parametrized.

To do so, we simulate data with a distribution of random coefficients following a mixture of Gaussians and we estimate the parameters of this mixture. We provide a method to estimate the parameters when the distribution of the RC is a mixture in Section 1.C.6 of the Appendix. In particular, we provide a new parametrization of the model, which yields substantial practical gains and may be of interest to researchers independent of the rest of the paper. The simulation design remains the same as previously. We assume that the random coefficient v_i is distributed according to the following mixture: $v_i \sim D_i\, \mathcal{N}(-2, 0.5) + (1 - D_i)\, \mathcal{N}(4, 0.5)$ with $\mathbb{P}(D_i = 1) = 0.25$. Thus, there are 5 parameters associated with the distribution of RC: the means and variances of each component of the mixture and the mixing probability. Our objective is to compare the performance of the global and local interval instruments with the instruments commonly used by practitioners: the differentiation instruments from Gandhi and Houde (2019) and the "optimal instruments" from Reynaert and Verboven (2014). In Table 1.4, we report the empirical biases and the square root of the MSE for the estimators of the non-linear parameters for each set of instruments and for the different sample sizes. In Appendix 1.D.3, we report the same information for the linear parameters (see Tables 1.14, 1.15, and 1.17) as well as the distribution of the empirical distribution of the non-linear estimates. Table 1.4 allows us to directly compare the performances of the three sets of instruments in estimating the non-linear parameters. We first observe that for all the sets of instruments, the empirical biases and $\sqrt{MSE}$ of the estimators decrease when the sample size increases, which is reassuring. Furthermore, it appears clearly that the differentiation instruments perform worse than the "optimal instruments" and the interval instruments. The empirical $\sqrt{MSE}$ of the estimators with the differentiation instruments is up to 12 times larger than with the interval instruments and up to 6 times larger than with the "optimal instruments". We reach the same conclusions when we study empirical biases. The interval instruments appear to perform better than the "optimal instruments" even if the difference is less significant than with the differentiation instruments. For the sake of conciseness, we do not report the results obtained with a mixture of 3 components but the observations we make with two components are even more exacerbated. In Appendix 1.D.3, as a means of comparison, we perform the same exercise when the distribution of random coefficients is a simple Gaussian and here, we do not observe any significant differences between the different sets of instruments, which confirms that the interval instruments make a difference when the distribution of RCs is flexible.

Table 1.4: Estimation non-linear parameters of the mixture (1000 replications)

	Instruments	Differentiation					"Optimal"					Interval Global					Interval Local				
	Parameter	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25
T=50, J=12	bias	0.214	0.184	-0.022	-0.045	0.027	0.076	0.059	0.026	-0.111	0.01	0.017	0	-0.045	0.004	0.005	-0.006	-0.005	-0.039	-0.001	0.003
	$\sqrt{MSE}$	0.633	0.734	0.281	0.35	0.075	0.361	0.483	0.212	0.281	0.036	0.277	0.391	0.227	0.259	0.024	0.251	0.34	0.214	0.244	0.019
T=50, J=20	bias	0.189	0.347	0.022	-0.081	0.025	0.074	0.11	0.028	-0.089	0.01	0.013	0.042	-0.018	-0.003	0.004	0.019	0.033	-0.023	0.01	0.003
	$\sqrt{MSE}$	0.566	0.887	0.184	0.291	0.059	0.328	0.563	0.163	0.228	0.033	0.248	0.415	0.166	0.22	0.021	0.228	0.38	0.15	0.184	0.018
T=100, J=12	bias	0.233	0.226	0.02	-0.066	0.027	0.054	0.037	0.019	-0.066	0.007	0.004	-0.012	-0.027	0.005	0.002	0	0	-0.028	0.007	0.001
	$\sqrt{MSE}$	0.592	0.703	0.256	0.305	0.072	0.279	0.4	0.154	0.211	0.028	0.167	0.282	0.157	0.201	0.013	0.127	0.225	0.143	0.164	0.005
T=100, J=20	bias	0.198	0.423	0.047	-0.101	0.025	0.074	0.107	0.033	-0.074	0.01	-0.009	-0.005	-0.008	-0.009	0.001	-0.003	0.004	-0.01	0.004	0.001
	$\sqrt{MSE}$	0.552	0.89	0.164	0.27	0.055	0.311	0.52	0.129	0.194	0.034	0.115	0.264	0.115	0.169	0.005	0.104	0.226	0.103	0.125	0.004
T=200, J=12	bias	0.184	0.167	0.011	-0.049	0.019	0.026	0.011	0.021	-0.061	0.004	-0.006	-0.027	-0.015	-0.001	0.001	0.002	-0.007	-0.016	0.006	0.001
	$\sqrt{MSE}$	0.466	0.601	0.176	0.262	0.053	0.184	0.313	0.113	0.172	0.018	0.088	0.219	0.108	0.164	0.003	0.091	0.174	0.099	0.123	0.003

1.7 Empirical application

The objective of the empirical exercise is twofold. First, we want to verify how well our instruments perform at estimating a flexible distribution of RCs using a real data set. Second, we want to study how the shape of the distribution of RCs can modify key counterfactual quantities such as the price elasticities or the pass-through, and check whether the results we obtain are consistent with the findings in Miravete et al. (2022). To do so, we estimate demand for cars using data on new car registrations in Germany from 2012 to 2018.[24] There are many reasons to focus on the car market. First, cars are highly differentiated products, which makes the BLP framework particularly adapted to this market. As a result, the BLP demand model has been widely applied to study the car industry (e.g., Berry et al. (1995), Grigolon et al. (2018), Petrin (2002)) and one can easily compare our results with previous results obtained in the literature under different specifications. Second, there are many policy-relevant questions related to this market. In particular, the role of road transport in air pollution is significant and many countries have

[24]The dataset was kindly provided to us by Kevin Remmy https://kevinremmy.com/research/.

implemented tax policies to reduce the CO2 emissions generated by car transportation.[25] An important strand of the literature has investigated the performance of these different taxation schemes (Alberini and Horvath (2021), Allcott and Wozny (2014), D'Haultfœuille, Givord, and Boutin (2014), Durrmeyer (2022), Durrmeyer and Samano (2018), Gillingham and Houde (2021), Grigolon et al. (2018), Huse and Koptyug (2022), Kunert (2018)). Other policy-relevant questions include the impact of import tariffs (Miravete et al. (2018)) and the determinants of market power (Berry et al. (1995), Grieco, Murry, and Yurukoglu (2022)). To answer these questions, the researcher must often estimate the demand for cars. The credibility of the implied analysis depends critically on how well the model can reproduce the underlying substitution patterns and the shape of the demand curve. To this end, it is essential to have a demand system that is sufficiently flexible, and particularly so with respect to the random coefficient on price. In this section, we use our instruments to estimate a Gaussian mixture as the random coefficient associated with price. Moreover, we use our test to assess how moving from the usual Gaussian RC to the Gaussian mixture decreases the degree of misspecification. Finally, we compare the counterfactual quantities under a Gaussian mixture and the traditional specifications (Gaussian RC and logit). In line with the findings in Miravete et al. (2022), our results indicate that the Gaussian mixture yields higher pass-through rates and curvatures.

1.7.1 The Data

The data set includes state-level new car registrations, publicly available by the German Federal Motor Transport Authority (KBA) from 2012 to 2018. This gives us 112 markets defined by state-year pairs. Data on car characteristics and price are scraped from General German Automobile Club and include horsepower, engine type, size, weight, fuel cost, CO2 emission, number of doors, segment, and body type. The data set is at a granular level where every car is uniquely identified by its manufacturer and its type key code (HSN/TSN) that is defined according to the characteristics of the car. Following the

[25] In 2017, road transport was responsible of approximately 19% of total greenhouse has emissions in EU-28 Retrieved from https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases/transport-emissions-of-greenhouse-gases-12 on October 21, 2022.

literature, we aggregate products with the same brand, model, engine type, and body combination (e.g. BMW-1 Series-Diesel-Hatchback).[26] Likewise, we follow the literature and define the market size as the number of households in the market. To construct market shares, we simply divide new car registrations of a given product by the market size. The data set is complemented by information on demographics such as the number of households or the average income per household at the state-year level and yearly average gas price data from ADAC.[27]

Summary statistics. The shares of products sold by engine type are presented in Table 1.5. We focus our analysis on combustion engine vehicles as in our sample period electric-vehicle cars constitute a small market share (always less than 5% of the sold vehicles) and can be seen as a distinct market. Between diesel and gasoline cars, we observe that the market share for diesel cars decreases over time, starting from 2016. The timing is in line with the emissions scandal, known as the Dieselgate, which started in September 2015.

Table 1.5: Shares (%) of new registrations by engine type

	Year						
Fuel Type	2012	2013	2014	2015	2016	2017	2018
Diesel	46.8	46.1	46.3	46.4	43.9	36.2	30.0
Gasoline	52.6	52.9	52.6	52.3	54.4	60.8	66.5
Battery EV	0.1	0.2	0.3	0.4	0.3	0.7	1.1
Hybrid EV	0.5	0.8	0.7	0.6	1	1.4	1.6
Plug-in hybrid EV	0	0	0.1	0.3	0.4	0.9	0.9

Table 1.6 provides sales-weighted averages for prices and observed characteristics. We observe that the difference in fuel consumption and resulting fuel costs steadily ranks diesel above gasoline. However,

[26]In aggregating the products from the HSN/TSN level, we use the characteristics of the car with the highest sales.
[27]State level income https://ec.europa.eu/eurostat/web/products-datasets/-/nama_10r_2hhinc

the average price of diesel cars sold is higher than gasoline cars. This implies a potential trade-off in terms of the costs of car ownership at the time of purchase. With a fixed mileage in mind, a consumer with high sensitivity to fuel costs might be willing to pay a higher price for a more fuel-efficient car. We also observe that the horsepower and the size of the newly registered cars increase over time.

Table 1.6: Summary Statistics (Sales weighted)

	Year						
	2012	2013	2014	2015	2016	2017	2018
Diesel							
Price/income	0.74	0.72	0.73	0.72	0.71	0.69	0.68
Size (m2)	8.31	8.31	8.32	8.36	8.42	8.48	8.53
Horsepower (kW/100)	1.09	1.07	1.11	1.11	1.14	1.16	1.21
Fuel cost (euros/100km)	7.90	7.18	6.63	5.53	4.94	5.25	5.83
Fuel cons. (Lt./100km)	5.19	4.98	4.89	4.73	4.61	4.61	4.71
CO2 emission (g/km)	136.19	130.50	127.69	123.58	120.42	120.49	123.27
Nb. of products/market	133	138	146	150	151	149	143
Gasoline							
Price/income	0.46	0.46	0.46	0.46	0.46	0.45	0.43
Size (m2)	7.23	7.27	7.28	7.30	7.36	7.46	7.53
Horsepower (kW/100)	0.79	0.78	0.80	0.82	0.85	0.88	0.91
Fuel cost (euros/100km)	9.48	8.61	8.11	7.27	6.69	7.06	7.40
Fuel cons. (Lt./100km)	5.76	5.47	5.40	5.31	5.25	5.34	5.38
CO2 emission (g/km)	135.80	128.18	125.27	122.89	121.22	122.86	123.26
Nb. of products/market	157	171	179	185	186	193	188

Note: Provided statistics are sales weighted averages across products. Total number of markets (State*Year) is 112 .

Inter-market variation. Our dataset contains both geographical variation and time variation, as we observe the sales in every state in Germany over the period 2012-2018. States in Germany differ significantly in terms of income per inhabitant, population density and average distance driven.[28] It is

[28]For the population density 2019 (inh/km^2): 69 (Mecklenburg-Vorpommern) to 4118 (Berlin) (from Federal Statistical Office of Germany (Destatis)), GDP per capita 2019: 28.9k (Mecklenburg-Vorpommern) to 67k (Hamburg)

fundamental to take this inter-market variation into account in our empirical specification for two reasons. First, our model postulates that consumers' preferences are the same across markets. However, we observe that the market shares vary from one state to the other even if the choice set remains the same. This feature of the data can only be explained if we let the preferences vary from one market to the other. Second, in Section 1.2.3, we saw that there needs to be sufficient variation in the product characteristics across markets to identify the distribution of RCs. By interacting product characteristics with state demographics, we achieve both objectives: we shift the preferences to a more common representation and we create variation in the product characteristics across markets. To choose which interaction terms to include in the utility function, we first create market specific sales-weighted characteristics for the following variables: price, fuel cost, size, horsepower, height, gasoline dummy, and foreign dummy (equal to one if the manufacturer of the car is not German). Then, we regress these quantities on the demographics of interest: average income, population density, and a time trend. Last, we select the interaction terms that explain the best the variation in sales-weighted characteristics (namely, the variables with a p-value below $1e^{-10}$). The results of these regressions are presented in Table 1.7. They suggest that income shifts positively the preferences for price, size, and horsepower (i.e. higher income is associated with larger cars, and higher horsepower). In contrast, income shifts negatively the preferences for foreign status, height, and gasoline status.[29] Although weaker, a similar pattern is observed for the effect of population density on car characteristics.

(retrieved from https://www.ceicdata.com/en/germany/esa-2010-gdp-per-capita-by-region/gdp-per-capita-bremen on 05 November 2022). For average driving distance in 2019: 13079 km (Mecklenburg-Vorpommern) to 9531 (Berlin) retrieved from https://de.statista.com/statistik/daten/studie/644381/umfrage/fahrleistung-privater-pkw-in-deutschland-nach-bundesland/ on 19 September 2022.

[29]In the main analysis, we use price/income to capture the income effect.

Table 1.7: Linear regressions of sales-weighted car characteristics on demographic characteristics

	Income(/1000)	Population density (/100)	Time trend
Price($\times$1000)	0.138**	0.069*	0.286*
	(0.013)	(0.011)	(0.059)
Fuel cost (euros/100km)	-0.0069	-0.0036	0.3587**
	(0.0063)	(0.0056)	(0.0296)
Size(m^2)	0.0058**	0.0018*	0.0176*
	(0.00079)	(0.00070)	(0.00371)
Horsepower (KW/100)	0.0028**	0.0012*	0.0129**
	(0.00028)	(0.00025)	(0.00132)
Foreign	-0.0050^{**}	-0.0014^{*}	0.0295**
	(0.00052)	(0.00046)	(0.00246)
Height(m)	-0.00051^{**}	-0.00043^{**}	0.00181*
	(0.000061)	(0.000054)	(0.000286)
Gasoline	-0.0067^{**}	-0.0024^{*}	0.0131*
	(0.00059)	(0.00053)	(0.00280)

Note: * p-value lower than 0.01, ** p-value lower than $1e^{-10}$.

Instruments for the endogeneity of price. To instrument for price, we use a combination of variables on the intensity of competition and cost shifters. To measure the intensity of competition, we consider the number of competing products of the same class and engine type in a given market, and the number of competing products of the same engine type in a given market. As for cost shifters, we use three complementary datasets: the mean hourly labor cost, the price of steel (interacted with the weight of the car), and exchange rates between Germany and the country of assembly.

1. Labor cost: we use the mean nominal hourly labor cost per employee in the manufacturing sector of the country of assembly of the models. The data on labor costs come from International Labor Organization Statistics (ILOSTAT).[30]

[30] Retrieved from https://www.ilo.org/ilostat-files/Documents/Excel/INDICATOR/LAC_4HRL_

2. Price of steel: we collect the price of steel futures in January of each year.

3. Exchange rates: we construct the exchange rates between Germany and the country of assembly of each car model using exchange rate data from OECD.[31]

1.7.2 Empirical specification

The indirect utility of consumer i, purchasing product j in market t (defined as a state-year pair) is given by:

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi^*_{jt}}_{\delta_{jt}} + x'_{2jt}\alpha_i + \varepsilon_{ijt}.$$

The mean utility $\delta_{jt} = x'_{1jt}\beta + \xi^*_{jt}$ captures homogeneous preferences. The variables in x_{1jt} consist of the product characteristics for which we assume that there is no preference heterogeneity and the interaction terms that explain the best the geographical variation observed in Table 1.7.[32]
The demand shock on product j is decomposed as follows:

$$\xi^*_{jt} = Brand_j + State_t + Year_t + \xi_{jt},$$

where $Brand_j$ is a brand fixed effect that captures the unobserved quality of the brand of product j, $State_t$ captures state specific demand shocks that are fixed across time and products and $Year_t$ captures year-specific demand shocks. Therefore, $State_t$ and $Year_t$ play a role in explaining the variation in the overall demand for cars (or equivalently, in the share of the outside option).

The variables in x_{2jt} are the product characteristics that display preference heterogeneity and which we augment with a RC. In our specification, we include the price, the size, and the gasoline dummy in x_{2jt}. We estimate the model assuming different specifications for the distribution of RCs. First, we estimate the model without any consumer heterogeneity. Second, we assume that all the RCs are normally

ECO_CUR_NB_A_EN.xlsx

[31] Retrieved from https://data.oecd.org/conversion/exchange-rates.htm

[32] The choice of the variables that display preference heterogeneity is based on our understanding of the car market and follows current empirical practices for this specific market. However, we understand the limitations of this approach, and we are working on an iterative procedure to select the variables that display consumer heterogeneity.

distributed. Finally, we consider a Gaussian mixture on price to increase flexibility with respect to the preferences on price. For each different specification, we perform the specification test developed in Section 1.5 to see how the degree of misspecification evolves as we increase flexibility on the distribution of RCs.

1.7.3 Estimation

Estimation conditional logit (no heterogeneity). First, we estimate the logit model, and we report the results in Table 1.8.[33] As expected, we find a negative effect of price and fuel cost on the utility. The interaction terms indicate that the utility derived from size, horsepower, foreign status and gasoline all decrease with income. Moreover, we observe that the aversion to fuel cost decreases over time, which is likely an artifact implied by increasing fuel cost over the years. In contrast, the utility derived from horsepower appears to increase with time. However, these time effects are smaller in comparison with the heterogeneity due to income. To facilitate the interpretation of these results, we consider a household with a €47,000 income in 2018. This corresponds to the mean income in 2018. For this household, the implied effect of size on the utility is negative, whereas a positive utility is derived from higher horsepower, the car's brand being German, height, and gasoline engines.

[33] In Appendix 1.E, we provide results for baseline specifications including the simple conditional logit and the nested logit (with and without state and year fixed effects).

Table 1.8: Logit estimation

	Baseline		**× Income (/1000)**		**× Pop. density(/100)**		**× Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-2.4	1.3e-01	-	-	-	-	-	-
Fuel Cost	-0.25	8.6e-03	-	-	-	-	0.014	1.7e-03
Size(m^2)	0.15	4.2e-02	-0.0055	8.5e-04	-	-	-	-
Horsepower(KW/100)	2.7	1.8e-01	-0.019	2.4e-03	-	-	-0.081	7e-03
Foreign	0.18	7.1e-02	-0.017	1.4e-03	-	-	-	-
Height(m)	3.5	2.3e-01	-0.0015	4.6e-03	-0.036	4.7e-03	-	-
Gasoline	1.1	6.3e-02	-0.011	1.2e-03	-	-	-	-

Note: Brand, Year and State FE's are included.

Estimation with Gaussian random coefficients. We now increase the flexibility in the *traditional* manner, by assuming that the RCs on the price, the size and the gasoline indicator follow a Gaussian distribution. We report the estimates obtained under this new specification in Table 1.9. The signs for the homogeneous preference parameters in x_{1jt} remain the same and the magnitude of the effects do not change significantly. The sign associated with the mean effect of price remains negative. In contrast, the sign on the mean effects of the size and the gasoline dummy are inverted with respect to the logit specification. This last observation illustrates an important empirical finding: average effects are not invariant to the introduction of preference heterogeneity. In other words, the logit estimates do not necessarily match the means, when we introduce a Gaussian RC. Moreover, the three RCs display high variances and particularly so for the gasoline dummy, which indicate a high level of heterogeneity with respect to these three characteristics.[34]

[34]The estimation is performed using the parametrization proposed in Ketz (2019), which avoids boundary issues at 0 for the variances of the RCs.

Table 1.9: Traditional BLP (Gaussian RC)

	Baseline		$\times$ **Income (/1000)**		$\times$ **Pop. density(/100)**		$\times$ **Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-	-	-	-	-	-	-	-
Fuel Cost	-0.29	5.1e-03	-	-	-	-	0.031	9.2e-04
Size(m^2)	-	-	-0.0053	3.1e-04	-	-	-	-
Horsepower(KW/100)	0.77	1.5e-02	0.0078	6.8e-04	-	-	-0.12	5.6e-03
Foreign	0.21	5.4e-02	-0.019	1.1e-03	-	-	-	-
Height(m)	3.4	1.1e-02	-0.0088	1.2e-03	-0.032	3.6e-04	-	-
Gasoline	-	-	-0.0028	8.6e-04	-	-	-	-
Gaussian RC	$\hat{\beta}$	S.E	$\hat{\sigma}$	S.E				
Price/income	-2.4	2e-02	0.96	5.9e-03	-	-	-	-
Size(m^2)	-0.37	1.5e-02	0.43	3.6e-03	-	-	-	-
Gasoline	-2.3	4.4e-02	4	4.1e-04	-	-	-	-

Note: Brand, Year and State FE's are included.

Estimation with a Gaussian mixture on the price. Finally, we increase the flexibility of the model, by replacing the Gaussian RC on the price variable with a Gaussian mixture of 2 components. We focus on the price as the literature shows that the distribution of price sensitivity is absolutely key for many quantities of interest in IO, including the price elasticities and the pass-through. We report the estimates obtained under this new specification in Table 1.10. The results point out the presence of two distinct modes in the distribution of the RC associated with price. The two modes reveal the presence of two groups of consumers: the first one with high price sensitivity (with the mean component at -9.6) and the second one with low price sensitivity (with the mean component at -2.5). Moreover, the distribution is heavily asymmetric with the probability of the first mode being 0.9, which entails that the majority of consumers are highly sensitive to price. This last feature is completely absent in the logit and Gaussian specifications, which seem to capture only the first mode of the distribution as we can see in Figure 1.6. Once again the homogeneous parameters are relatively unchanged with respect to the previous specifications. The Gaussian RC on the gasoline still displays a high variance (the standard deviation of the RC equals 2.8).

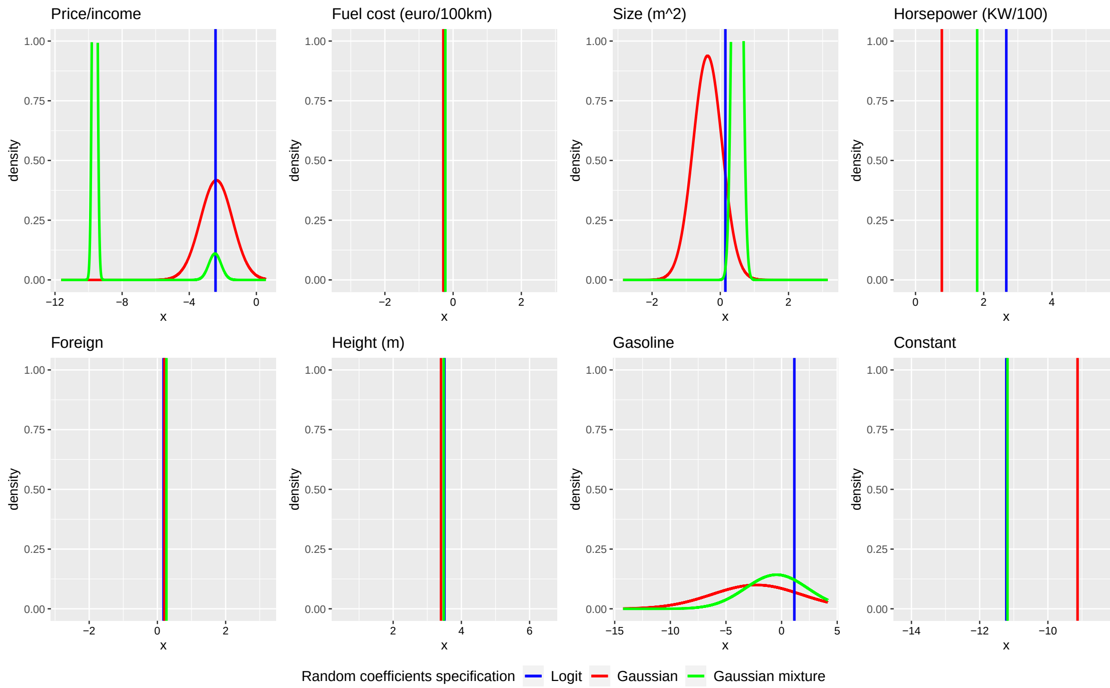
Table 1.10: Estimation Gaussian mixture on Price

	Baseline		**× Income (/1000)**		**× Pop. density(/100)**		**× Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-	-	-	-	-	-	-	-
Fuel Cost	-0.23	5.8e-03	-	-	-	-	0.026	1e-03
Size(m^2)	-	-	-0.0055	3.7e-04	-	-	-	-
Horsepower(KW/100)	1.8	3.6e-02	-0.0016	1.1e-03	-	-	-0.1	7e-03
Foreign	0.26	6.1e-02	-0.021	1.2e-03	-	-	-	-
Height(m)	3.5	1.1e-02	-0.012	1.2e-03	-0.032	3.7e-04	-	-
Gasoline	-	-	-0.026	1.3e-03	-	-	-	-
Gaussian RC	$\hat{\beta}$	S.E	$\hat{\sigma}$	S.E				
Size(m^2)	0.5	1.9e-02	0.1	6.7e-02	-	-	-	-
Gasoline	-0.45	3.8e-03	2.8	9.1e-03	-	-	-	-
Gaussian Mixture	$\hat{\beta}_1$	S.E	$\hat{\sigma}_1$	S.E	$\hat{\beta}_2$	S.E	$\hat{\sigma}_2$	S.E
Price/income	-9.6	1.8e-02	0.1	1.8e-03	-2.5	1.8e-02	0.35	5.2e-04
Probability	0.9	6.8e-05						

Note: Brand, Year and State FE's are included.

In Figure 1.6, we plot the estimated distribution of random coefficients under the three specifications we consider. We observe little to no variation in the homogeneous parameters from one specification to the other. The main difference comes from the introduction of the Gaussian mixture on price, which reveals the presence of a large group of highly price sensitive consumers.

Figure 1.6: Estimated distributions of RCs in the three specifications



Specification Test. By increasing the flexibility on the distribution of RCs, we recover less precise estimates and the model becomes more difficult to estimate. Thus, it is important to show that the additional flexibility substantially reduces the misspecification of the model. To quantify the degree of misspecification accross the different models, we keep the same set of estimation instruments across the different specifications of RCs and we report the value of the associated Sargan-Hansen J statistics in each case. Moreover, for every model, we follow the procedure developed in Section 1.5 to test if the distribution of RCs on price is well specified. We use the global interval instruments and we denote this test "Interval test". We report the values of the test statistics and the degrees of freedom of the chi-square under the null in Table 1.11. We observe an important decrease in the Sargan-Hansen J statistic when we transition from the logit to the Gaussian RC. However, the decrease in the Sargan-Hansen J statistic is much larger when we transition from the Gaussian RC on price to the Gaussian mixture, which indicates that the Gaussian mixture performs much better than the simple Gaussian at capturing the underlying heterogeneity in price sensitivity. The interval test displays a similar behavior, with the largest decrease in the test statistic stemming from the transition from the Gaussian RC to the Gaussian mixture.

Table 1.11: Evolution of misspecification with flexibility

Instruments	Logit			Gaussian RC			Gaussian mixture		
Test	Stat.	Critical val.	DF	Stat.	Critical val.	DF	Stat.	Critical val.	DF
J test	2755.7	40.1	27	2341.7	36.4	24	950.3	33.9	21
Interval test	1331.9	14.1	7	999.4	14.1	7	244.0	14.1	7

1.7.4 Counterfactual quantities

The objective of this subsection is to illustrate how changes in the distribution of the RC associated with price affect many counterfactual quantities of interest in empirical IO, such as the price elasticities, the marginal costs faced by car manufacturers, and the pass-through of cost. In order to compare our empirical results with the findings in Miravete et al. (2022), we also calculate the demand curvature under the different specifications. They show that a large demand curvature is necessary to recover a pass-through larger than one.

In the following, we study the effect of different specifications on the price elasticities, demand curvature, marginal costs and mark-ups, and finally on the pass-through. To recover marginal costs and mark-ups, we assume that multi-product firms pricing under Bertrand-Nash pricing. For the pass-through, we calculate the new equilibrium prices using fixed point iterations and, following the literature, study the effect of increasing the marginal costs of each product by 1% and recomputing the marginal cost. In Appendix 1.E, we provide details on the calculation of counterfactual quantities. In our computations, we use the year 2018, which is the last year of our sample.

Summary of results. We report the median values for the five counterfactual quantities of interest in Table 1.12. Several remarks are in order. First, the Gaussian mixture yields a much lower price elasticity than the two other specifications. This is related to the emergence of a group of highly price sensitive consumers in the mixture specification, which we fail to detect with the logit and Gaussian RC specifications. Moreover, the low price elasticities that we recover in the Gaussian and logit specifications, generate unreasonably low marginal costs (even negative ones as we can see in Figure 1.7) and excessive mark-ups. In contrast, this problem does not appear with the Gaussian mixture. Finally, to link our results with the findings in Miravete, Seim, and Thurk (2022), we now focus on the demand curvature and the pass-through of cost. As expected, the logit displays a curvature and a pass-through equal to 1. In contrast, we can see that the Gaussian mixture displays a larger demand curvature than the other two specifications. This comes from the skewness that the mixture induces in the distribution of price sensitivity. This last feature implies that the Gaussian mixture yields a pass-through much greater than 1 (1.5 on average). Unfortunately, the negative marginal costs we recover with the Gaussian RC prevent us from computing the pass-through in this case.[35]

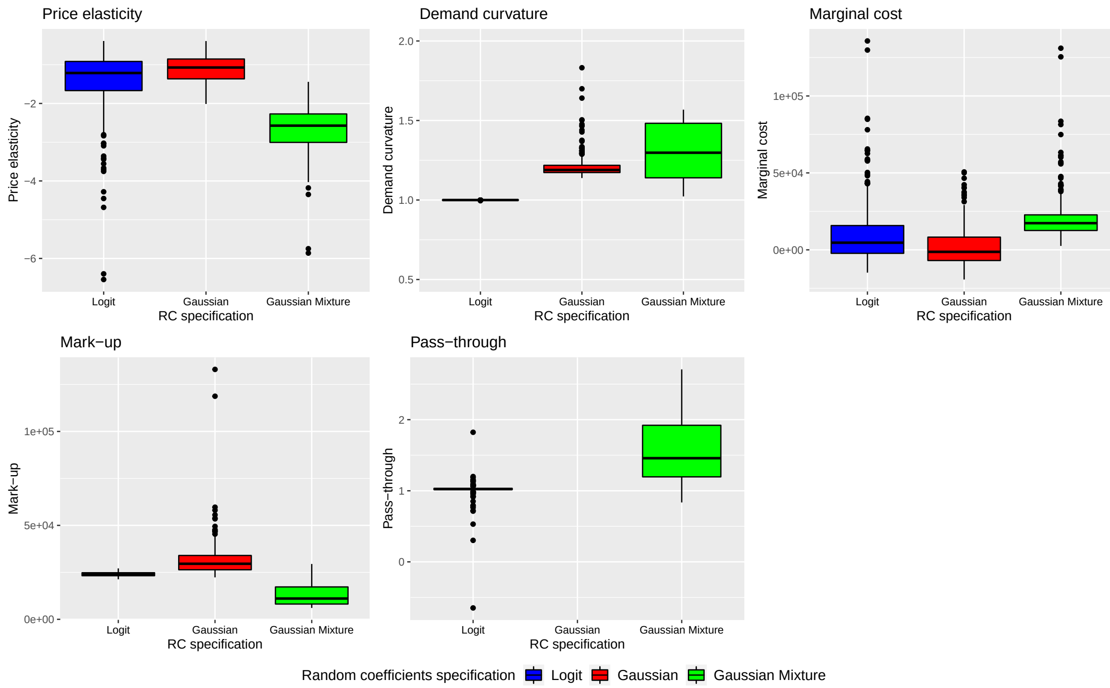
Table 1.12: Median counterfactual quantities under different specifications on RCs

RC distribution on price	Logit	Gaussian	Gaussian Mixture
Own price-elasticity	-1.2	-1.1	-2.6
Demand curvature	1.0	1.2	1.3
Marginal cost	9,366	1,929	20,105
Mark-up	24,048	29,572	11,066
Pass-through	1.0	-	1.5

In Figure 1.7, we plot the empirical distributions of the counterfactual quantities. We can see in the plot featuring the distribution of marginal costs that the logit and Gaussian specifications generate negative marginal costs for some of the cars. This is an indication that the price elasticities implied by these specifications are too low in absolute value.

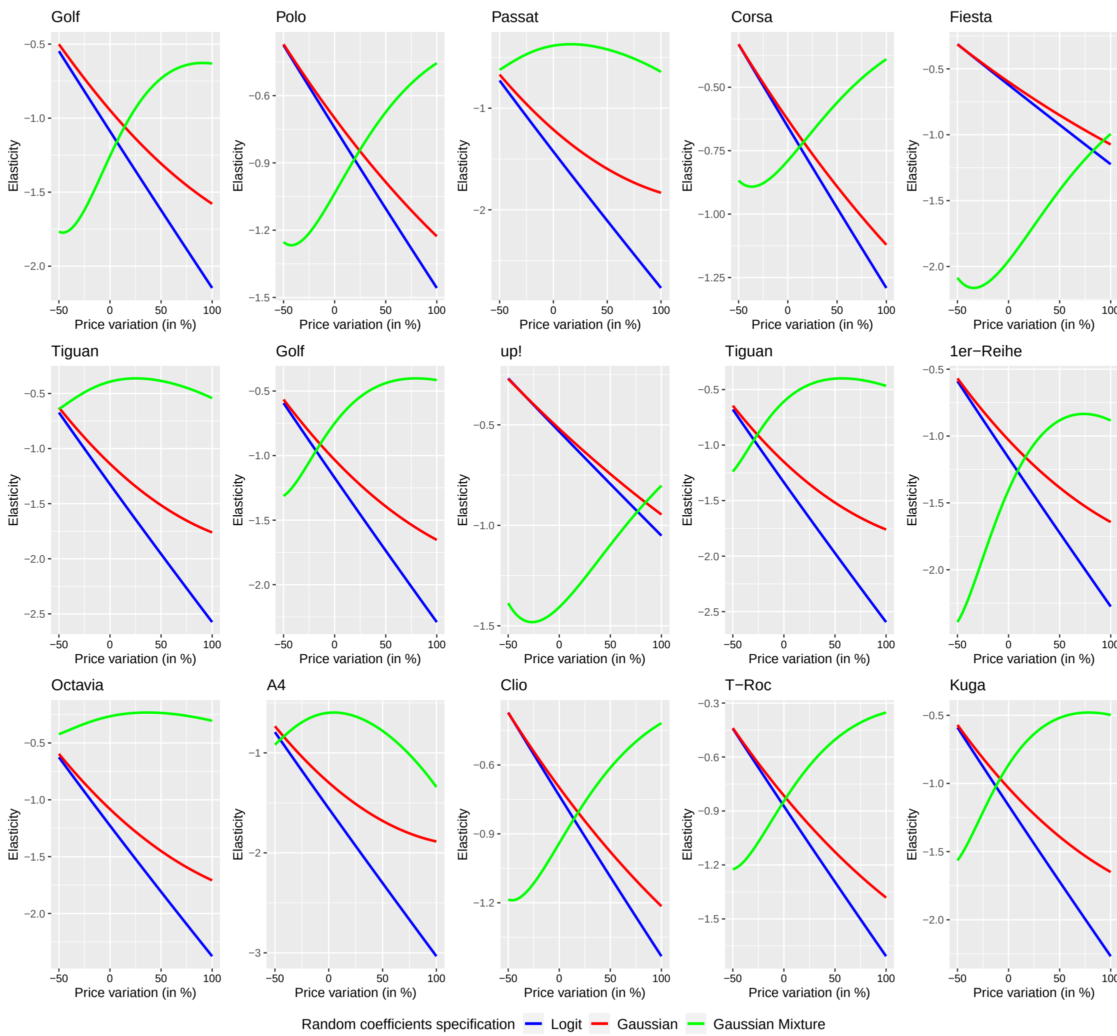
[35] Our algorithm to compute the new equilibrium prices after the change in cost does not converge.

Figure 1.7: Empirical distribution of counterfactual quantities under different specifications



Finally, in Figure 1.8, we plot the elasticity functions implied by the different specifications for the 15 most popular cars in our sample. We observe important differences in the elasticities. The Gaussian mixture generates lower price elasticities than the other two specifications. We do the same exercise with the demand curves in Appendix 1.E.

Figure 1.8: Estimated elasticities under different specifications



1.8 Conclusion

In this paper, we develop novel econometric tools to parsimoniously increase the flexibility of the distribution of random coefficients in the BLP demand model initiated by Berry et al. (1995). Specifically, we construct a formal moment-based specification test on the distribution of random coefficients, which allows researchers to test the chosen specification without having to re-estimate the model under a more flexible parametrization. The moment conditions (or equivalently the instruments) are designed to maximize the power of the test when the distribution of RC is misspecified. By exploiting the duality between estimation and testing, we show that these instruments can also improve the estimation of the BLP model under a flexible parametrization. Our Monte Carlo simulations confirm that the interval instruments we develop in this paper outperform the traditional instruments both for testing and estimating purposes. Finally, we apply these new tools to flexibly estimate the demand for cars in Germany. We show that these tools can be applied to the equally popular mixed logit demand model with individual-level data.

In future work, we plan to see if we can generalize these instruments to other non-linear moment-based models, as well as to the general problem of testing distributional assumptions in structural models. From a broader perspective, our paper is part of an existent discussion on the most effective way to model unobserved preference heterogeneity in structural models. Most empirical frameworks feature a clear trade-off between the degree of flexibility one chooses and the precision of the estimates one obtains. It is thus critical to understand how misspecification on the unobserved heterogeneity affects the counterfactual quantities of interest. In the case of the BLP demand model, our paper and others show that misspecification in the distribution of random coefficients substantially distorts the substitution patterns as well as the shape of the demand curve and, thus, is likely to significantly alter the counterfactual quantities.

Bibliography

ALBERINI, A. AND M. HORVATH (2021): "All car taxes are not created equal: evidence from Germany," *Energy Economics*, 100.

ALLCOTT, H. AND N. WOZNY (2014): "Gasoline prices, fuel economy, and the energy paradox," *Review of Economics and Statistics*, 96, 779–795.

ALLEN, R. AND J. REHBECK (2020): "Identification of random coefficient latent utility models," *arXiv preprint: 2003.00276*.

AMEMIYA, T. (1974): "The nonlinear two-stage least-squares estimator," *Journal of Econometrics*, 2, 105–110.

ANDREWS, D. W. AND P. GUGGENBERGER (2019): "Identification and singularity-robust inference for moment condition models," *Quantitative Economics*, 10, 1703–1746.

ARMSTRONG, T. B. (2016): "Large market asymptotics for differentiated product demand estimators with economic models of supply," *Econometrica*, 84, 1961–1980.

BAHADUR, R. R. (1960): "Stochastic comparison of tests," *The Annals of Mathematical Statistics*, 31, 276–295.

BARAHONA, N., C. OTERO, S. OTERO, AND J. KIM (2020): "Equilibrium effects of food labeling policies," *SSRN working paper*.

BERRY, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

BERRY, S. T. AND P. A. HAILE (2009): "Nonparametric identification of multinomial choice demand models with heterogeneous consumers," *NBER working paper 15276*.

——— (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

BERRY, S. T., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

BERRY, S. T., O. B. LINTON, AND A. PAKES (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *The Review of Economic Studies*, 71, 613–654.

CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

CHETVERIKOV, D. AND D. WILHELM (2017): "Nonparametric instrumental variable estimation under monotonicity," *Econometrica*, 85, 1303–1320.

COMPIANI, G. (2018): "Nonparametric demand estimation in differentiated products markets," *SSRN working paper*.

CONLON, C. AND J. GORTMAKER (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51, 1108–1161.

CRAWFORD, G. S., O. SHCHERBAKOV, AND M. SHUM (2019): "Quality overprovision in cable television markets," *American Economic Review*, 109, 956–995.

D'HAULTFŒUILLE, X., P. GIVORD, AND X. BOUTIN (2014): "The environmental effect of green taxation: the case of the French *bonus/malus*," *The Economic Journal*, 124, F444–F480.

DUBÉ, J.-P., J. T. FOX, AND C.-L. SU (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80, 2231–2267.

DUBOIS, P., R. GRIFFITH, AND M. O'CONNELL (2018): "The effects of banning advertising in junk food markets," *The Review of Economic Studies*, 85, 396–436.

DUFOUR, J.-M. AND M. L. KING (1991): "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR (1) errors," *Journal of Econometrics*, 47, 115–143.

DUNKER, F., S. HODERLEIN, AND H. KAIDO (2022): "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models," *arXiv preprint: 2201.06140*.

DURRMEYER, I. (2022): "Winners and losers: the distributional effects of the French feebate on the automobile market," *The Economic Journal*, 132, 1414–1448.

DURRMEYER, I. AND M. SAMANO (2018): "To rebate or not to rebate: fuel economy standards vs. feebates," *The Economic Journal*, 128, 3076–3116.

FOX, J. T. AND A. GANDHI (2011): "Identifying demand with multidimensional unobservables: a random functions approach," *NBER working paper 17557*.

FOX, J. T., K. IL KIM, S. P. RYAN, AND P. BAJARI (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212.

FREYBERGER, J. (2015): "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, 185, 162–181.

GANDHI, A. AND J.-F. HOUDE (2019): "Measuring substitution patterns in differentiated-products industries," *NBER working paper 26375*.

GEWEKE, J. (1981): "The approximate slopes of econometric tests," *Econometrica*, 49, 1427–1442.

GILLINGHAM, K. AND S. HOUDE (2021): "Consumer myopia in vehicle purchases: evidence from a natural experiment," *American Economic Journal: Economic Policy*, 207–238.

GOURIEROUX, C. AND A. MONFORT (1995): *Statistics and econometric models*, vol. 2, Cambridge University Press.

GRENNAN, M. (2013): "Price discrimination and bargaining: empirical evidence from medical devices," *American Economic Review*, 103, 145–177.

GRIECO, P. L., C. MURRY, AND A. YURUKOGLU (2022): "The evolution of market power in the US automobile industry," *NBER working paper 29013*.

GRIGOLON, L., M. REYNAERT, AND F. VERBOVEN (2018): "Consumer valuation of fuel costs and tax policy: evidence from the European car market," *American Economic Journal: Economic Policy*, 10, 193–225.

HANSEN, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

HO, K. AND A. PAKES (2014): "Hospital choices, hospital prices, and financial incentives to physicians," *American Economic Review*, 104, 3841–3884.

HUSE, C. AND N. KOPTYUG (2022): "Salience and policy instruments: evidence from the auto market," *Journal of the Association of Environmental and Resource Economists*, 9, 345–382.

ICHIMURA, H. AND T. S. THOMPSON (1998): "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 86, 269–295.

KETZ, P. (2019): "On asymptotic size distortions in the random coefficients logit model," *Journal of Econometrics*, 212, 413–432.

KUNERT, U. (2018): "Diesel fuel and passenger cars receive preferential tax treatment in Europe: reform of taxation needed in Germany," *DIW Weekly Report*, 8, 289–298.

LEE, J. AND K. SEO (2015): "A computationally fast estimator for random coefficients logit demand models using aggregate data," *The RAND Journal of Economics*, 46, 86–102.

LEWBEL, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

LU, Z., X. SHI, AND J. TAO (2021): "Semi-nonparametric estimation of random coefficient logit model for aggregate demand," *SSRN working paper*.

MCFADDEN, D. AND K. TRAIN (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

MILLER, N. H., G. SHEU, AND M. C. WEINBERG (2021): "Oligopolistic price leadership and mergers: the United States beer industry," *American Economic Review*, 111, 3123–3159.

MILLER, N. H. AND M. C. WEINBERG (2017): "Understanding the price effects of the MillerCoors joint venture," *Econometrica*, 85, 1763–1791.

MIRAVETE, E. J., M. J. MORAL, AND J. THURK (2018): "Fuel taxation, emissions policy, and competitive advantage in the diffusion of European diesel automobiles," *The RAND Journal of Economics*, 49, 504–540.

MIRAVETE, E. J., K. SEIM, AND J. THURK (2022): "Robust pass-through estimation in discrete choice models," *Working paper*.

NEVO, A. (2000): "Mergers with differentiated products: the case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, 31, 395–421.

NEWEY, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, 58, 809–837.

——— (2004): "Efficient semiparametric estimation via moment restrictions," *Econometrica*, 72, 1877–1897.

PETRIN, A. (2002): "Quantifying the benefits of new products: the case of the minivan," *Journal of Political Economy*, 110, 705–729.

REYNAERT, M. (2021): "Abatement strategies and the cost of environmental regulation: emission standards on the European car market," *The Review of Economic Studies*, 88, 454–488.

REYNAERT, M. AND F. VERBOVEN (2014): "Improving the performance of random coefficients demand models: the role of optimal instruments," *Journal of Econometrics*, 179, 83–98.

ROODMAN, D. (2009): "A Note on the theme of too many instruments," *Oxford Bulletin of Economics and Statistics*, 71, 135–158.

SALANIÉ, B. AND F. A. WOLAK (2019): "Fast, "robust", and approximately correct: estimating mixed demand systems," *NBER working paper 25726*.

SILVEY, S. D. (1959): "The Lagrangian multiplier test," *The Annals of Mathematical Statistics*, 30, 389–407.

TEBALDI, P., A. TORGOVITSKY, AND H. YANG (2019): "Nonparametric estimates of demand in the California health insurance exchange," *NBER working paper 25827*.

TRAIN, K. E. (2008): "EM algorithms for nonparametric estimation of mixing distributions," *Journal of Choice Modelling*, 1, 40–69.

WANG, A. (2022): "Sieve BLP: A semi-nonparametric model of demand for differentiated products," *Journal of Econometrics*.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

1.A Extension to the mixed logit demand model

The main difference between the BLP demand model and the mixed logit model is that the latter one assumes that the econometrician observes individual data. Let us consider the baseline mixed logit model with no endogeneity and consumer level data.[36] We observe the choices of N consumers. The indirect utility function of consumer i making choice $j \in \{0, 1, \dots, J\}$ is given by:

$$u_{ij} = x'_{1ij}\beta_0 + x'_{2ij}v_i + \varepsilon_{ij}, \tag{1.A.12}$$

where

- ε_{ij} is a preference shock that follows a type I extreme value distribution independent of all other variables and across i, j;
- x_{1ij} is a vector of product characteristics interacted with consumer characteristics of dimension K_1 which display no preference heterogeneity;
- x_{2ij} is a vector of product characteristics interacted with consumer characteristics of dimension K_2 which display preference heterogeneity;
- v_i is a vector of random coefficients of dimension K_2 which jointly follows a distribution characterized by a density f;

Each consumer chooses the product that maximizes his or her utility. For any couple $(\tilde{f}, \tilde{\beta})$, demand for product j from consumer i writes:

$$\forall j \neq 0, \quad \rho_j(x_i, \tilde{f}, \tilde{\beta}) = \int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{1ij}\tilde{\beta} + x'_{2ij}v\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\tilde{\beta} + x'_{2ik}v\right\}} \tilde{f}(v)dv. \tag{1.A.13}$$

For the outside option, we have:

$$\text{for } j = 0, \quad \rho_j(x_i, \tilde{f}, \tilde{\beta}) = \int_{\mathbb{R}^{K_2}} \frac{1}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\tilde{\beta} + x'_{2ik}v\right\}} \tilde{f}(v)dv. \tag{1.A.14}$$

[36] In the mixed logit case, the absence of endogenous variables is not an unrealistic assumption as the econometrician can always model unobserved product quality by incorporating product fixed effects into the utility function.

Structural error. As we did in the case of the BLP demand model, we can define the structural error generated by $(\tilde{\beta}, \tilde{f})$ as follows. Let y_{ij} equal to 1 if individual i chooses good $j = 0, 1, \ldots, J$, the structural error writes:

$$\xi_{ij}(\tilde{f}, \tilde{\beta}) = y_{ij} - \rho_j(x_i, \tilde{\beta}, \tilde{f}).$$

By construction, at the true (f, β), we have $\mathbb{E}[\xi_{ij}(\beta, f)|x_i] = \mathbb{E}[y_{ij}|x_i] - \rho_j(x_i, \beta, f) = 0$ a.s..The notation x_i refers to $(x_{ij})_{j=1,\ldots,J}$.

Most powerful instrument and approximations. As in the aggregate demand model, we want to derive the instruments with the greatest ability to detect misspecification in the distribution of RCs. Given that the model displays no endogeneity, the set of exogenous variables is simply x_i. Our objective is to find the functions of x_i, which provides the most detection power against a wrong distribution. With this objective in mind, we consider a situation where the econometrician has a candidate (f_0, β_0) and wants to test $\bar{H}_0 : (f, \beta) = (f_0, \beta_0)$ against $H_a : (f, \beta) \neq (f_0, \beta_0)$. We proceed as in the BLP case and we derive the instrument that maximizes the power of the associated moment based test. Second, we propose feasible approximations of the MPI that do not depend on the fixed alternative.

For any set of testing instruments $h_D(x_i)$, we have the following implication:

$$\overline{H}_0 : (f, \beta) = (f_0, \beta_0) \implies \overline{H}'_0 : \quad \mathbb{E}[h_D(x_i)\xi_{ij}(f_0, \beta_0)] = 0.$$

We propose to test $\overline{H}_0$ indirectly through its implication $\overline{H}'_0$, which is a set of unconditional moment conditions. We test $\overline{H}'_0$ with a moment-based test and the test statistic writes as follows:

$$S_N(h_D, f_0, \beta_0) = NJ \left(\frac{1}{NJ} \sum_{i,j} \xi_{ij}(f_0, \beta_0) h_D(x_i) \right)' \hat{\bar{\Omega}}_0^{-1} \left(\frac{1}{NJ} \sum_{i,j} \xi_{ij}(f_0, \beta_0) h_D(x_i) \right), \quad (1.A.15)$$

with $\hat{\bar{\Omega}}_0$ a consistent estimator of $\overline{\Omega_0}$ the asymptotic variance-covariance matrix of $\frac{1}{NJ} \sum_{i,j} \xi_{ij}(f_0, \beta_0) h_D(x_i)$ under $\overline{H}_0$:

$$\bar{\Omega}_0 = \mathbb{E} \left[\left(\frac{1}{\sqrt{J}} \sum_j \xi_{ij}(f_0, \beta_0) h_D(x_i) \right) \left(\frac{1}{\sqrt{J}} \sum_j \xi_{ij}(f_0, \beta_0) h_D(x_i) \right)' \right].$$

Assuming that (x_i, y_i) are i.i.d. across individuals and consistent with the probability model defined by equations (1.A.12,1.A.13, 1.A.14) evaluated at (f, β), $\mathbb{E}[\|\xi_{ij}(f_0, \beta_0)h_D(x_i)\|^2] < +\infty$, and Ω_0 has full rank, we can show:

- under $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$, $\quad S_N(h_D, f_0, \beta_0) \xrightarrow[T\to+\infty]{d} \chi^2_{|h_D|_0}$,
- under $H'_a :\ \mathbb{E}\left[h_D(x_i)\xi_{jt}(f_0, \beta_0)\right] \neq 0, \quad \forall q \in \mathbb{R}^+,\ \mathbb{P}(S_N(h_D, f_0, \beta_0) > q) \xrightarrow[T\to+\infty]{} 1$,

with $|\cdot|_0$ being the counting norm. The proof is almost identical to the proof of Proposition 3.1 and thus, we omit it. Following the same steps as the proof of Proposition 3.2, the expression for the Most Powerful Instrument (that maximizes the slope of the test) writes:

$$h^*_D(x_i) = \mathbb{E}[\xi_{ij}(f_0, \beta_0)^2|x_i]^{-1}\Delta(x_i, f_0, \beta_0, f_a, \beta_a),$$

where each component j of the correction term $\Delta(x_i, f_0, \beta_0, f_a, \beta_a)$ writes:

$$\begin{aligned}\Delta(x_i, f_0, \beta_0, f_a, \beta_a)_j &= \rho_j(x_i, \beta_0, f_0) - \rho_j(x_i, \beta_a, f_a) \\ &= \int_{\mathbb{R}} \left[\rho_j(x_i, \beta_0, f_0) - \frac{\exp\{x'_{1ij}\beta_a + x'_{2ij}v\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\beta_a + x'_{2ik}v\right\}}\right] f_a(v).\end{aligned}$$

Several remarks are in order. First, contrary to the BLP case, the correction term $\Delta^{\xi_j}_{0,a}$ is a function of the exogenous variables x_i, and thus we don't need to compute its conditional expectation as in the BLP model. The conditional variance term can be estimated, even if it is challenging in practice. For the sake of exposition, we drop this term in the rest of the analysis. As we did for the BLP case, we propose two feasible approximations of the MPI, which don't require the knowledge (β_a, f_a) and, which can be computed in practice.

- **Local approximation.** First, we provide a local approximation, which is accurate when f_0 is close to the true density f_a. To derive this local approximation, we need to impose additional restrictions on β_0 and β_a so that $\|\beta_a - \beta_0\| = O\left(\int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv\right)$. This is the purpose of Assumption 1

Assumption 1 *We assume that* $\beta_0 = \beta_0^*$ *and* $\beta_a = \beta_a^*$ *where* (β_0^*, β_a^*) *are both pseudo true values, which maximize the conditional expectation of their respective population log-likelihoods. Namely,*

$$\beta_0^* = \underset{\tilde{\beta} \in \mathbb{R}^{K_1}}{argmax}\ \mathbb{E}\left[L(x_i, y_i, \tilde{\beta}, f_0) \middle| x_i\right] \ with\ \ L(x_i, y_i, \tilde{\beta}, f_0) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} \log(\rho_j(x_i, \tilde{\beta}, f_0))$$

$$\beta_a^* = \underset{\tilde{\beta} \in \mathbb{R}^{K_1}}{argmax}\ \mathbb{E}\left[L(x_i, y_i, \tilde{\beta}, f_a) \middle| x_i\right] \ with\ \ L(x_i, y_i, \tilde{\beta}, f_a) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} \log(\rho_j(x_i, \tilde{\beta}, f_a))$$

Now we can derive the following first order approximation of the $\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a)$ around f_0.

Proposition 1.1
Under Assumption 1, a first order expansion of $\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a)$ *around* f_0 *writes:*

$$\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a) = \int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{1ij}\beta_0 + x'_{2ij}v\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\beta_0 + x'_{2ik}v\right\}}(f_0(v) - f_a(v))dv + \frac{\partial \rho_j(x_i, \beta, f_a)}{\partial \tilde{\beta}}\bigg|_{\beta=\beta_0} (\beta_a - \beta_0) + \mathcal{R}_0$$

with $\mathcal{R}_0 = o\left(\int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv\right)$.

The proof is in Appendix 1.B. Building on this approximation, we can discretize the integrals as we did in the BLP case to circumvent the fact that we do not know f_a.

$$\mathbb{E}[\Delta_j(x_i, f_0, f_a)|x_i] \approx \sum_{l=1}^{L} \bar{\omega}_{1l}(f_a) \underbrace{\left[\rho_j(x_i, \beta_0, f_0) - \frac{\exp\{x'_{1ij}\beta_0 + x'_{2ij}v_l\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\beta_0 + x'_{2ik}v_l\right\}}\right]}_{\pi_{1,j,l}(x_i)}$$
$$+ \sum_{l=1}^{L} \bar{\omega}_{2l}(f_a) \underbrace{\frac{\partial}{\partial \beta} \left\{\frac{\exp\{x'_{1ij}\beta_0 + x'_{2ij}v_l\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\beta_0 + x'_{2ik}v_l\right\}}\right\}}_{\pi_{2,j,l}(x_i)},$$

with $\{v_l\}_{l=1,\dots,L}$ L points chosen in the domain of definition of f_a, and $\bar{\omega}_l(f_a)$ the unknown weights associated with each point. The local interval instruments, in the mixed logit case, write: $(\pi_{1,j,l}(x_i), \pi_{2,j,l}(x_i))$.

- **Global approximation.** As we did, in the BLP case, we can also write a global approximation of the corection term. To do so, we replace the unknown β_a by a known substitute β_0.[37] To circumvent the fact that f_a is unknown, we replace the integral with a finite sum. Namely, we have:

$$\mathbb{E}[\Delta_j(x_i, f_0, f_a)|x_i] \approx \sum_{l=1}^{L} \omega_l \underbrace{\left[\rho_j(x_i, \beta_0, f_0) - \frac{\exp\{x_{1ij}'\beta_0 + x_{2ij}'v_l\}}{1 + \sum_{k=1}^{J} \exp\left\{x_{1ik}'\beta_0 + x_{2ik}'v_l\right\}}\right]}_{\bar{\pi}_{j,l}(x_i)}$$

  with $\{v_l\}_{l=1,\ldots,L}$ L points chosen in the support of f_a, and $\omega_l(f_a)$ the unknown weights associated with each point. The local interval instruments, in the mixed logit case, write: $(\bar{\pi}_{1,j,l}(x_i))$

Composite hypothesis. In practice, as in the BLP model, the researcher must make a parametric assumption on the distribution of random coefficients to estimate the model. Formally, the econometrician assumes f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$, where $\tilde{\lambda}$ is a parameter that must be estimated. In applied work, researchers typically assume that f is normally distributed. The researcher may be interested in testing the validity of the specification. The mixed logit is often estimated by conditional MLE. To test the validity of the specification, the research must follow the same steps as the ones highlighted in Section 1.5. First, the researcher must estimate a pseudo-true value $\theta_0 = (\beta_0', \lambda_0')' \in \mathbb{R}^{|\theta|_0}$, which maximizes the conditional expectation of their respective population log-likelihoods under $H_0 : f \in \mathcal{F}_0$. Namely:

$$\theta_0 = \underset{\tilde{\theta} \in \mathbb{R}^{|\theta|_0}}{\operatorname{argmax}} \mathbb{E}\left[L(x_i, y_i, \tilde{\beta}, f_0(\cdot|\tilde{\lambda}))\middle|x_i\right] \text{ with } L(x_i, y_i, \tilde{\beta}, f_0(\cdot|\tilde{\lambda})) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} \log(\rho_j(x_i, \tilde{\beta}, f_0(\cdot|\tilde{\lambda}))$$

Next, we test $H_0' : \mathbb{E}[h_D(x_i)\xi_{ij}(f_0(\cdot|\tilde{\lambda}), \beta_0)] = 0$ with the moment based test exhibited above. To derive the asymptotic distribution of $S_N(h_D, f_0(\cdot|\hat{\lambda}, \hat{\beta})$, we must now take into account the parameter uncertainty stemming from the first stage estimation. As in the BLP demand model, the integrals must

[37] in our simulations, we find that the homogeneous parameters are usually close to each other even when the distributions are somewhat distant from each other.

be numerically computed to recover the theoretical probabilities implied by the model, and compute the conditional likelihood. Thus, one must also take into account the numerical approximations in the derivation of the asymptotic distribution.

1.B Proofs

1.B.1 Identification

In this subsection, we prove that under Assumption A, the distribution of random coefficients f is non-parametrically point identified.

Proof. Proof of Proposition 2.1

We want to show that under Assumptions A, the following implication holds:

$$(\tilde{f},\tilde{\beta}) = (f,\beta) \iff \mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}] = 0 \; a.s.$$
$$\iff \mathbb{E}\left[\rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x'_{1jt}\tilde{\beta}\middle| z_{jt}\right] = 0 \; a.s..$$

Step 1: First, we show that for any random permutation of indexes $j \to j'$, the following equivalence holds:

$$\mathbb{E}[\xi_{jt}|z_{jt}] = 0 \; a.s. \iff \mathbb{E}[\xi_{jt}|z_{j't}] = 0 \; a.s. \;\forall\, j'.$$

As the new indexation is done exogenously, we have for any j':

$$\mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}] = \mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}, j \to j'] \equiv \mathop{\mathbb{E}}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}] \; a.s.,$$

with $j \to j'$ indicates index j has been changed into j'. Consequently, we have:

$$\mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}] = 0 \; a.s. \iff \forall j' \; \mathop{\mathbb{E}}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}] = 0 \; a.s.$$

This last equivalence allows us to come back to the exogeneity condition assumed in Berry and Haile (2014) and in Wang (2022): $\forall k, \;\; \mathbb{E}\left[\xi_{jt}|z_{jt}, j = k\right] = 0 \; a.s.$. The only difference being that here j' is determined completely randomly. Intuitively, the exogeneity condition required for non-parametric identification of the demand functions is stronger than the one needed for the non-parametric identification of the distribution of RC.

Step 2: We now need to show the following equivalence:

$$(\tilde{f},\tilde{\beta}) = (f,\beta) \iff \forall j',\ \mathbb{E}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}] = 0\ a.s..$$

Given the random permutation $j \to j'$, which is market dependent, we must redefine our matrices and vectors as follows: $\hat{x}_t = M_t x_t$ with $(M_t)_{i,k} = \mathbf{1}\{i = j_t, k = j'_t\}$. Likewise $\hat{s}_t = M_t s_t$. M_t is a random matrix. It is straight forward to show the direct implication.

$$(\tilde{f},\tilde{\beta}) = (f,\beta) \implies \forall j',\quad \mathbb{E}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f}) - x'_{1j't}\tilde{\beta}\middle|z_{j't}\right] = \mathbb{E}_{j'}[\xi_{j't}(f,\beta)|z_{j't}] = 0\ a.s.$$

The reverse implication is much more intricate to prove and we will exploit other results in the literature. We want to show:

$$(\tilde{f},\tilde{\beta}) \neq (f,\beta) \implies \exists j' \ \Big|\ \mathbb{E}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f}) - \hat{x}'_{1jt}\tilde{\beta}\middle|z_{j't}\right] = 0 \text{ a.s. does not hold.}$$

Case 1: First, let us assume that $\tilde{f} = f$ and $\tilde{\beta} \neq \beta$, then we have:

$$\rho^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f}) - \hat{x}_{1t}\tilde{\beta} = \underbrace{\rho^{-1}(\hat{s}_t,\hat{x}_{2t},f) - x_{1t}\beta}_{\hat{\xi}_t(f,\beta)} + \hat{x}_{1t}(\beta - \tilde{\beta})$$

By assumption, we have: $P(x'_{1t}x_{1t}\ \ dp) > 0$. M_t is symmetric, idempotent and full rank. As a consequence,

$$P(\hat{x}'_{1t}\hat{x}_{1t}\ \ dp) = P(x'_{1t}M_t x_{1t}\ \ dp) = P(x'_{1t}x_{1t}\ \ dp) > 0$$

Therefore, we have $\forall\ \gamma \neq 0 \in \mathbb{R}^K$,

$$\begin{aligned} P(\gamma'\hat{x}'_{1t}\hat{x}_{1t}\gamma > 0) > P(\hat{x}'_{1t}\hat{x}_{1t}\ \ dp) > 0 &\iff P(\|\hat{x}_{1t}\gamma\|^2 > 0) > 0 \\ &\iff P(\hat{x}_{1t}\gamma \neq 0) > 0 \end{aligned}$$

Thus, $\exists j' \mid x'_{1j't}(\beta - \tilde{\beta}) = 0$ *a.s.* does not hold. To conclude, there exists j' such that:

$$\mathop{\mathbb{E}}_{j'}[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x'_{1j't}\tilde{\beta}|z_{j't}] = \underbrace{\mathop{\mathbb{E}}_{j'}[\xi_{j't}(f,\beta)|z_{j't}]}_{=0} + \underbrace{\mathop{\mathbb{E}}_{j'}[x'_{1j't}(\beta - \tilde{\beta})|z_{j't}]}_{=0 \ a.s \text{ does not hold from the completeness}}$$

Case 2: Now let us assume that $\tilde{f} \neq f$ and we want to show that $\forall \tilde{\beta} \in \mathbb{R}^k$, $\exists j'$ such that:

$$\mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - x'_{1j't}\tilde{\beta}\middle|z_{jt}\right] = 0 \ a.s. \text{ does not hold.}$$

First, let us observe that $\forall j'$,

$$\mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - x'_{1j't}\tilde{\beta}|z_{j't}\right] = \underbrace{\mathop{\mathbb{E}}_{j'}\left[\xi_{j't}(f,\beta)|z_{j't}\right]}_{=0} + \mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x'_{1j't}(\tilde{\beta} - \beta)|z_{j't}\right].$$

As a consequence, we need to show that $\exists j'$ such that $\mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x'_{1j't}(\tilde{\beta} - \beta)\right] = 0$ a.s. does not hold. From the completeness condition, a sufficient condition is: $\exists j'$ such that $\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x'_{1j't}(\tilde{\beta} - \beta) = 0$ *a.s.* does not hold.

Let $\gamma = (\tilde{\beta} - \beta)$. By contradiction, it can be easily be shown that $\rho(\hat{\delta}_t, \hat{x}_{2t}, f) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) \neq 0 \implies \exists j' \ \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) \neq \rho_j^{-1}(\hat{s}_t, \hat{x}_{2t}, f) + \gamma' x_{1j't}$. Indeed, assume that $\rho(\hat{\delta}_t, \hat{x}_{2t}, f) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) \neq 0$ and $\forall j' \ \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) = \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) + \gamma' x_{1j't}$. Then, we have: $\rho(\rho^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}), \hat{x}_{2t}, \tilde{f}) = \rho(\rho^{-1}(\hat{s}_t, \hat{x}_{2t}, f) + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) = \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) \neq \rho(\hat{\delta}_t, \hat{x}_{2t}, f) = \hat{s}_t$. Therefore, we have a contradiction.

Hence, the next step is to show that $\forall \gamma$, $\tilde{f} \neq f \implies \rho(\hat{\delta}_t, \hat{x}_{2t}, f_0) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, f) = 0$ a.s. does not hold.

To this end, we are going to exploit the identification result shown by Wang (2022). Following the notations in this paper, we define $\mu_i = \hat{x}_{1t}\Gamma + \hat{x}_{2t}v_i = \hat{x}_t\mathbf{v}$ with $\mathbf{v}_i = (\Gamma, v_i)$. Here Γ is a degenerate random variable characterized by constant c such that $P(\Gamma = c) = 1$. Let $G_{\mu|\hat{x}_t}$ the distribution of $\mu_i|\hat{x}_t$ under $f^\dagger = (c = 0, f)$ and $G_{\tilde{\mu}|\hat{x}_t}$ the distribution of $\mu_i|\hat{x}_t$ under $\tilde{f}^\dagger = (c = \gamma, \tilde{f})$. The following result is shown in Wang (2022): for any $\hat{x}_t \in Supp(\hat{x}_t)$,

$$\exists j' \mid \rho_{j'}(\hat{\delta}_t, G_{\mu|\hat{x}_t}) - \rho_{j'}(\hat{\delta}_t, G_{\tilde{\mu}|\hat{x}_t}) = 0 \text{ on open set } \mathcal{D} \subset \mathbb{R}^J \implies G_{\mu|\hat{x}_t} = G_{\tilde{\mu}|\hat{x}_t}.$$

Thanks to the real analytic property of the demand functions ρ, Wang (2022) does not require a full support assumption on $\hat{\delta}_t$.

Fix the value of $\hat{x}_t$ as follows: $\hat{x}_t = \bar{M}_t\bar{x}_t = \hat{\bar{x}}_t$. By assumption, there exists $\bar{x}_t \in Supp(x_t)$ such that $\bar{x}_t'\bar{x}_t$ is dp and $\delta_t = \bar{x}_{1t}\beta + \xi_t$ varies on an open set $\bar{\mathcal{D}}$ almost surely.These properties naturally transmit to $\hat{\bar{x}}_t$. The chosen permutation $\bar{M}_t$ doesn't matter. Given the result in Wang (2022), in order to prove that $\rho(\hat{\delta}_t, \hat{x}_{2t}, f_0) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, f) = 0$ $a.s.$ does not hold, we just need to prove that $\forall\gamma,\ \tilde{f} \neq f \implies G_{\tilde{\mu}|\hat{x}_t} \neq G_{\mu|\hat{x}_t}$. By definition (see assumption A (iv)), $\tilde{f} \neq f \implies \exists v^* \in \mathbb{R}^{K_2}\ \ \tilde{F}(v^*) \neq F(v^*)$. Take $x^* = (0_{K_1}, \hat{\bar{x}}_{2t}v^*)' = \hat{\bar{x}}_t(0_{K_1}, v^*)'$:

$$G_{\mu|\hat{\bar{x}}_t}(x^*) = P(x_t\mathbf{v}_i \leq x^*|x_t = \hat{\bar{x}}_t) = P((x_t'x_t)^{-1}x_t'x_t\mathbf{v}_i \leq (x_t'x_t)^{-1}x_t'\bar{x}_t(0_{K_1}, v^*)'|x_t = \hat{\bar{x}}_t).$$
$$= (1_{K_1}, P(v_i \leq v^*|x_t = \hat{\bar{x}}_t))' = (1_{K_1}, F(v^*))'$$

The last equality comes from independence of v_i and x_t. Likewise, $G_{\tilde{\mu}|\hat{\bar{x}}_t}(x^*) = (1\{\gamma > 0\}, \tilde{F}(v^*))'$

Therefore, $\exists x^*,\ \forall\gamma\ \ G_{\tilde{\mu}|\hat{\bar{x}}_t}(x^*) \neq G_{\mu|\hat{\bar{x}}_t}(x^*)$. Following the result in Wang (2022), we have that for all $\gamma \in \mathbb{R}^{K_1}$, $\rho(\hat{\delta}_t, \hat{x}_{2t}, f) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) = 0$ $a.s.$ does not hold, which in turn implies that for all $\gamma \in \mathbb{R}^{K_1}$, $\exists j'\ \ \rho_j^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_j^{-1}(\hat{s}_t, \hat{x}_{2t}, f) + \hat{x}_{1jt}'\gamma = 0$ $a.s.$ does not hold.

To conclude: $\forall\beta \in \mathbb{R}^k$, there exists j' such that:

$$\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x_{1j't}'(\tilde{\beta} - \beta) = 0 \ a.s. \text{ does not hold,}$$

which is what we wanted to show.

□

In Section 1.5, we used the following equivalence between the composite hypothesis and the pseudo-true value to construct the specification test.

Corollary 2.1 *Under Assumption A, and assume $h_E(z_{jt})$ and W are such that the pseudo-true value θ_0 is unique, then we have:*

$$H_0 : f \in \mathcal{F}_0 \iff \overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0).$$

Proof. Proof of Corollary 2.1

Let us assume that under specification $\mathcal{F}_0$, instruments $h_E(z_{jt})$ and weighting matrix W , the pseudo true value is unique.

- Under $H_0 : f \in \mathcal{F}_0$ and there exists λ such that $f = f_0(\cdot|\lambda)$. By the mean independence assumption on the unobserved quality ξ_{jt}, we have at the true $\theta = (\beta, \lambda)$:

  $$\xi_{jt}(f_0(.|\lambda), \beta) = \rho_j^{-1}(s_t, x_{2t}, f_0(.|\lambda)) - x'_{1jt}\beta = \xi_{jt} \implies \mathbb{E}[(\xi_{jt}(f_0(.|\lambda), \beta)h_E(z_{jt})] = 0.$$

  Thus, θ is solution to the previous minimization problem and as the solution is unique: $\theta_0 = \theta$. As a consequence, $\xi_{jt}(f_0(.|\lambda_0), \beta_0)) = \xi_{jt}$ and $\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}] = 0$ *a.s.*.

- Under an alternative specification: $f \notin \mathcal{F}_0$, we know from the Proposition 2.1 that $\forall\tilde{\theta} = (\tilde{\beta}, \tilde{\lambda})$,

  $$\mathbb{E}\left[\rho_j^{-1}(s_t, x_{2t}, f_0(.|\tilde{\lambda})) - x'_{1jt}\tilde{\beta}\middle|z_{jt}\right] = 0 \; a.s. \text{ does not hold.}$$

  In particular, the last equation doesn't hold at the true value $\tilde{\theta} = \theta_0$.

□

1.B.2 Detecting misspecification: the most powerful instrument

Proof of Proposition 3.1.

- Under $\overline{H}_0 : (f,\beta) = (f_0,\beta_0)$. By assumption, the data are i.i.d. across markets, $\mathbb{E}[\|\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\|^2] = \frac{1}{J}\mathbb{E}[\sum_j \|\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\|^2] < +\infty$, the CLT applies:

$$\frac{1}{\sqrt{TJ}}\sum_{j,t} h_D(z_{jt})\xi_{jt}(f_0,\beta_0) = \frac{1}{\sqrt{TJ}}\sum_{j,t} h_D(z_{jt})\xi_{jt} \underset{T\to+\infty}{\longrightarrow} \mathcal{N}(0,\tilde{\Omega}_0),$$

with:

$$\begin{aligned}
\tilde{\Omega}_0 &= \mathbb{E}\left[\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J} h_D(z_{jt})\xi_{jt}\right)\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J} h_D(z_{jt})\xi_{jt}\right)'\right] \\
&= \frac{1}{J}\mathbb{E}\left[\sum_{j=1}^{J} h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2 + \sum_{j=1}^{J}\sum_{k\neq j} h_D(z_{jt})h_D(z_{kt})'\xi_{jt}\xi_{kt}\right] \\
&= \frac{1}{J}\mathbb{E}\left[\sum_{j=1}^{J} h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2\right] + \frac{1}{J}\sum_{j=1}^{J}\sum_{k\neq j}\mathbb{E}\left[h_D(z_{jt})h_D(z_{kt})'\underbrace{\mathbb{E}[\xi_{jt}\xi_{kt}|z_{jt},z_{kt}]}_{=0}\right] \\
&= \mathbb{E}\left[h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2\right] \\
&= \Omega_0.
\end{aligned}$$

Third line comes from $\xi_{jt} \perp\!\!\!\perp \xi_{kt}|z_t$. By assumption, Ω_0 has a full rank. Thus, we have by the CMT:

$$S_T(h_D,f_0,\beta_0) = TJ\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\right)'\hat{\Omega}_0^{-1}\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\right) \underset{T\to+\infty}{\overset{d}{\longrightarrow}} \chi^2_{|h_D|_0}.$$

- Under $H_a' :\ \mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right] \neq 0$. The data are i.i.d. across markets, by the law of large numbers: $\frac{1}{TJ}\sum_{j,t} h_D(z_{jt})\xi_{jt}(f_0,\beta_0) \overset{\mathbb{P}}{\to} \mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right]$. It follows by the continuous mapping theorem:

$$\begin{aligned}
\frac{S_T(h_D,f_0,\beta_0)}{T} &\overset{\mathbb{P}}{\to} J\mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right]'\Omega_0^{-1}\mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right] \\
&= J\underbrace{\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right]'\Omega_0^{-1}\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right]}_{\kappa(h_D,f_0,\beta_0)}
\end{aligned}$$

Under H'_a, $\kappa(h_D, f_0, \beta_0)$ is strictly positive because Ω_0 is positive definite. Thence,

$$\begin{aligned}\forall q \in \mathbb{R},\ \ \underset{T\to\infty}{lim}\ \mathbb{P}(S_T(h_D, f_0, \beta_0) > q) &= \underset{T\to\infty}{lim}\ \mathbb{P}\left(\frac{S(h_D, f_0, \beta_0) - q}{T} > 0\right)\\ &= \mathbb{P}(J\kappa(h_D, f_0, \beta_0) > 0)\\ &= 1,\end{aligned}$$

where the second equality holds because convergence in probability implies convergence in distribution.

□

Proof of Proposition 3.2. To shorten notations, let $\xi_{jt0} \equiv \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)$, $\xi_{jta} \equiv \xi_{jt}(f_a, \beta_a)$ and ξ_{t0} and ξ_{ta} their stacked versions over j. Likewise, we define $h_D(z_t) = (h_D(z_{1t}), ..., h_D(z_{Jt}))'$. Under $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$, the asymptotic slope of the test writes:

$$\begin{aligned} c_{h_D}(f_a, \beta_a) &= \mathbb{E}\left(\sum_j \xi_{jt0} h_D(z_{jt})\right)' \mathbb{E}\left(\left(\sum_j \xi_{jt0} h_D(z_{jt})\right)\left(\sum_{j'} \xi_{j't0} h_D(z_{j't})\right)'\right)^{-1} \mathbb{E}\left(\sum_j \xi_{jt0} h_D(z_{jt})\right)\\ &= \mathbb{E}(\xi'_{t0} h_D(z_t)) \mathbb{E}(h_D(z_t)'\xi_{t0}\xi'_{t0} h_D(z_t))^{-1} \mathbb{E}(h_D(z_t)'\xi_{t0})\\ &= \mathbb{E}(\Delta_{0,a}^{\xi_t}{}' h_D(z_t)) \mathbb{E}(h_D(z_t)'\mathbb{E}(\xi_{t0}\xi'_{t0}|z_t) h_D(z_t))^{-1} \mathbb{E}(h_D(z_t)'\Delta_{0,a}^{\xi_t}). \end{aligned}$$

Third line comes from $\mathbb{E}(\Delta_{0,a}^{\xi_t}{}' h_D(z_t)) = \mathbb{E}((\xi_{t0} - \xi_{ta})' h_D(z_t)) = \mathbb{E}(\xi'_{t0} h_D(z_t))$ because ξ_{ta} is the true structural error. Then the slope of the test taking $h^*_D(z_t) = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{-1}\mathbb{E}(\Delta_{0,a}^{\xi_t}|z_t)$ is equal to:

$$c_{h^*_D}(f_a, \beta_a) = \mathbb{E}\left(\mathbb{E}(\Delta_{0,a}^{\xi_t}|z_t)'\mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{-1}\mathbb{E}(\Delta_{0,a}^{\xi_t}|z_t)\right)$$

To finish the proof, we must show that for any set of instruments h_D, we have: $c_{h^*_D}(f_a, \beta_a) \geq c_{h_D}(f_a, \beta_a)$.

Denote $\tilde{h}_D(z_t) = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{1/2} h_D(z_t)$ and $\tilde{h}^*_D(z_t) = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{1/2} h^*_D(z_t)$.With these new notations, we have:

$$\begin{aligned} c_{h^*_D}(f_a, \beta_a) - c_{h_D}(f_a, \beta_a) &= \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}^*_D(z_t)\right) - \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}_D(z_t)\right)\mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)^{-1}\mathbb{E}\left(\tilde{h}_D(z_t)'h^*_D(z_t)\right)\\ &= G'\begin{pmatrix}\mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}^*_D(z_t)\right) & \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}_D(z_t)\right)\\ \mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}^*_D(z_t)\right) & \mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)\end{pmatrix} G\\ &= G'\mathbb{E}\left(\tilde{H}\tilde{H}'\right) G \geq 0, \end{aligned}$$

with $\tilde{H} = (\tilde{h}_D^*(z_t), \tilde{h}_D(z_t))'$ and $G = \left(1, -\mathbb{E}\left(\tilde{h}_D^*(z_t)'\tilde{h}_D(z_t)\right)\mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)^{-1}\right)'$.

Special case: when we assume for $k \neq j$, $\xi_{jt} \perp\!\!\!\perp \xi_{kt}|z_t$, and take $\hat{\Omega}_0 = \frac{1}{JT}\sum_{j,t}\xi_{jt0}^2 h_D(z_{jt})h_D(z_{jt})'$ as our weighting matrix (as we do for illustrations purposes in the main text), we find that the slope under $\overline{H}_a$ writes:

$$c_{h_D}(f_a, \beta_a) = \mathbb{E}(\Delta_{0,a}^{\xi_{jt}} h_D(z_{jt}))'\mathbb{E}(h_D(z_{jt})h_D(z_{jt})'\mathbb{E}(\xi_{jt0}^2|z_{jt}))^{-1}\mathbb{E}(h_D(z_{jt})\Delta_{0,a}^{\xi_{jt}}).$$

Using the same arguments as previously, one can show that a maximizer of the slope of the test is obtained by taking $h_D^*(z_{jt}) = \mathbb{E}(\xi_{jt}^2|z_{jt})^{-1}\mathbb{E}(\Delta_{0,a}^{\xi_{jt}}|z_{jt})$.

□

Proof of Proposition 3.3.

Under Assumption A, Proposition 2.1 implies the following:

$$\begin{aligned}
\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0) &\implies \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}] \neq 0 \;\; a.s. \\
&\implies \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]^2 > 0 \; a.s. \\
&\implies \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]^2\right] > 0 \\
&\implies \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0, \beta_0)\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]|z_{jt}]\right] > 0 \\
&\implies \mathbb{E}\left[\xi_{jt}(f_0, \beta_0)\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]\right] > 0 \\
&\implies \overline{H}_a' : \; \mathbb{E}\big[\xi_{jt}(f_0, \beta_0)\underbrace{\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]}_{h_D^*(z_{jt})}\big] \neq 0.
\end{aligned}$$

Under the same assumptions as 3.1, we have the following:

$$\overline{H}_a' : \; \mathbb{E}\left[\xi_{jt}(f_0, \beta_0)h_D^*(z_{jt})\right] \neq 0 \implies \forall q \in \mathbb{R}^+, \; \mathbb{P}(S_T(h_D^*, \mathcal{F}_0, \hat{\theta}) > q) \to 1.$$

□

Proof of Proposition 3.4.

Let $\mathcal{H}$ the set of measurable functions of z_{jt}, we want to show under $\bar{H}_a$:

$$\forall \alpha \in \mathbb{R}^*, \ \ \alpha \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] \in \arg\max_{h \in \mathcal{H}} \operatorname{corr}(\xi_{jt}(f_0, \beta_0), h(z_{jt})).$$

We proceed in 2 steps. First, we derive the upper bound by showing that for any $h \in \mathcal{H}$, we have:

$$\operatorname{corr}\left(\xi_{jt}(f_0, \beta_0), h(z_{jt})\right) \leq \sqrt{\frac{\operatorname{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\operatorname{var}(\xi_{jt}(f_0, \beta_0))}}.$$

To do so, we use the definition of the conditional expectation and the Cauchy Schwarz inequality. First notice that we have: $\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] = \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]$. By definition of the conditional expectation, we have for any $h \in \mathcal{H}$,

$$\mathbb{E}[h(z_{jt})\xi_{jt}(f_0, \beta_0)] = \mathbb{E}[h(z_{jt})\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]].$$

It follows that:

$$\left|\operatorname{cov}\left(h(z_{jt}), \xi_{jt}(f_0, \beta_0)\right)\right| = \operatorname{cov}\left(h(z_{jt}), \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]\right) \leq \sqrt{\operatorname{var}(h(z_{jt}))\operatorname{var}\left(\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]\right)}.$$

The inequality comes from the Cauchy Schwarz inequality. The result follows by using the definition of the correlation coefficient.

Second, we show that the upper bound is reached by taking for any $\alpha \in \mathbb{R}^*$, $h_D^*(z_{jt}) = \alpha \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]$.

$$\begin{aligned} \operatorname{cov}\left(\xi_{jt}(f_0, \beta_0), \alpha \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right) &= \alpha \operatorname{cov}\left(\Delta_{0,a}^{\xi_{jt}}, \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right) \\ &= \alpha \operatorname{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right). \end{aligned}$$

Consequently,

$$\operatorname{corr}\left(\xi_{jt}(f_0, \beta_0), h_D^*(z_{jt})\right) = \frac{\alpha}{\sqrt{\alpha^2}} \sqrt{\frac{\operatorname{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\operatorname{var}(\xi_{jt}(f_0, \beta_0))}} \implies \left|\operatorname{corr}\left(\xi_{jt}(f_0, \beta_0), h_D^*(z_{jt})\right)\right| = \sqrt{\frac{\operatorname{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\operatorname{var}(\xi_{jt}(f_0, \beta_0))}}.$$

□

Connection with optimal instruments

In the parametric case and assuming that the model is well specified ($f \in \mathcal{F}_0$) the BLP parameter θ_0 is identified by the following non-linear conditional moment restriction $\mathbb{E}[\xi_{jt}(\theta_0)|z_{jt}] = 0$. The derivation of the optimal instruments in this context has been studied by Amemiya (1974). For an arbitrary choice of $h_E(z_{jt})$, the GMM estimator with the 2-step efficient weighting matrix has the following asymptotic distribution:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, (\Gamma(\mathcal{F}_0, \theta, h_E)'\Omega(\mathcal{F}_0, h_E)^{-1}\Gamma(\mathcal{F}_0, \theta, h_E))^{-1}\right),$$

with the same notations as previously:

$$\Omega(\mathcal{F}_0, h_E) = \mathbb{E}\left[\left(\sum_j \xi_{jt}(\theta)h_E(z_{jt})\right)\left(\sum_j h_E(z_{jt})\xi_{jt}(\theta)\right)'\right]$$

$$\Gamma(\mathcal{F}_0, \theta_0, h_E) = \mathbb{E}\left[\sum_j h_E(z_{jt})\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}'}\right].$$

For the sake of exposition, we will assume that unobserved demand shock ξ_{jt} is independent across observations, namely: $\mathbb{E}\left[\xi_{jt}(\theta_0)\xi_{j't}(\theta)|z_t\right] = 0$ for $j \neq j'$. The general case extends naturally. The optimal instrument $h^*_E(z_{jt})$ are chosen to minimize the asymptotic variance covariance matrix. We derive the form of the optimal instruments in the context of BLP by applying well known results in Chamberlain (1987) and Amemiya (1974)

Lemma 2.2 *Optimal instruments in the BLP model.*

In our setting and assuming $f \in \mathcal{F}_0$, the optimal instruments $h^*_E(z_{jt})$ write:

$$h^*_E(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]^{-1}\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}\bigg|z_{jt}\right].$$

and the corresponding efficiency bound (obtained by setting $h_E = h^*_E$) writes:

$$V^* = \mathbb{E}\left[\sum_j \mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}\bigg|z_{jt}\right]\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}\bigg|z_{jt}\right]'\mathbb{E}[\xi_{jt}(\theta)^2|z_{jt}]^{-1}\right]^{-1}$$

Proof. To shorten the notations, we denote: $\sigma^2(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]$ and $d(z_{jt}) = \mathbb{E}\left[\frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}\middle|z_{jt}\right]$. Likewise, we define:

$$\Omega_0(h_E) = \mathbb{E}\left[\sum_j \mathbb{E}[\xi_{jt}(\theta_0)^2|z_{jt}]h_E(z_{jt})h_E(z_{jt})'\right].$$

We want to prove that for any set of instruments $h_E(z_{jt})$ that $V^*(z_{jt}) - \Gamma_0(h_E)'\Omega_0(h_E)^{-1}\Gamma_0(h_E)$ matrix is semi definite positive.

$$\begin{aligned}
&V^*(z_{jt}) - \Gamma_0(h_E)\Omega_0(h_E)^{-1}\Gamma_0(h_E)' = \\
&= \mathbb{E}\left[\sum_j d(z_{jt})d(z_{jt})'\sigma^2(z_{jt})\right] - \mathbb{E}\left[\sum_j \frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}h_E(z_{jt})'\right]\Omega_0(h_E)^{-1}\mathbb{E}\left[\sum_j h_E(z_{jt})\frac{\partial\xi_{jt}(\theta_0)}{\partial\tilde{\theta}}'\right] \\
&= \mathbb{E}\left[\sum_j d(z_{jt})d(z_{jt})'\sigma^{-2}(z_{jt})\right] - \mathbb{E}\left[\sum_j d(z_{jt})h_E(z_{jt})'\right]\mathbb{E}\left[\sum_j \sigma^2(z_{jt})h_E(z_{jt})h_E(z_{jt})'\right]\mathbb{E}\left[\sum_j h_E(z_{jt})d(z_{jt})'\right] \\
&= \mathbb{E}\left[\tilde{D}(z_{jt})'\tilde{D}(z_{jt})\right] - \mathbb{E}\left[\tilde{D}(z_{jt})'\tilde{H}_E(z_{jt})\right]\mathbb{E}\left[\tilde{H}_E(z_{jt})'\tilde{H}_E(z_{jt})\right]^{-1}\mathbb{E}\left[\tilde{H}_E(z_{jt})'\tilde{D}(z_{jt})\right].
\end{aligned}$$

The second line comes from law of iterated expectations. The third line is a matricial way to rewrite the second line. $\tilde{D}(z_{jt})$ a matrix which stacks $d(z_{jt})/\sigma(z_{jt})$ over the set of products (each line corresponds to one product j). Likewise, let $\tilde{H}_E(z_{jt})$ a matrix which stacks $h_E(z_{jt})\sigma(z_{jt})$ over the set of products (each line corresponds to one product j). Now let us define the following matrices.

$$\tilde{X} = \begin{pmatrix}\tilde{D}(z_{jt}) & \tilde{H}_E(z_{jt})\end{pmatrix} \text{ and } \tilde{M} = \begin{pmatrix} I_{|\theta_0|} & -\mathbb{E}\left[\tilde{D}(z_{jt})'\tilde{H}_E(z_{jt})\right]\mathbb{E}\left[\tilde{H}_E(z_{jt})'\tilde{H}_E(z_{jt})\right]^{-1}\end{pmatrix}'$$

We have: $V^*(z_{jt}) - \Gamma_0(h_E)\Omega_0(h_E)^{-1}\Gamma_0(h_E) = \tilde{M}'\mathbb{E}[\tilde{X}'\tilde{X}]\tilde{M}$.

The matrix above is clearly semi definite positive. □

1.B.3 Feasible most powerful instrument

Local approximation of the MPI

Proof of Proposition 4.1. First, we define $s_t^0 = \rho(\delta_t, x_{2t}, f_0(.|\lambda_0))$ with δ_t the true mean utility. From lemma 2.4 ρ^{-1} is $\mathcal{C}^\infty$ and in particular, ρ^{-1} is $\mathcal{C}^1$. Thus, the Taylor expansion of $\rho^{-1}(s_t^0, x_{2t}, f_0(.|\lambda_0))$ around s_t writes:

$$\rho^{-1}(s_t^0, x_{2t}, f_0(.|\lambda_0)) = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)) + \left.\frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}{\partial s}\right|_{s=s_t} (s_t^0 - s_t) + o\Big(||s_t^0 - s_t||\Big)$$
$$\delta_t = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0)) + \left.\frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0))}{\partial s}\right|_{s=s_t} (s_t^0 - s_t) + o\Big(||s_t^0 - s_t||\Big)$$

We now derive an expression for the first derivative of the inverse function. We make use of lemma 2.5: for any $\delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta}$ is invertible.

$$\frac{\partial \rho(\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0)), x_{2t}, f_0(.|\lambda_0))}{\partial s} = I_J \iff \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0))}{\partial s}\left(\frac{\partial \rho(\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0)), x_{2t}, f_0(\cdot|\lambda_0))}{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0))}\right) = I_J$$
$$\iff \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0))}{\partial s} = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0(\cdot|\lambda_0))}{\partial \delta}\right)^{-1}$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$.Consequently,

$$\underbrace{\rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)) - \delta_t}_{\Delta(s_t, x_{2t}, f_0, f_a)} = -\left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0(.|\lambda_0))}{\partial \delta}\right)^{-1} (s_t^0 - s_t) + o\Big(||s_t^0 - s_t||\Big) \tag{1.B.16}$$

with $\delta_t^0 = \rho_j^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$

Now let us show that there exists a constant M such that $||s_t^0 - s_t|| \leq M\tau(f_0(.|\lambda_0) - f_a)$. with $\tau(f_0 - f_a) = \int_{\mathbb{R}^{K_2}} |f_0(v|\lambda_0) - f_a(v)|dv$. Norms are equivalent in a finite vectorial space and without loss of generality, we will derive the results with the L_1 norm. By definition:

$$s_t^0 - s_t = \int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta_t + x_{2t}v\}}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2jk}v\right\}} (f_0(v|\lambda_0) - f_a(v))dv$$

Taking the L_1 norm of this vector:

$$||s_t^0 - s_t||_1 = \sum_{j=1}^{J} \left| \int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta_{jt} + x_{2jt}' v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt} + x_{2jk}' v\right\}} (f_0(v|\lambda_0) - f_a(v))dv \right|$$
$$\leq \sum_{j=1}^{J} \int_{\mathbb{R}^{K_2}} \underbrace{\left| \frac{\exp\{\delta_{jt} + x_{2jt}' v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt} + x_{2jk}' v\right\}} \right|}_{\leq 1} |f_0(v|\lambda_0) - f_a(v)|dv.$$
$$\leq J \int_{\mathbb{R}^{K_2}} |f_0(v|\lambda_0) - f_a(v)|dv = J\tau(f_0(.|\lambda_0) - f_a).$$

This proves the statement. As a consequence, we have: $||s_t^0 - s_t||_1 = O(\tau(f_0(.|\lambda_0) - f_a))$ and $o\left(||s_t^0 - s_t||\right) = o(\tau(f_0(.|\lambda_0) - f_a))$.

The problem with the term $s_t^0 - s_t$ is that it is an expression of δ_t which we do not know under misspecification. As we want to be able to compute this approximation of the error term, it is not convenient in practice to have an expression which depends on δ_t. On the other hand, we know δ_t^0 and thus, the simple idea that we exploit is to take a Taylor expansion of the term above around δ_t^0. First, let us remark that from equation 1.B.16, we have that:

$$||\delta_t - \delta_t^0|| = ||\delta_t - \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)|| = O(||s_t^0 - s_t||) = O(\tau(f_0(.|\lambda_0) - f_a)).$$

Now let us take the Taylor expansion of $s_t^0 - s_t$ around δ_t^0:

$$s_t^0 - s_t = \int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta_t^0 + x_{2t} v\}}{1 + \sum_{k=1}^{J} \exp\left\{\tilde{\delta}_{kt} + x_{2jk}' v\right\}} (f_0(v|\lambda_0) - f_a(v))dv$$
$$+ \underbrace{\int_{\mathbb{R}^{K_2}} \frac{\partial}{\partial \delta'} \left\{ \frac{\exp\{\delta_t^0 + x_{2t} v\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}^0 + x_{2jk}' v\right\}} \right\} (\delta_t - \delta_t^0)(f_0(v|\lambda_0) - f_a(v))dv}_{B} + o\Big(||\delta_t - \delta_t^0||\Big) .$$

From what precedes, we know that $o\left(||\delta_t - \delta_t^0||\right) = o(\tau(f_0(.|\lambda_0) - f_a))$. Now, let us show that term B in the previous expansion is also $o(\tau(f_0(.|\lambda_0) - f_a))$. Again taking the L_1 norm:

$$||B||_1 = \sum_{j=1}^{J}\left|\sum_{l=1}^{J}\int_{\mathbb{R}^{K_2}} \frac{\partial}{\partial \delta_l}\left\{\frac{\exp\{\delta^0_{jt}+x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\tilde{\delta}_{kt}+x'_{2jk}v\right\}}\right\}(\delta_{lt}-\delta^0_{lt})(f_0(v|\lambda_0)-f_a(v))dv\right|$$

$$\leq \sum_{j=1}^{J}\sum_{l=1}^{J}\int_{\mathbb{R}^{K_2}} \underbrace{\left|\frac{\partial}{\partial \delta_l}\left\{\frac{\exp\{\delta^0_{jt}+x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{kt}+x'_{2jk}v\right\}}\right\}\right|}_{\leq 1}|\delta_{lt}-\tilde{\delta}_{lt}||f_0(v|\lambda_0)-f_a(v)|dv$$

$$\leq J^2\tau(f_0(.|\lambda_0)-f)O(\tau(f_0(.|\lambda_0)-f_a)) = O(\tau(f_0(.|\lambda_0)-f_a)^2) = o(\tau(f_0(.|\lambda_0)-f_a))\,.$$

Thus, $||B||_1 = o(\tau(f_0(.|\lambda_0)-f_a))$ and by combining all the results together, we get the final result. When $f_0(.|\lambda_0)$ gets ''close " to f_a, we have the following approximation:

$$\Delta(s_t, x_{2t}, f_0, f_a) = \left(\frac{\partial\rho(\delta^0_t, x_{2t}, f_0(.|\lambda_0))}{\partial\delta}\right)^{-1}\int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta^0_t + x_{2t}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{kt}+x'_{2jk}v\right\}}(f_a(v)-f_0(v|\lambda_0))dv$$
$$+\, o(\tau(f_a - f_0(.|\lambda_0)))\,,$$

with $\delta^0_t = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$ and $\tau(f_a - f_0(.|\lambda_0)) = \int_{\mathbb{R}^{K_2}}|f_a(v)-f_0(.|\lambda_0)(v)|dv\,.$ □

Global approximation of the MPI

Lemma 2.3 *Analytical expression for $\Delta_j(s_t, x_{2t}, f_0, f_a)$. Let $\delta^0_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_0)$ and $\delta^a_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_a)$. We have the following:*

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}}\frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^a_{kt}+x'_{2kt}v\right\}}f_a(v)dv}{\int_{\mathbb{R}^{K_2}}\frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2kt}v\right\}}f_0(v)dv}\right)\,.$$

Proof of Lemma 2.3.

$$1 = \frac{\rho_j(\delta_t, x_{2t}, f_a)}{\rho_j(\delta_t^0, x_{2t}, f_0)} \iff 1 = \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta_{jt}+x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{kt}^a+x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{\delta_{jt}^0+x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{kt}^0+x'_{2kt}v\right\}} f_0(v)dv}$$

$$\iff \frac{\exp\{\delta_{jt}^0\}}{\exp\{\delta_{jt}^a\}} = \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x_{2t}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{kt}^a+x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{jt}^0+x'_{2kt}v\right\}} f_0(v)dv}$$

$$\iff \Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{kt}^a+x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\left\{\delta_{jt}^0+x'_{2kt}v\right\}} f_0(v)dv}\right).$$

□

Approximation of the MPI in the mixed logit case

Proof of Proposition 1.1. By definition, we have:

$$g_j(x_i, \cdot, f) : \mathbb{R}^{K_1} \to [0,1]$$
$$\tilde{\beta} \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{ij1}\tilde{\beta} + x'_{2ij}v\right\}}{1+\sum_{k=1}^J \exp\left\{x'_{ik1}\tilde{\beta} + x'_{2ik}v\right\}} f(v)dv$$

g is $\mathcal{C}^\infty$ on $\mathbb{R}^{K_1}$. Thus, we can take a first order Taylor expansion of $g_j(x_i, ., f_a)$ around β_0:

$$g_j(x_i, \beta_a, f_a) = g_j(x_i, \beta_0, f_0) + \left.\frac{\partial g_j(x_i, \tilde{\beta}, f_0)}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta_0} (\beta_a - \beta_0) + o(||\beta_a - \beta_0||)$$

This yields immediately,

$$g_j(x_i, \beta_0, f_0) - g_j(x_i, \beta_a, f_a) = \int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{1ij}\beta_0 + x'_{2ij}v\}}{1+\sum_{k=1}^J \exp\left\{x'_{1ik}\beta_0 + x'_{2ik}v\right\}} (f_0(v) - f_a(v))dv +$$
$$\left.\frac{\partial g_j(x_i, \tilde{\beta}, f_a)}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta_0} (\beta_a - \beta_0) + o(||\beta_a - \beta_0||)$$

Finally, we need to show $||\beta_a - \beta_0|| = o\left(\int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv\right)$. From Assumption 1 by definition of β_0^* and β_a^*, we have $f_0 = f_a \Rightarrow \beta_0^* = \beta_a^*$. Going further it can be shown using the Kullback divergence that for any $e_1 > 0$ such that $\int_{\mathbb{R}K_2} |f_0(v) - f_a(v)|dv < e_1$ there exists some $e_2 > 0$ such that $||\beta_0^* - \beta_a^*|| < e_2$. In other words if $\int_{\mathbb{R}K_2} |f_0(v) - f_a(v)|dv$ is small then $||\beta_a^* - \beta_0^*||$ is also small, thus any small o of $||\beta_a - \beta_0|| = ||\beta_a^* - \beta_0^*||$ can be replaced by a small o of $\int_{\mathbb{R}K_2} |f_0(v) - f_a(v)|dv$ even if the two quantities are not proportional. Consequently

$$
\begin{aligned}
g_j(x_i, \beta_0, f_0) - g_j(x_i, \beta_a, f_a) = & \int_{\mathbb{R}^{K_2}} \frac{\exp\{x_{1ij}'\beta_0 + x_{2ij}'v\}}{1 + \sum_{k=1}^J \exp\left\{x_{1ik}'\beta_0 + x_{2ik}'v\right\}} (f_0(v) - f_a(v))dv + \\
& \left.\frac{\partial g_j(x_i, \tilde{\beta}, f_a)}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta_0} (\beta_a - \beta_0) + o\left(\int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv\right)
\end{aligned}
$$

□

1.B.4 Specification Test: composite hypothesis

In this section, we prove theorem 5.1, which is the main asymptotic result of the paper. The section is organized as follows. First, we establish the equivalence between the moment condition around which we build our test $\mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)h_D(z_{jt})\right] = 0$ and the one characterizing $H_0' : \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)h_D(z_{jt})\right] = 0$. Then, we introduce the notations used in the proofs and we decompose $\widehat{\xi}$ according to the BLP approximations. Second we provide technical lemmas which prove that under the assumptions in E, the BLP approximations vanish asymptotically. Third, we prove that the BLP estimator is consistent and asymptotically normal. Finally, we prove the main theorem and we show that under the null the test is pivotal in the 2 polar cases described in the main text.

Equivalence between moment conditions

Let $h_D(z_{jt})$ our detection instruments. For conciseness, we omit the dependence in f_0 and denote $\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = \xi_{jt}(\theta_0)$. We want to prove that the following two moment conditions are equivalent:

$$\mathbb{E}\left[\xi_{jt}(\theta_0)h_D(z_{jt})\right] = 0 \iff \mathbb{E}\left[\sum_{j=1}^{J}\xi_{jt}(\theta_0)h_D(z_{jt})\right] = 0$$

Let R_t a categorial random variable which exogenously selects a product j with probability $\frac{1}{J}$. Formally, we have $(\xi_{jt}(\theta_0), z_{jt}) \perp\!\!\!\perp R_{jt}$. By construction, we have:

$$\begin{aligned}\mathbb{E}\left[\xi_{jt}(\theta_0)h_D(z_{jt})\right] &= \sum_{k=1}^{J}\mathbb{E}\left[\xi_{kt}(\theta_0)h_D(z_{kt})R_{kt}\right] = \sum_{k=1}^{J}\mathbb{E}\left[\xi_{kt}(\theta_0)h_D(z_{kt})\right]\mathbb{E}[R_{kt}] \\ &= \frac{1}{J}\mathbb{E}\left[\sum_{k=1}^{J}\xi_{kt}(\theta_0)h_D(z_{kt})\right]\end{aligned}$$

Second line results from independence of $(\xi_{jt}(\theta_0), z_{jt})$ and R_{jt}. This proves the result.

Notations

In the proofs, we will adopt the following notations. If the derivations are done under the parametric assumption $H_0 : f \in \mathcal{F}_0$ then we omit the dependence in f_0 and interchangeably use $\xi_{jt}(f_0(.|\lambda), \beta)$ and $\xi_{jt}(\theta)$. We also omit the dependence of the BLP pseudo true value in W and $h_E(z_{jt})$ [38]. Then define the following objectives of the GMM minimization

$$\begin{aligned}\hat{\mathcal{Q}}_T(\tilde{\theta}) &= \left(\frac{1}{T}\sum_{j,t}\widehat{\xi}_{jt}(\tilde{\theta})h_E(z_{jt})\right)'\hat{W}\left(\frac{1}{T}\sum_{j,t}\widehat{\xi}_{jt}(\tilde{\theta})h_E(z_{jt})\right) \\ \mathcal{Q}_T(\tilde{\theta}) &= \left(\frac{1}{T}\sum_{j,t}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right)'\hat{W}\left(\frac{1}{T}\sum_{j,t}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right) \\ \mathcal{Q}(\tilde{\theta}) &= \mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right]'W\mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right]\end{aligned}$$

[38] The BLP pseudo true value depends on W and $h_E(z_{jt})$ when the model is misspecified

We also define the following moments

$$\hat{g}_T(\tilde{\theta},h) = \frac{1}{T}\sum_{jt}\hat{\xi}_{jt}(\tilde{\theta})h(z_{jt})$$
$$g_T(\tilde{\theta},h) = \frac{1}{T}\sum_{jt}\xi_{jt}(\tilde{\theta})h(z_{jt})$$
$$g(\tilde{\theta},h) = \mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h(z_{jt})\right]$$

And recall the definition of $\Gamma(\mathcal{F}_0,\tilde{\theta},h)$ which is used interchangeably with $\Gamma(\tilde{\theta},h)$

$$\hat{\Gamma}_T(\tilde{\theta},h) = \frac{1}{T}\sum_{j,t}h(z_{jt})\frac{\partial}{\partial\theta}\widehat{\xi}_{jt}(\tilde{\theta})'$$
$$\Gamma_T(\tilde{\theta},h) = \frac{1}{T}\sum_{j,t}h(z_{jt})\frac{\partial}{\partial\theta}\xi_{jt}(\tilde{\theta})'$$
$$\Gamma(\tilde{\theta},h) = \mathbb{E}\left[\sum_{j}h(z_{jt})\frac{\partial}{\partial\theta}\xi_{jt}(\tilde{\theta})'\right]$$

Furthermore, unless specified, all limits are taken with respect to T; Additionally, we denote by the expression $X = o_P(T^{\kappa})$ a random variable or statistic which is asymptotically degenerate of order T^a, ie $X = o_P(T^{\kappa}) \Leftrightarrow \forall e > 0\ \mathbb{P}(|X|T^{-\kappa} > e) \underset{T\to\infty}{\to} 0$, and denote by $X = O_p(T^{\kappa})$ a random variable which is (bounded in probability) of order T^{κ}, ie $\forall e_1 > 0 \exists e_2 > 0, \exists T_N : \forall T \geqslant T_N\ \mathbb{P}(|X|T^{-\kappa} > e_2) < e_1$. Properties of o_P and O_P random variables are used throughout these proofs.

Feasible Structural Error and BLP approximations

We now decompose the difference between the true structural error $\xi_{jt}(\tilde{\theta})$ and the feasible structural error $\hat{\xi}_{jt}(\tilde{\theta})$ in terms of the different approximations involved in the derivation of the feasible structural error $\hat{\xi}_{jt}(\tilde{\theta})$. In market t given an assumption $\mathcal{F}_0$, a parameter $\tilde{\lambda}$, market shares s_t and product characteristics with preference heterogeneity x_{2t} there exists a unique $\delta_t \in \mathbb{R}^J$ such that $s_t = \rho(\delta_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ (Brouwer's fixed point theorem, see Berry (1994)) so that $\delta_t = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$. There is no closed form for $\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ so the NFP algorithm is used. Denote as C the contraction used to find the mean utilities which solve the demand equal market share constraint

$$C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) : \delta \in \mathbb{R}^J \mapsto \delta + log(s_t) - log(\rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda})))$$

So that for some starting mean utility $\delta_0 \in \mathcal{B} \subset \mathbb{R}^J$ where $\mathcal{B}$ is bounded, the mean utility obtained via NFP at the limit is equal to the unique vector which solves the constraint

$$\delta_t(f_0(\cdot|\tilde{\lambda})) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) = \lim_{H\to\infty} C^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$$

Similarly the error generated by $(f_0(\cdot|\tilde{\lambda}, \tilde{\beta}))$ can be obtained from NFP at the limit

$$\xi_t(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) = \delta_t(f_0(\cdot|\tilde{\lambda})) - x_{1t}\tilde{\beta} = \lim_{H\to\infty} C^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - x_{1t}\tilde{\beta}$$

This way we obtain a vector of mean utilities for each market t. There are 3 approximations to consider, market shares are not truly observed, the demand integral has to be simulated, and the contraction is never taken to its limit, so define $\widehat{\xi}(f_0, \tilde{\lambda}))$ $\widehat{\delta}(f_0, \tilde{\lambda}))$ and $\hat{C}$ for some starting value δ_0

$$\widehat{\xi}_t(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) = \hat{C}^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - x_{1jt}\tilde{\beta}, \qquad \widehat{\delta}(f_0, \tilde{\lambda})) = \hat{C}^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0, \tilde{\lambda}))$$
$$\hat{C} : \delta \mapsto \delta + log(\hat{s}_t) - log(\hat{\rho}(\delta, x_{2t}, f_0(\cdot|\lambda_0)))$$

Consequently we decompose the difference between the error generated by $(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$ and its feasible approximation into 3 differences

$$\begin{aligned}
\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) - \widehat{\xi}_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) &= \delta_{jt}(f_0(\cdot|\tilde{\lambda})) - \widehat{\delta}_{jt}(f_0(\cdot|\tilde{\lambda})) \\
&= \lim_{H\to\infty} C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - \hat{C}_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&= \lim_{H\to\infty} C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\quad + C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - C_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\quad + C_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \hat{C}_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\equiv \rho_j^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - D_j(\rho, s_t, \tilde{\lambda}) \\
&\quad + D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\theta}) \\
&\quad + D_j(\rho, \hat{s}_t, \tilde{\theta}) - D_j(\hat{\rho}, \hat{s}_t, \tilde{\theta})
\end{aligned}$$

In the fourth line, we simply introduce shortened notations for the same objects.

Technical Lemmas

The 1st and 2nd lemma establish the smoothness of ρ^{-1} and the invertibility of the Jacobian matrix of ρ with respect to δ. In the 3rd lemma, we derive the Lipschitz constant of the contraction and we prove that it is bounded away from 0 and 1. The 4th lemma ensures that for key moments and quantities the BLP approximations can be ignored uniformly asymptotically.

Lemma 2.4 *ρ^{-1} is $\mathcal{C}^{\infty}$*

Proof. We know that the demand function ρ is $\mathcal{C}^{\infty}$ and invertible on $\mathbb{R}^J$. Moreover,$\forall \delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta} \neq 0$. As a consequence, $\rho^{-1} : [0,1]^J \to \mathbb{R}^J$ the inverse demand function is also $\mathcal{C}^{\infty}$. □

Lemma 2.5 *For any $\delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta}$ is invertible.*

Proof. $\frac{\partial \rho}{\partial \delta}$ is a $J \times J$ matrix such that $\left(\frac{\partial \rho}{\partial \delta}\right)_{j,k}$ is:

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}} = \begin{cases} \int \mathcal{T}_{jt}(v)\left(1 - \mathcal{T}_{kt}(v)\right) f(v) dv \quad if \;\; j = k \\ -\int \mathcal{T}_{jt}(v) \mathcal{T}_{kt}(v) f(v) dv \quad if \quad j \neq k \end{cases}$$

with $\mathcal{T}_{jt}(v) \equiv \frac{exp\{\delta_{jt} + x'_{2jt} v\}}{1 + \sum_{j'=1}^{J} exp\{\delta_{j't} + x'_{2j't} v\}}$

One can easily check that $\frac{\partial \rho}{\partial \delta}$ is strictly diagonally dominant. Indeed for each row j:

$$\left| \frac{\partial \rho_j\left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}} \right| - \sum_{k \neq j} \left| \frac{\partial \rho_j\left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}} \right| = \int \mathcal{T}_{jt}(v) \underbrace{\left(1 - \sum_{k=1}^{J} \mathcal{T}_{kt}(v)\right)}_{>0} f(v) dv > 0$$

□

Lemma 2.6 (Contraction Mapping Lipschitz Constant)
Given parametric assumption $\mathcal{F}_0$, under assumptions B-E, assume that starting mean utility δ_0 is in $\mathcal{B}$ where $\mathcal{B}$ is compact, then without loss of generality there exists some $(\underline{a}, \bar{a}) \in \mathbb{R}^2$ with $\bar{a} > \underline{a}$ such that

for any $b \in \mathcal{B}$ *for any* $j = 1, \ldots, J$ $\underline{a} \leqslant b_j \leqslant \bar{a}$*, furthermore denote by* $\mathcal{X}$ *the compact support of* x_{2jt}*. Then on* $\mathcal{B}$ *the map* $C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0))$ *is a contraction with Lipschitz constant*

$$\epsilon = \max_{j=1,\ldots,J} \sup_{a\in\mathcal{B}, b\in[0;\bar{a}-\underline{a}]^J, x_2\in\mathcal{X}, \tilde{\lambda}\in\Lambda_0} 1 - \frac{\int \frac{\exp\{a_j+b_j+x'_{2j}v\}}{\left(1+\sum_k \exp\{a_k+b_k+x'_{2k}v\}\right)^2} f_0(v|\tilde{\lambda})dv}{\int \frac{\exp\{a_j+b_j+x'_{2j}v\}}{1+\sum_k \exp\{a_k+b_k+x'_{2k}v\}} f_0(v|\tilde{\lambda})dv}$$

which is in $(0;1)$

Proof. This proof is inspired by the proof of the Theorem in Appendix 1 of Berry et al. (1995). Let $C_j(\cdot) \equiv C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0))$, we first determine the partial derivative of $C_j(\cdot)$

$$\frac{\partial C_j(a)}{\partial a_j} = 1 - \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{\exp\{a_j + x'_{2kt}v\}(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}) - \exp\{2(a_j + x'_{2kt}v)\}}{(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$$
$$= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{\exp\{2(a_j + x'_{2jt}v)\}}{(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$$
$$\frac{\partial C_j(a)}{\partial a_{j'}} = \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{\exp\{a_j + x'_{2jt}v\}\exp\{a_{j'} + x'_{2j't}v\}}{(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$$

Note that for any $j = 1, \ldots, J$ all partial derivatives of $C_j(\cdot)$ are strictly positive and that the sum of its derivatives evaluated in a equals

$$\sum_{k=1}^J \frac{\partial C_j(a)}{\partial a_k} = \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{\exp\{a_j + x'_{2jt}v\} \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}}{(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$$
$$= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{\exp\{a_j + x'_{2jt}v\}(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\} - 1)}{(1 + \sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$$
$$= 1 - \frac{\int \frac{\exp\{a_j + x'_{2jt}v\}}{(1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}$$

For any $(a_1, a_2) \in \mathcal{B}^2$ let $\tilde{a} = (||a_1 - a_2||_\infty, \ldots, ||a_1 - a_2||_\infty) \in \mathbb{R}^J$ then

$$\begin{aligned} C_j(a_1) - C_j(a_2) &= C_j(a_2 + a_1 - a_2) - C_j(a_2) \leqslant C_j(a_2 + \tilde{a}) - C_j(a_2) \\ &\leqslant \int_{0^J}^{||a_1 - a_2||_\infty^J} \frac{\partial C_j(a_2 + b)}{\partial a} db \\ &\leqslant ||a_1 - a_2||_\infty \sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J} \sum_{k=1}^{J} \frac{\partial C_j(a+b)}{\partial a_k} \\ &\leqslant ||a_1 - a_2||_2 \max_{j=1,..J} \sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J, x_2 \in \mathcal{X}, \tilde{\lambda} \in \Lambda_0} \sum_{k=1}^{J} \frac{\partial C_j(a+b)}{\partial a_k} \\ &\equiv ||a_1 - a_2||_2 \, \epsilon \end{aligned}$$

where the 1st inequality holds because $C_j(\cdot)$ is increasing in all its inputs, the 2nd inequality holds by the fundamental theorem of calculus and by the total derivative formula, the 3rd and 4th inequalities hold by properties of norms.

We now prove that $\sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J, \tilde{\lambda} \in \Lambda_0} \sum_{k=1}^{J} \frac{\partial C_j(a+b)}{\partial a_k} \in (0; 1)$ which will imply that $\epsilon \in (0; 1)$. To do so we have to prove that $\sum_{k=1}^{J} \frac{\partial C_j(a, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k}$ is continuous in $(a, x_{2t}, \tilde{\lambda})$ and takes values in $(0; 1)$ almost surely, this way because $\mathcal{B}$, $\mathcal{X}$ and Λ_0 are compact by Weierstrass' Extreme Value Theorem the sum of partial derivatives will also take values in a compact which is inside $(0; 1)$, then the supremum will become a maximum which can be attained and which is inside $(0; 1)$. The sum of partial derivatives is almost surely in $(0; 1)$ because

$$
\begin{aligned}
&\int \frac{\exp\{a_j + x'_{2jt}v\}}{(1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv - \rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))\\
&= \int \frac{\exp\{a_j + x'_{2jt}v\}}{(1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv - \int \frac{\exp\{a_j + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}} f_0(v|\tilde{\lambda})dv\\
&= -\int \frac{\exp\{a_j + x'_{2jt}v\}\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}}{(1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv < 0\\
&\Rightarrow \frac{\int \frac{\exp\{a_j + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 1\\
&\Rightarrow \sum_{k=1}^J \frac{\partial C_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k} = 1 - \frac{\int \frac{\exp\{a_j + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} > 0\\
&\Rightarrow -\frac{\int \frac{\exp\{a_j + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 0\\
&\Rightarrow \sum_{k=1}^J \frac{\partial C_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k} = 1 - \frac{\int \frac{\exp\{a_j + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\}} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 1
\end{aligned}
$$

Continuity of the sum of the partial derivatives in (a, x_{2t}) is trivial, continuity in $\tilde{\lambda}$ also holds because $f_0(\cdot|\tilde{\lambda})$ must be continuously differentiable via Assumption D. $\forall e_1 > 0, \exists e_2 : \forall(\lambda_1, \lambda_2) : ||\lambda_1 - \lambda_2||_2 \leqslant e_2$ implies $|f_0(v|\lambda_1) - f_0(v|\lambda_2)| < e_1$ for all v which in turn implies

$$
\begin{aligned}
\forall x_2 \in \mathcal{X}, \forall a \in \mathcal{B} \quad &\left| \int \frac{\exp\{a_j + x'_{2j}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2k}v\}} (f_0(v|\lambda_1) - f_0(v|\lambda_2))dv \right|\\
&\leqslant \int \frac{\exp\{a_j + x'_{2j}v\}}{1+\sum_{k=1}^J \exp\{a_k + x'_{2k}v\}} |f_0(v|\lambda_1) - f_0(v|\lambda_2)|dv\\
&\leqslant e_1
\end{aligned}
$$

Thus both $\tilde{\lambda} \mapsto \rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ and $\tilde{\lambda} \mapsto \int \frac{\exp\{a_j + x'_{2jt}v}{(1+\sum_{k=1}^J \exp\{a_k + x'_{2kt}v\})^2} f_0(v|\tilde{\lambda})dv$ are continuous and so is their ratio. □

Lemma 2.7 (Uniform Convergence of Objective Function wrt BLP Approximations)

Given parametric assumption $\mathcal{F}_0$, under assumptions B-E and $\forall h$ which satisfies D

$$\underset{\tilde{\theta}\in\Theta_0}{sup}\ \sqrt{T}||\hat{g}_T(\tilde{\theta},h)-g_T(\tilde{\theta},h)||_2 \xrightarrow{\mathbb{P}} 0$$
$$\underset{\tilde{\theta}\in\Theta_0}{sup}\,||\hat{\Gamma}_T(\tilde{\theta},h)-\Gamma_T(\tilde{\theta},h)||_2 \xrightarrow{\mathbb{P}} 0$$
$$\underset{\tilde{\theta}\in\Theta_0}{sup}\,|\hat{\mathcal{Q}}_T(\tilde{\theta})-\mathcal{Q}(\tilde{\theta})| \xrightarrow{\mathbb{P}} 0$$

Proof. Parts of this proof are inspired from Freyberger (2015). We prove the 3 statements of the Lemma in order

1. Using the properties of the sup, the fact that $\forall(A,B)$ rv, $\forall e>0$, $\forall\alpha\in(0,1)$, $\mathbb{P}(A+B>e)\leqslant \mathbb{P}(A>\alpha e)+\mathbb{P}(B>(1-\alpha)e)$ and the previous decomposition of the difference between ξ and $\widehat{\xi}$ we can find an upper bound on the probability that that the difference between $\hat{g}_T(\cdot)$ and $g_T(\cdot)$ is above a deviation: For any $e_1>0$

$$\begin{aligned}\mathbb{P}(\underset{\tilde{\theta}}{sup}\ \sqrt{T}||\hat{g}_T(\theta,h)-g_T(\theta,h)||_2>e_1) &= \mathbb{P}(\underset{\tilde{\theta}}{sup}\ \sqrt{T}\frac{1}{T}||\sum_{j,t}(\widehat{\xi}_t(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})-\xi_t(f_0(\cdot|\tilde{\lambda}),\tilde{\beta}))h(z_{jt})||_2>e_1)\\ &\leqslant \mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-D_j(\rho,s_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3})\\ &+\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,\hat{s}_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3})\\ &+\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(D_j(\rho,\hat{s}_t,\tilde{\lambda})-D_j(\hat{\rho},\hat{s}_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3})\end{aligned}$$

   Then we can prove that each element of the upper bound converges to 0

   (a) By properties of contractions and using Lemma 2.6 we have

   $$|\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-D_j(\rho,s_t,\tilde{\lambda})|\leqslant\epsilon^H|\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-\delta_0|\leqslant\epsilon^H\kappa$$

   for some constant κ which exists due to the compactness of Λ_0, $\mathcal{X}$ and $\mathcal{B}$. Thus using the iid nature of the data **??**(i), the speed of the NFP algorithm Assumption E(iii), the triangle inequality, Markov inequality and Cauchy-Schwarz inequality the 1st element converges to 0

$$\begin{aligned}
&\mathbb{P}(\sup_{\tilde{\lambda}} \sqrt{T}||\frac{1}{T}\sum_{j,t}(\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0)) - D_j(\rho, s_t, \tilde{\lambda}))h(z_{jt})||_2 > \frac{e_1}{3}) \\
&\leqslant \mathbb{P}(\sqrt{T}\epsilon^H \kappa||\frac{1}{T}\sum_{j,t}h(z_{jt})||_2 > \frac{e_1}{3}) \leqslant \mathbb{P}(\sqrt{T}\epsilon^H \frac{1}{T}\sum_{j,t}||h(z_{jt})||_2 > \frac{e_1}{3}) \\
&\leqslant \frac{3\kappa}{e_1}\sqrt{T}\epsilon^H \sum_j \sqrt{\mathbb{E}(||h(z_{jt})||_2^2)} \underset{T\to\infty}{\to} 0
\end{aligned}$$

(b) Note that D_j is continuously differentiable in $s \in (0;1)$ so that it is uniformly continuous in s. Indeed C is $\mathcal{C}^\infty$ in s so that

$$\frac{\partial D(\rho, s_t, \tilde{\lambda})}{\partial s} = \prod_{h=1}^{H} \frac{\partial C(C^{(h-1)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})), s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial s}$$

Next because Λ_0 is compact it can be covered by some finite union of closed balls in $\mathbb{R}^{K_2}$, ie $\Lambda_0 \subset \cup_{c=1}^{N}\Lambda_{0,c}^N$ with $\forall c = 1,\ldots,N \;\; \Lambda_{0,c}^N = \{\tilde{\lambda} : ||\tilde{\lambda} - \lambda_c||_2 \leqslant r_N\}$, $\lambda_c \in \Lambda_0$ and $r_N \underset{N\to\infty}{\to} 0$. Consequently

$$\begin{aligned}
&\mathbb{P}(\sup_{\tilde{\lambda}} \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\lambda}))h_E(z_{jt})||_2 > \frac{e_1}{3}) \\
&\leqslant \mathbb{P}(\max_{c=1,\ldots,N} \sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \tilde{\theta}) - D_j(\rho, \hat{s}_t, \tilde{\theta}))h_E(z_{jt})||_2 > \frac{e_1}{3}) \\
&\leqslant \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\lambda})| \; ||h_E(z_{jt})||_2 > \frac{e_1}{3}) \\
&\leqslant \sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \lambda_c) - D_j(\rho, \hat{s}_t, \lambda_c))h_E(z_{jt})||_2 > \frac{e_1}{9}) \\
&\quad + \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, s_t, \lambda_c)| \; ||h_E(z_{jt})||_2 > \frac{e_1}{9}) \\
&\quad + \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, \hat{s}_t, \lambda_c) - D_j(\rho, \hat{s}_t, \tilde{\lambda})| \; ||h_E(z_{jt})||_2 > \frac{e_1}{9})
\end{aligned}$$

where the last inequality was obtained using the triangle inequality. Then by uniform continuity of D_j in s it follows that $\exists e_2 > 0$ such that $\forall c \; \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \lambda_c) - D_j(\rho, \hat{s}_t, \lambda_c))h_E(z_{jt})||_2 > \frac{e_1}{9}$ implies $\frac{1}{\sqrt{T}}||\sum_{j,t}(s_t - \hat{s_t})||_2 > e_2$ thence letting $\mathbb{P}^* = \mathbb{P}(\cdot|n_t, x_t, \xi_t)$

$$\mathbb{P}^*(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,s_t,\lambda_c)-D_j(\rho,\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})\leqslant\mathbb{P}^*(\frac{1}{\sqrt{T}}||\sum_{j,t}(s_t-\hat{s}_t)||_2>e_2)$$
$$\leqslant\frac{J\sum_t\mathbb{E}^*(||s_t-\hat{s}_t||_2)}{e_2\sqrt{T}}=\frac{J\sum_t\mathbb{E}^*\left(\sqrt{\sum_j(s_{jt}-\hat{s}_{jt})^2}\right)}{e_2\sqrt{T}}\leqslant\frac{J\sum_t\sqrt{\sum_j\mathbb{E}^*\left((s_{jt}-\hat{s}_{jt})^2\right)}}{e_2\sqrt{T}}$$
$$\leqslant\frac{J\sum_t\sqrt{\sum_j\mathbb{E}^*\left((\frac{1}{n_t}\sum_{i=1}^{n_t}y_{ijt}-\mathbb{E}^*(y_{ijt}))^2\right)}}{e_2\sqrt{T}}=\frac{J\sum_t\sqrt{\sum_j Var^*(\frac{1}{n_t}\sum_{i=1}^{n_t}y_{ijt})}}{e_2\sqrt{T}}$$
$$\leqslant\frac{J\sum_t\sqrt{\sum_j\frac{1}{n_t}Var^*(y_{ijt})}}{e_2\sqrt{T}}\leqslant\frac{J^{3/2}}{e_2}\frac{1}{\sqrt{T}}\sum_t\frac{1}{\sqrt{n_t}}$$

where Markov inequality, Jensen inequality, the fact that $y_{ijt}\in\{0;1\}$, that ε_{ijt} is iid extreme-value type 1 distributed across i, j and t, and the fact that n_t is iid and independent of all other variables have been used. Then taking the expectations and summing over N on both sides implies by Assumption E(i)

$$\sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,s_t,\lambda_c)-D_j(\rho,\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})\leqslant\frac{J^{3/2}N}{e_2}\sqrt{T}\mathbb{E}(n_t^{-1/2})\underset{T\to\infty}{\to}0$$

Next using continuity of D_j in $\tilde{\lambda}$ it must be that for any $e_1>0$ there exists some N such that $\forall\tilde{\lambda}\in\Lambda_{0,c}^N$ such that $||\tilde{\lambda}-\lambda_c||_2\leqslant r_N$ implies

$$\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\ ||h_E(z_{jt}||_2\leqslant\frac{e_1}{9}$$

because $r_N\underset{N\to\infty}{\to}0$. By definition of the supremum it also implies that

$$\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\ ||h_E(z_{jt}||_2\leqslant\frac{e_1}{9}$$

The contraposition is that

$$\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\ ||h_E(z_{jt}||_2>\frac{e_1}{9}$$

implies $\forall\tilde{\lambda}\in\Lambda_{0,c}^N\quad||\tilde{\lambda}-\lambda_c||_2>r_N$ which is impossible by definition of $\Lambda_{0,c}^N$. Consequently

$$\sum_{c=1}^{N}\mathbb{P}(\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\ ||h_E(z_{jt}||_2>\frac{e_1}{9})$$
$$\leqslant\sum_{c=1}^{N}\mathbb{P}(\cap_{\tilde{\lambda}\in\Lambda_{0,c}^N}||\tilde{\lambda}-\lambda_c||_2>r_N)=0$$

Similarly

$$\sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde\lambda\in\Lambda_{0,c}}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat s_t,\tilde\lambda)-D_j(\rho,\hat s_t,\lambda_c)|\,||h_E(z_{jt}||_2>\frac{e_1}{9})=0$$

(c) With the same arguments as in (b)

$$\begin{aligned}
&\mathbb{P}(\sup_{\tilde\lambda}\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat s_t,\tilde\lambda)-D_j(\hat\rho,\hat s_t,\tilde\lambda))h_E(z_{jt})||_2>\frac{e_1}{3})\\
&\leqslant\sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat s_t,\lambda_c)-D_j(\hat\rho,\hat s_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})\\
&\quad+\sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde\lambda\in\Lambda^N_{0,c}}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat s_t,\tilde\lambda)-D_j(\rho,\hat s_t,\lambda_c)|\,||h_E(z_{jt})||_2>\frac{e_1}{9})\\
&\quad+\sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde\lambda\in\Lambda^N_{0,c}}\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\hat\rho,\hat s_t,\lambda_c)-D_j(\hat\rho,\hat s_t,\tilde\lambda)|\,||h_E(z_{jt})||_2>\frac{e_1}{9})\\
&=\sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat s_t,\lambda_c)-D_j(\hat\rho,\hat s_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})
\end{aligned}$$

where $D_j(\rho,s_t,\lambda_c)=C^{(H)}(\delta_0,s_t,x_{2t},f_0(\cdot|\lambda_c))$. D_j is $\mathcal{C}^\infty$ in $\rho\in(0;1)$, moreover $\rho_j(\delta_t,x_{2t},f_0(\cdot|\tilde\lambda))$ and $\hat\rho_j(\delta_t,x_{2t},f_0(\cdot|\tilde\lambda))$ are continuously differentiable in Λ_0. Therefore there exists some $e_2>0$ such that

$$\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat s_t,\lambda_c)-D_j(\hat\rho,\hat s_t,\lambda_c)|\,||h_E(z_{jt})||_2>\frac{e_1}{9}$$

implies $\sup_{a\in\mathcal{B}}\frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a,x_{2t},f_0(\cdot|\lambda_c)-\hat\rho(a,x_{2t},f_0(\cdot|\lambda_c))||_2>e_2$, and as $\mathcal{B}$ is compact we can cover it by $\tilde N$ closed balls $\mathcal{B}_b^{\tilde N}=\{a\in\mathcal{B}:||a-a_b||\leqslant r_{\tilde N}\}$ with $a_b\in\mathcal{B}$ for any $b=1,\dots,\tilde N$ so that

$$\begin{aligned}
&\sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat s_t,\lambda_c)-D_j(\hat\rho,\hat s_t,\lambda_c)|\,||h_E(z_{jt})||_2>\frac{e_1}{9})\\
&\leqslant\sum_{c=1}^{N}\mathbb{P}(\sup_{a\in\mathcal{B}}\frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a,x_{2t},f_0(\cdot|\lambda_c)-\hat\rho(a,x_{2t},f_0(\cdot|\lambda_c))||_2>e_2)\\
&\leqslant\sum_{c,b}\mathbb{P}(\sup_{a\in\mathcal{B}_b^{\tilde N}}\frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a,x_{2t},f_0(\cdot|\lambda_c)-\hat\rho(a,x_{2t},f_0(\cdot|\lambda_c))||_2>e_2)\\
&=\sum_{c,b}\mathbb{P}(\frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a_b,x_{2t},f_0(\cdot|\lambda_c)-\hat\rho(a_b,x_{2t},f_0(\cdot|\lambda_c))||_2>e_2)
\end{aligned}$$

where the last equality was obtained reusing arguments from (b). As a consequence let $F_{jt}(v) = \frac{\exp\{a_{bj}+x'_{2jt}v\}}{1+\sum_k \exp\{a_{bk}+x'_{2kt}v\}}$ and $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot|x_t, \xi_t)$ then using Markov inequality and Cauchy-Schwarz inequality

$$\begin{aligned}
&\mathbb{P}^*(\frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \hat{\rho}(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda}))||_2 > e_2)\\
&\leqslant \frac{J\sum_t \mathbb{E}^*(||\hat{\rho}(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \rho(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda}))||_2)}{e_2\sqrt{T}}\\
&\leqslant \frac{J\sum_t \sqrt{\sum_j \mathbb{E}^*\left((\frac{1}{R}\sum_{r=1}^R F_{jt}(v_R) - \mathbb{E}^*(F_{jt}(v_R)))^2\right)}}{e_2\sqrt{T}} = \frac{J\sum_t \sqrt{\sum_j Var^*(\frac{1}{R}\sum_{r=1}^R F_{jt}(v_r))}}{e_2\sqrt{T}}\\
&\leqslant \frac{J^{3/2}}{e_2}\sqrt{\frac{T}{R}}
\end{aligned}$$

where the fact that v_r are iid draws from $f_0(\cdot|\tilde{\lambda})$ independent from all other variables has been used. It follows by taking the expectation and summing over N and $\tilde{N}$ that

$$\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, \hat{s}_t, \tilde{\lambda}) - D_j(\hat{\rho}, \hat{s}_t, \tilde{\lambda})|\ h_E(z_{jt})||_2 \underset{T\to\infty}{\to} 0$$

by Assumption E(i).

2. The 2nd statement is not formally proven as it largely builds on the proof of the 1st statement. To see why recall that

$$\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h) = \frac{1}{T}\sum_{jt} h(z_{jt})\frac{\partial}{\partial\theta}(\widehat{\xi}(\tilde{\theta}) - \xi_{jt}(\tilde{\theta}))'$$

More precisely let $e'_j = (0\ \dots\ 0\ \underbrace{1}_{\text{j-th coordinate}}\ 0\ \dots\ 0)$ then

$$\frac{\partial\xi_{jt}(\tilde{\theta})}{\partial\beta} = -x_{1jt}, \quad \frac{\partial}{\partial\lambda}\xi_{jt}(\tilde{\theta}) = -e'_j\left(\frac{\partial\rho(\delta_t(\tilde{\lambda}), x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial\delta}\right)^{-1}\int \frac{\exp\{\delta_{jt}(\tilde{\lambda}) + x'_{2jt}v\}}{1+\sum_{k=1}^J \exp\{\delta_{kt}(\tilde{\lambda}) + x'_{2kt}v\}}\frac{\partial}{\partial\lambda}f_0(v|\tilde{\lambda})dv$$

Thus the columns of the matrix $\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h)$ associated to the derivative in β are equal to 0. Furthermore using an uniform continuity argument $\left|\frac{\partial\hat{\xi}_{jt}(\tilde{\theta})}{\partial\lambda} - \frac{\partial\xi_{jt}(\tilde{\theta})}{\partial\lambda}\right| > e_1$ is implied by $||\widehat{\delta}_t(\tilde{\lambda}) - \delta_t(\tilde{\lambda})||_2 > e_2$ for some $e_2 > 0$. Using the compactness of Λ_0 and Assumption E it is straightforward that $\underset{\tilde{\lambda}}{sup}\ ||\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h)||_2 \overset{\mathbb{P}}{\to} 0$ for any h which satisfies the conditions in Assumption D.

3. The 3rd statement follows from the 1st. Indeed using Cauchy-Schwarz and properties of the supremum

$$\begin{aligned}
\sup_{\tilde{\theta}\in\Theta_0} |\hat{\mathcal{Q}}_T(\tilde{\theta}) - \mathcal{Q}_T(\tilde{\theta})| =& |(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E))'\hat{W}(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E)) \\
& - 2(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E))'\hat{W}g_T(\tilde{\theta},h_E))| \\
\leqslant & \sup_{\tilde{\theta}\in\Theta_0} ||(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E))||_2^2\bar{\mu}(\hat{W}) \\
& + 2\sup_{\tilde{\theta}\in\Theta_0} ||(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E))||_2 \sup_{\tilde{\theta}\in\Theta_0} ||g_T(\tilde{\theta},h_E))||_2\bar{\mu}(\hat{W})
\end{aligned}$$

where $\bar{\mu}(\cdot)$ maps a square matrix towards its maximum eigenvalue. By D(iv) and definition of the L_2 matrix norm, $\bar{\mu}(\hat{W}) \xrightarrow{\mathbb{P}} \bar{\mu}(W)$. Then we apply Jennrich's ULLN: the data is iid, Θ_0 is compact, and $g_T(\tilde{\theta},h_E) = \sum_j \xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})h_E(z_{jt})$ has an enveloppe with finite absolute 1st moment because $\xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta}) = \rho^{-1}(s_t,x_{2t},\tilde{\lambda}) - x'_{1jt}\tilde{\beta}$ and $\rho^{-1}(\cdot)$ has a maximum because it is continuous and its input are in a compact and because $\tilde{\beta}$ is in a compact and x_{1jt} has finite 4th moments, see Assumption B; Thus by the CMT $\sup_{\tilde{\theta}\in\Theta_0} ||g_T(\tilde{\theta},h_E))||_2 \xrightarrow{\mathbb{P}} \sup_{\tilde{\theta}\in\Theta_0} ||g(\tilde{\theta},h_E)||_2$; Finally using the 1st statement we have $||(\hat{g}_T(\tilde{\theta},h_E) - g_T(\tilde{\theta},h_E))||_2 \xrightarrow{\mathbb{P}} 0$ therefore by the CMT

$$\sup_{\tilde{\theta}\in\Theta_0} |\hat{\mathcal{Q}}_T(\tilde{\theta}) - \mathcal{Q}_T(\tilde{\theta})| \xrightarrow{\mathbb{P}} 0$$

□

Asymptotic Properties of the BLP estimator

Lemma 2.8 (Consistency of BLP Estimator)

Given parametric assumption $\mathcal{F}_0$ and under assumptions B-E,

$$\hat{\theta} \xrightarrow{\mathbb{P}} \theta_0$$

.

Proof. We prove consistency using arguments for the consistency of M-estimators. For any $e_1 > 0$ such that $|\hat{\theta} - \theta_0| > e_1$ then by Assumption D(iii) there exists some $e_2 > 0$ such that $\mathcal{Q}(\hat{\theta}) - \mathcal{Q}(\theta_0) > e_2$ as

θ_0 is the unique minimizer of the objective. Thence for any $e_1 > 0$, $\exists e_2 > 0$ such that

$$\begin{aligned}
\mathbb{P}(|\hat{\theta} - \theta_0| > e_1) &\leqslant \mathbb{P}(\mathcal{Q}(\hat{\theta}) - \mathcal{Q}(\theta_0) > e_2) \\
&= \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) + \mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) + \hat{\mathcal{Q}}_T(\hat{\theta}) - \hat{\mathcal{Q}}_T(\theta_0) > e_2) \\
&\leqslant \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) + \mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) > e_2) \\
&\leqslant \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) > (1-\alpha)e_2) + \mathbb{P}(\mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) > \alpha e_2)
\end{aligned}$$

where $\alpha \in (0;1)$, the 2nd inequality comes from the fact that $\hat{\mathcal{Q}}_T(\hat{\theta}) - \hat{\mathcal{Q}}_T(\theta_0)$ is almost surely negative by definition of $\hat{\theta}$, and the 3rd inequality is obtained by utilizing properties of indicator functions. Then by a direct implication of Lemma 2.7 the right-hand-side converges to 0.

□

Lemma 2.9 (Asymptotic Normality of BLP Estimator)
Given parametric assumption $\mathcal{F}_0$, under assumptions B-E and under $H_0 : f \in \mathcal{F}_0$

$$\sqrt{T}(\hat{\theta} - \theta_0) = \left(\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E)\right)^{-1} \sqrt{T}\Gamma'(\theta_0, h_E) W g_T(\theta_0, h_E) + o_P$$

Furthermore under $H_0; f \in \mathcal{F}_0$

$$\begin{aligned}
\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} &\mathcal{N}(0, (\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1} \Gamma'(\theta_0, h_E) W \Omega(\mathcal{F}_0, h_E) W \Gamma(\theta_0, h_E) \\
&(\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1})
\end{aligned}$$

Proof. We prove asymptotic normality using arguments from M-estimators asymptotics. From Taylor's theorem there exists some $\tilde{\theta}$ such that $||\tilde{\theta} - \theta_0||_2 \leqslant ||\hat{\theta} - \theta_0||_2$ and

$$\begin{aligned}
&\hat{g}_T(\hat{\theta}, h_E) = \hat{g}_T(\theta_0, h_E) + \hat{\Gamma}_T(\tilde{\theta}, h_E)(\hat{\theta} - \theta_0) \\
&\Rightarrow \sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\hat{\theta}, h_E) = \sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\theta_0, h_E) + \hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{\Gamma}_T(\tilde{\theta}, h_E)\sqrt{T}(\hat{\theta} - \theta_0) = 0 \\
&\Leftrightarrow \sqrt{T}(\hat{\theta} - \theta_0) = -\left(\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{\Gamma}_T(\tilde{\theta}, h_E)\right)^{-1} \sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\theta_0, h_E)
\end{aligned}$$

where the 1st implication is due to the FOC Assumption D(v). Then, we apply the CMT to $(A, B) \mapsto (A'BA)^{-1}A'B$ which is a continuous mapping if A and B are full rank so that when taking $A =$

$\hat{\Gamma}_T(\hat{\theta}, h_E)$ and $B = \hat{W}$ we obtain:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\left(\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E)\right)^{-1} \sqrt{T}\Gamma'(\theta_0, h_E) W g_T(\theta_0, h_E) + o_P$$

To prove that $plim\ \hat{\Gamma}_T(\hat{\theta}, h_E) = plim\ \hat{\Gamma}_T(\tilde{\theta}, h_E) = \Gamma(\theta_0, h_E)$ we make the following decomposition

$$\hat{\Gamma}_T(\hat{\theta}, h_E) - \Gamma(\theta_0, h_E) = \hat{\Gamma}_T(\hat{\theta}, h_E) - \Gamma_T(\hat{\theta}, h_E) + \Gamma_T(\hat{\theta}, h_E) - \Gamma(\hat{\theta}, h_E) + \Gamma(\hat{\theta}, h_E) - \Gamma(\theta_0, h_E)$$

where the 1st difference is o_P by Lemma 2.7, the 3rd difference is o_P by the CMT and the consistency of $\hat{\theta}$, see Lemma 2.8, and the 2nd difference is o_P by Jennrich's ULLN. The ULLN can be applied if and only if $\sum_j h_E(z_{jt}) \frac{\partial \xi_{jt}(\theta)}{\partial \theta}$ has an enveloppe with finite 1st absolute moments: $\xi_{jt}(\theta) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda)) - x'_{1jt}\beta$ and $\frac{\partial \xi_{jt}(\theta)}{\partial \beta} = x_{1jt}$ with x_{1jt} has finite moments of order 4 by Assumption B(iv), whereas $\frac{\partial \xi_{jt}(\theta)}{\partial \lambda} = \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial \lambda}$ and ρ^{-1} is C^∞ with arguments (s_t, x_{2t}, λ) which take values in a compact thus $\frac{\partial \rho^{-1}}{\partial \lambda}$ has bounds.

Thence $plim\ \hat{\Gamma}_T(\hat{\theta}, h_E) = plim\ \hat{\Gamma}_T(\tilde{\theta}, h_E) = \Gamma(\theta_0, h_E)$ which is full rank by Assumption D(ii), $plim\ \hat{W} = W$ which is full rank by Assumption D(iv), and by Lemma 2.7 $plim\ \sqrt{T}(\hat{g}_T(\theta_0, h_E) - g_T(\theta_0, h_E)) = 0$ so we can apply the aforementioned CMT and by the CLT which can be applied because $g(\theta_0, h_E) = 0$ under the null

$$\begin{aligned}
\sqrt{T}(\hat{\theta} - \theta_0) &= -\left(\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E)\right)^{-1} \sqrt{T}\Gamma'(\theta_0, h_E) W g_T(\theta_0, h_E) + o_P \\
&\xrightarrow{d} \mathcal{N}(0, (\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1} \Gamma'(\theta_0, h_E) W \Omega(\mathcal{F}_0, h_E) W \Gamma(\theta_0, h_E) \\
&\quad (\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1})
\end{aligned}$$

□

Asymptotic distribution of the test statistic

Proof of Theorem 5.1

Proof. This proof leans heavily on the proof of Lemma 2.9. By Taylor's theorem there exists $\tilde{\theta}$ such that

$$||\tilde{\theta}-\theta_0||_2 \leqslant ||\hat{\theta}-\theta_0||_2$$

$$\begin{aligned}
\sqrt{T}\hat{g}_T(\hat{\theta},h_D) &= \sqrt{T}\hat{g}_T(\theta_0,h_D)+\hat{\Gamma}_T(\tilde{\theta},h_D)\sqrt{T}(\hat{\theta}-\theta_0)\\
&= (I_{|h_D|_0} \quad -\Gamma(\theta_0,h_D)(\Gamma'(\theta_0,h_D)W\Gamma(\theta_0,h_D))^{-1}\Gamma'(\theta_0,h_D)W)\sqrt{T}\begin{pmatrix} g_T(\theta_0,h_D)\\ g_T(\theta_0,h_E)\end{pmatrix}+o_P\\
&\equiv (I_{|h_D|_0} \quad G)\sqrt{T}\begin{pmatrix} g_T(\theta_0,h_D)\\ g_T(\theta_0,h_E)\end{pmatrix}+o_P
\end{aligned}$$

The second equality is obtained by relying on the proof of Lemma 2.9 to express $\sqrt{T}(\hat{\theta}-\theta_0)$ as a function of moments, by relying on Lemma 2.7 so that $plim\ \sqrt{T}\hat{g}_T(\theta_0,h_D) = plim\ \sqrt{T}g_T(\theta_0,h_D)$ and $plim\ \hat{\Gamma}_T(\tilde{\theta},h_D) = plim\ \Gamma_T(\theta_0,h_D)$, and by using the CMT.

- Under $H_0 : f \in \mathcal{F}_0$ then $\mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(\theta_0)\right] = 0$ by LIE. So using the CLT and Slutsky's Lemma we obtain

$$\sqrt{T}\hat{g}_T(\hat{\theta},h_D) \xrightarrow{d} Z \sim \mathcal{N}(0,\Omega_0)$$

where

$$\Omega_0 = \begin{pmatrix} I_{|h_D|_0} & G\end{pmatrix}\begin{pmatrix}\Omega(\mathcal{F}_0,h_D) & \Omega(\mathcal{F}_0,h_D,h_E)\\ \Omega(\mathcal{F}_0,h_D,h_E)' & \Omega(\mathcal{F}_0,h_E)\end{pmatrix}\begin{pmatrix} I_{|h_D|_0}\\ G'\end{pmatrix}$$

with

$$\begin{aligned}
\Omega(\mathcal{F}_0,h_D) &= \mathbb{E}\Bigg[\Bigg(\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_D(z_{jt})\Bigg)\Bigg(\sum_j h_D(z_{jt})\xi_{jt}(f_0(.|\lambda_0),\beta_0)\Bigg)'\Bigg]\\
\Omega(\mathcal{F}_0,h_D,h_E) &= \mathbb{E}\Bigg[\Bigg(\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_D(z_{jt})\Bigg)\Bigg(\sum_j h_E(z_{jt})\xi_{jt}(f_0(.|\lambda_0),\beta_0)\Bigg)'\Bigg]\\
G &= -\Gamma(\theta_0,h_D)\left[\Gamma(\theta_0,h_E)'W\Gamma(\theta_0,h_E)\right]^{-1}\Gamma(\theta_0,h_E)'W
\end{aligned}$$

Thence by the continuous mapping theorem:

$$S(h_D,\mathcal{F}_0,\hat{\theta}) = \hat{g}_T(\hat{\theta},h_D)'\hat{\Sigma}\hat{g}_T(\hat{\theta},h_D) \xrightarrow{d} Z'\Sigma Z$$

□

- Under $H_a' : \mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right] \neq 0$, we have by Lemma 2.7, by consistency of $\hat{\theta} \xrightarrow{\mathbb{P}} \theta_0$ and the CMT:

$$\hat{g}_T(\hat{\theta}, h_D) = g_T(\theta_0, h_D) + o_P$$

Thus by Assumption D(iv) and the CMT

$$\frac{S(h_D, \mathcal{F}_0, \hat{\theta})}{T} \xrightarrow{\mathbb{P}} \underbrace{\mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right]' \Sigma \mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right]}_{\kappa(h_D,\mathcal{F}_0,\theta_0)}$$

Under H_a', $\kappa(h_D, \mathcal{F}_0, \theta_0)$ is strictly positive because Σ is positive definite. Thence,

$$\begin{aligned} \forall q \in \mathbb{R} \ \underset{T\to\infty}{lim} \ \mathbb{P}(S(h_D, \mathcal{F}_0, \hat{\theta}) > q) &= \underset{T\to\infty}{lim} \ \mathbb{P}\left(\frac{S(h_D, \mathcal{F}_0, \hat{\theta}) - q}{T} > 0\right) \\ &= \mathbb{P}(\kappa(h_D, \mathcal{F}_0, \theta_0) > 0) \\ &= 1 \end{aligned}$$

where the 2nd equality holds because convergence in probability implies convergence in distribution.

□

Application of Theorem 5.1 to the 2 polar cases

1. **Sargan-Hansen J test**

If $h_D = h_E$, with W and Σ are set to be equal to the GMM 2-step optimal weighting matrix

$$\Sigma = W = \mathbb{E}\left[\left(\sum_j \xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)h_E(z_{jt})\right)\left(\sum_j \xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)h_E(z_{jt})\right)'\right]^{-1} = \Omega(\mathcal{F}_0, h_E)^{-1}$$

Then under H_0:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_E|_0 - |\theta|_0}$$

Proof. By applying theorem 5.1, we have:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z'\Sigma Z \ \text{ with } \ Z \sim \mathcal{N}(0, \Omega_0)$$

If $h_D = h_E$ and $W = \Omega(\mathcal{F}_0, h_E)^{-1}$ then Ω_0 simplifies to

$$\Omega_0 = \Omega(\mathcal{F}_0, h_E) - \Gamma(\theta_0, h_E) \left[\Gamma(\theta_0, h_E)'\Omega(\mathcal{F}_0, h_E)^{-1}\Gamma(\theta_0, h_E)\right]^{-1} \Gamma(\theta_0, h_E)'$$
$$= \Omega(\mathcal{F}_0, h_E)^{1/2} M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)} \Omega(\mathcal{F}_0, h_E)^{1/2}$$

with $M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)} \equiv I_{|h_E|_0} - P_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)}$ is the orthogonal projection on the space orthogonal to $\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)$. Let $\tilde{Z} \sim \mathcal{N}(0, I_{|h_E|_0})$, we have by definition:

$$Z = \Omega(\mathcal{F}_0, h_E)^{1/2} M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)} \tilde{Z} \implies \Sigma^{1/2} Z = M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)} \tilde{Z}$$
$$\implies Z'\Sigma Z = \tilde{Z}' M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)} \tilde{Z}$$

Second line comes from symmetry and idempotence of $M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)}$. Orthogonal projections have eigenvalues equal to either 0 or 1 with the number of eigenvalues equal to one corresponding to the rank of the space it projects into, which in our case is $|h_E| - |\theta|_0$. If we denote by V the matrix of eigenvectors of $M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)}$ then note that $V'\tilde{Z} \sim \mathcal{N}(0, I_{|h_E|_0})$ so that

$$Z'\Sigma Z = \sum_{k=1}^{|h_E|_0 - |\theta|_0} (V'\tilde{Z})_k^2 \sim \chi^2_{|h_E|_0 - |\theta|_0}$$

□

2. **Non-redundant h_D and h_E**

If Ω_0 is full rank and if the econometrician sets $\Sigma = \Omega_0^{-1}$, then our test statistic has the following asymptotic distribution under H_0:

$$S(h_T, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_D|_0}$$

One sufficient condition for Ω_0 being full rank is $(\xi_{jt}(f(\cdot|\lambda_0), \beta_0))_{j=1}^{J}$ is independent across j and $(h_E(z_{jt}), h_D(z_{jt}))$ not being perfectly colinear.

Proof. The asymptotic result is direct; $(\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0))_{j=1}^{J}$ being independent across j and

$(h_E(z_{jt}), h_D(z_{jt}))$ not being perfectly colinear implies that

$$\Omega(\mathcal{F}_0, h_E, h_D) = \sum_j \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)^2 h_E(z_{jt}) h_D(z_{jt})'\right]$$

$$\Rightarrow \Omega_0 = \sum_j (I_{|h_D|_0} \quad G) Var\left(\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)\begin{pmatrix} h_D(z_{jt}) \\ h_E(z_{jt}) \end{pmatrix}\right)\begin{pmatrix} I_{|h_D|_0} \\ G' \end{pmatrix}$$

Thus Ω_0 is positive definite because it is the sum of positive definite matrices. □

1.B.5 Properties of the MPI in the composite specification test: $f \in \mathcal{F}_0$

Proposition 2.10 (Consistency of the test for the composite test with the MPI) *Under Assumption A and Assumptions B-E*

$$H_a : f \notin \mathcal{F}_0 \implies \forall q \in \mathbb{R}^+, \; \mathbb{P}(S(h_D^*, \mathcal{F}_0, \hat{\theta}) > q) \to 1.$$

Proof of Proposition 2.10.

From corollary 2.1. Under Assumption A,

$$\begin{aligned} H_a : f \notin \mathcal{F}_0 \implies & \quad \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)|z_{jt}] \neq 0 \;\; a.s \\ \implies & \quad \mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2 > 0 \; a.s \\ \implies & \quad \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2\right] > 0 \\ \implies & \quad \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]|z_{jt}]\right] > 0 \\ \implies & \quad \mathbb{E}\left[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]\right] > 0 \\ \implies & \forall \alpha \neq 0 \;\; H_a' : \; \mathbb{E}\left[\xi_{jt}(f_0(.|\lambda_0), \beta_0) \underbrace{\alpha \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]}_{h_D^*(z_{jt})}\right] \neq 0. \end{aligned}$$

From theorem 5.1, under Assumptions B-E,

$$H_a' : \mathbb{E}\left[\xi_{jt}(f_0(.|\lambda_0), \beta_0) h_D^*(z_{jt})\right] \neq 0 \implies \forall q \in \mathbb{R}^+, \; \mathbb{P}(S(h_D^*, \mathcal{F}_0, \hat{\theta}) > q) \to 1.$$

□

1.C Additional results and comments

1.C.1 Literature on the identification of the distribution of RC

In this section, we briefly summarize some recent findings on the identification of random coefficients in multinomial choice models. In their seminal paper, Berry and Haile (2014) shows the identification of the demand functions ρ in a framework that encompasses the BLP model but their result does not entail identification of the random coefficients' distribution per se. To achieve their identification result, they require a completeness condition on the instruments as well as additional conditions (eg: connected substitutes) to ensure invertibility of the demand functions. They also need to impose that at least one of the product characteristic has a coefficient that is not random and that is equal to 1. Notice that in BLP model, the structure implied by the logit shock guarantees invertibility of the demand functions.

Fox et al. (2012) provides conditions under which the distribution of random coefficients is identified in a mixed logit model with micro-level data and no endogeneity. Their identification result requires continuous characteristics in x_{2t} and rules out interaction terms (eg polynomial terms of x_{2jt}). Moreover, their result is restricted to distributions of random coefficients with a compact support - excluding for instance a normally distributed random coefficient.

Fox and Gandhi (2011) investigates the identification of the joint distribution of random coefficients v_i and idiosyncratic shocks ε_{ijt} in aggregate demand models without endogeneity. They also consider a setting where endogeneity is introduced in a very restrictive way. They show identification under a special regressor assumption and finite support of the unobserved heterogeneity. The special regressor assumption assumes that a variable in x_{1t} has full support and has an associated coefficient that is either 1 or -1. This special regressor assumption is very common in the literature on the identification of random coefficients (see Ichimura and Thompson (1998), Berry and Haile (2009), Matzkin (2007) and Lewbel (2000)). Their framework does not nest the standard BLP model as ϵ_{ijt} and v_i are both assumed to have a finite support but it is more general in other dimensions. They do not exploit the logit distributional assumption on ε_{ijt}, they do not impose independence between v_i and ε_{ijt}, their identification argument can be extended to the case where multiple goods are purchased.

In a setting much closer to ours, Dunker et al. (2022) studies the identification of the distribution of random coefficients in endogenous aggregate demand models which includes the BLP model as a special case (in particular, no parametric assumption is made on the idiosyncratic shock ε_{ijt}). They make a clever use of the Radon transform to identify f. The price they have to incur for flexibility is that they need to make stringent assumptions on the product characteristics: variables in x_t are required to be continuous and to satisfy a joint full support assumption. The idea is to exploit the variation in the covariates in order to trace out the distribution of rc f. Unfortunately, these requirements are rarely met in real data sets.

In contrast to the rest of the literature, Wang (2022) adopts all the parametric assumptions assumed in the standard BLP model and looks for the set of minimal assumptions under which the distribution of random coefficients is identified. This approach allows him to obtain sufficient conditions which are much less stringent than the rest of the literature (no special regressor assumption, no full support assumption, no continuity assumption). To be more specific, he shows that if the demand functions are identified on an open set of $\mathbb{R}^J$[39], then the distribution of random coefficients is identified. His proof astutely exploits the real analytic property of the demand functions[40].

1.C.2 Feasible MPI: conditional expectation

In this subsection, we briefly motivate our approach of approximating the conditional expectation by first projecting the endogenous variables on a relevant subset of the exogenous variables. The problem we encounter can be summarized as follows. We want to compute $\mathbb{E}[g(x_{1t}, x_{2t})|z_{jt}]$, where g is highly non-linear, x_{1t} are endogenous variables and x_{2t} are exogenous variables. Moreover, z_{jt} has a large dimension (in the BLP model, its order of magnitude the number of products$\times$ number of characteristics). Our approach consists in first projecting the endogenous variables x_{1t} on a relevant subset of z_t, before plugging them into $g(x_{1t}, x_{2t})$. The traditional approach consists in using a non-parametric estimator

[39] which can be achieved using theorem 1 in Berry and Haile (2014)

[40] In particular, the real analytic property yields that the local identification of ρ on $\mathcal{D} \subset \mathbb{R}^J$ implies identification of ρ on $\mathbb{R}^J$. From global identification of ρ, he is then able to show that the random coefficients' distribution is identified under a simple rank condition on x_{2t}

of $\mathbb{E}[g(x_{1t}, x_{2t})|z_{jt}]$. However, given the dimension of z_{jt}, this approach is likely to yield poor results in practice because of the huge curse of dimensionality. In contrast, we know that some endogenous variables in x_{1t} only depend on a subset of z_{jt}, which we denote $\tilde{z}_{1jt}$, then we can use this to our advantage to construct a more precise estimator of $\mathbb{E}[[g(x_{1t}, x_{2t})|z_{jt}]$. First, $\hat{x}_{1t} = \mathbb{E}[x_{1t}|\tilde{z}_{jt}]$ will be much more accurately estimated as we only condition on $\hat{z}_{1jt}$ (for instance, the price usually depends on its own cost shifters and its own product characteristics, while the dependence with respect to characteristics of other goods is much weaker and can be ignored empirically). For exposition, we further assume that $dim(x_{1t}) = 1$. We take a second order Taylor expansion of $\mathbb{E}[g(x_{1t}, x_{2t})|z_{jt}]$ around $\hat{x}_{1t}$.

$$
\begin{aligned}
\mathbb{E}[g(x_{1t}, x_{2t})|z_{jt}] &= \mathbb{E}[g(\hat{x}_{1t}, x_{2t})|z_{jt}] + \mathbb{E}\left[\frac{\partial g(\hat{x}_{1t}, x_{2t})}{\partial x_1}(x_{1t} - \hat{x}_{1t})|z_{jt}\right] + \mathbb{E}\left[\frac{\partial^2 g(\tilde{x}_{1t}, x_{2t})}{\partial x_1^2}(x_{1t} - \hat{x}_{1t})^2|z_{jt}\right] \\
&= g(\hat{x}_{1t}, x_{2t}) + \frac{\partial g(\hat{x}_{1t}, x_{2t})}{\partial x_1}\underbrace{\mathbb{E}[(x_{1t} - \hat{x}_{1t})|z_{jt}]}_{=0} + \mathbb{E}\left[\frac{\partial^2 g(\tilde{x}_{1t}, x_{2t})}{\partial x_1^2}(x_{1t} - \hat{x}_{1t})^2|z_{jt}\right]
\end{aligned}
$$

with $\tilde{x}_{1t} \in [x_{1t}; \hat{x}_{1t}]$. Our approach yields an estimator that converges faster to $g(\hat{x}_{1t}, x_{2t})$, which is a first order approximation of $\mathbb{E}[g(x_{1t}, x_{2t})|z_{jt}]$.

1.C.3 Choice of the large-T asymptotics

In this paper, we study the asymptotics of our test when the number of markets T grows to infinity. We could also study the asymptotic properties of the BLP estimator and of the test when J grows to infinity and T stays fixed. We do not pursue this route for several reasons. First, from an economic point of view, a market with a number of products that grows to infinity is hardly conceivable in industries with imperfect competition and barriers to entry. Second, from a theoretical point of view there is a tension between the identification of demand which require all market shares to be strictly positive, see Berry and Haile (2014), and the large market asymptotics which require all market shares to tend to 0 as J grows to infinity, see Berry et al. (2004). At the same time it is well established that a many (weak) instruments problem can easily occur in a BLP model with a fixed number of markets and many products especially when using the traditional BLP instruments, see Armstrong (2016).

Consequently only markets with perfect competition and a careful choice of instruments could somehow fit the assumptions necessary for the BLP model to yield consistent estimators and valid tests with large J. Yet in the majority of empirical IO papers the markets have imperfect competition, sometimes oligopolies, and use the traditional BLP instruments. Thus we establish our theory with a large number of independent markets, which is a natural setting for empirical IO papers and which is not plagued with the aforementioned theoretical problems.

1.C.4 Construction of the interval instruments in practice

We now provide more details on how to construct the interval instruments in practice. The procedure to construct the interval instruments is as follows:

1. Given $(\mathcal{F}_0, \hat{W}, h_E)$, the researcher derives the BLP estimator $\hat{\theta}$
2. Then the researcher chooses L points $(v_l)_{l=1}^{L} \in \mathbb{R}^L$ in the presumed support of $f_0(\cdot|\hat{\lambda})$.
3. Finally, the researcher can construct a set of L interval instruments based on the approximations of the MPI that we develop in sections 1.4.2 and 1.4.1.
   - Global approximation: $\{\pi_{j,l}(z_{jt})\}_{l=1,\ldots,L}$ interval instruments, which are such that:

$$\mathbb{E}\left[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}\right] \approx \log\left(\sum_{l=1}^{L} \omega_l \ \pi_{j,l}(z_{jt})\right) \text{ with } \pi_{j,l}(z_{jt}) = \frac{\frac{\exp\{x'_{2jt}v_l\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{kt}+x'_{2kt}v_l\right\}}}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{jt}+x'_{2kt}v\right\}} f_0(v)dv}$$

     with $\hat{\delta}^0_t$ the linear projection of δ^0_t on z_{jt} (or a carefully chosen subset of z_{jt}).
   - Local approximation: $\{\bar{\pi}_{j,l}(z_{jt})\}_{l=1,\ldots,L}$ interval instruments such that

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \sum_{l=1}^{L} \bar{\omega}_l \ \bar{\pi}_{j,l}(z_{jt})$$

$$\text{with } \bar{\pi}_{j,l}(z_{jt}) = \left(\frac{\partial\rho(\hat{\delta}^0_t, x_{2t}, f_0)}{\partial\delta}\right)^{-1}\left[\frac{\exp\{\hat{\delta}^0_t + x_{2t}v_l\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{kt} + x'_{2kt}v_l\right\}} - \rho_j(\hat{\delta}^0_t, x_{2t}, f_0)\right]$$

     with $\hat{\delta}^0_t$ the linear projection of δ^0_t on z_{jt} (or a carefully chosen subset of z_{jt}).

Choice of the L points in the domain of f_a The researcher doesn't know a priori the support of the true density f_a. Thus, he/she must choose points in the domain of definition of f_a. If this choice coincides with points of the support where $|f_0(\cdot|\lambda_0) - f_a|$ is large, then this choice generates more informative instruments. In practice, one can take points in the high density regions of $f_0(\cdot|\lambda_0)$ (eg if $\mathcal{F}_0$ is the Gaussian family, then one can take points around the mean λ_0). The choice of of the number of instruments N obeys a usual bias variance tradeoff. On the one hand, a large L allows to better approximate the MPI and thus increases the detection ability of the instruments. On the other hand, it is well-known that a larger number of instruments can induce finite sample bias and can distort asymptotic distributions of estimators and tests such as the over-identification test.[41] Moreover, we observe in our simulations, that when one takes points in the support that are too close to each other, the implied instruments suffer from high levels of colinearity. For these reasons we advise not to use too few or too many interval instruments, in our simulations and application we use between 6 and 10 instruments (in every dimension). We leave a formal analysis of the optimal choice of L and of the general approximations properties of the interval instruments for future work.

1.C.5 Feasible MPIs for estimation

In the estimation framework, the researcher assumes that $f \in \mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ and wants to estimate the true parameter $\theta_0 = (\beta_0', \lambda_0')'$ under this parametric restriction. From the connection between the MPI and the local instruments that we present in Section 1.3.3, we infer that good testing instruments $h_E(z_{jt})$ ought to approximate the MPI devoted to test $H_0 : \theta = \theta_0$ against any local alternative. If we have an initial estimator of θ_0, we can directly use the interval instruments presented in Section 1.4 to approximate the MPI devoted to test $H_0 : \theta = \theta_0$. However, this approach requires to estimate $\hat{\theta}$ in a first step. Here, we present an alternative approach based on the global approximation of the MPI we derived in Section 1.4.2, which has the advantage of not requiring a first stage estimate of θ_0. it is straightforward to show that for any true parameter θ_0 and any alternative θ_a, we can rewrite the global approximation of the non-linear part of the MPI as follows:

[41] See Roodman (2009) for a review on the effect of many possibly weak moments on estimation and testing.

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, \theta_0, \theta_a)|z_{jt}] \approx \log\left(\sum_{l=1}^{L} \bar{\bar{\omega}}_l(\theta_0,\theta_a)\ \hat{\bar{\pi}}_{j,l}(z_{jt})\right) \text{ with } \hat{\bar{\pi}}_{j,l}(z_{jt}) = \frac{\exp\{x'_{2jt}v_l\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\bar{\delta}}^0_{jt} + x_{2jk}v_l\right\}}$$

$$\text{and } \bar{\bar{\omega}}_l(\theta_0,\theta_a) = \frac{\bar{\omega}_l(\theta_a)}{\int_{\mathbb{R}^{K_2}} \frac{\exp\{x'_{2jt}v\}}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2jk}v\right\}} f_0(\cdot|\lambda_0)(v)dv}$$

, with $\hat{\bar{\delta}}^0_{jt}$ projected first stage estimates of δ^0_{jt}, which can be obtained, for example, under the logit specification. $\hat{\bar{\pi}}_{j,l}(z_{jt})$ do not depend on f_0 and can be used for estimation.

1.C.6 Estimation procedure when the distribution of RC is a mixture

In this section, we present a procedure to estimate the BLP model when the distribution of RC is parametrized as a mixture. Namely, we perform the estimation under $H_0 : f \in \mathcal{F}_0$ with $\mathcal{F}_0$ the family of Gaussian mixtures with L components. The pdf of a Gaussian mixture writes as follows:

$$\forall x \in \mathbb{R}\ ,\ f_0(x|\lambda_0) = \sum_{l=1}^{L} p_{l0} f_l(x|\lambda_{l0}) \qquad \sum_{l=1}^{L} p_{l0} = 1 \qquad L \geqslant 1$$

where $f_{l0}(\cdot|\lambda_{l0})$ is the pdf of a $\mathcal{N}(\mu_{l0}, \sigma^2_{l0})$.

As long as the means are different ($\mu_{l0} \neq \mu_{l'0}\ \forall l \neq l'$), the gaussian mixture is uniquely characterized by the vector $\lambda_0 = (p_{10}, \ldots, p_{L0}, \mu_{10}, \ldots, \mu_{L0}, \sigma^2_{10}, \ldots, \sigma^2_{L0})$ up to permutations of indexes[42]. The objective of our procedure is to estimate the parameters of the model $\theta_0 = (\beta_0, \lambda_0)$ where λ_0 characterizes the mixture. In general, the problem of estimating a density by a mixture is solved through the use of the well-known Expectation-Maximization (EM) algorithm. In our case, the application of this algorithm is made difficult by two main obstacles. First, we do not observe directly the random coefficients. Second, we do not have individual choice data which would have enabled us to construct a likelihood as in Train (2008). As an alternative, we propose to adapt the BLP estimation procedure to estimate the parameters of a mixture of gaussians instead of the single normal distribution. The mixture affects the derivation of the market shares. The random coefficient v_i is now a gaussian mixture.

[42] If for some $l \neq l'$ we have $\mu_{l0} = \mu_{l'0}$ then the Gaussian mixture becomes observationally equivalent to an infinite number of other Gaussian mixtures

Hence, $v_i = \sum_{l=1}^{L} \mathbf{1}\{D_i = l\} v_{il}$ where $(v_{il})_{i=1}^{n}$ are iid and have density $f_{l0}(\cdot|\lambda_{l0})$ known up to λ_{l0} for $l = 1, \ldots, L$, and where $(D_i)_{i=1}^{n}$ are iid categorically distributed with pmf $\mathbb{P}(D_i = l) = p_{l0}$. For all market t and product j, the demand functions are as follows:

$$\begin{aligned}\rho_j(\delta_t, x_{2t}, f_0(.|\lambda_0)) &= \mathbb{P}(j \text{ chosen in market } t \text{ by } i | x_{1t}, x_{2t}, \xi_t) \\ &= \int_{\mathbb{R}} \frac{\exp\{x'_{1jt}\beta_0 + x'_{2jt}v + \xi_{jt}\}}{1 + \sum_{j'=1}^{J} \exp\{x'_{1j't}\beta_0 + x'_{2j't}v + \xi_{j't}\}} f_0(v|\lambda_0) dv \\ &= \sum_{l=1}^{L} p_{l0} \int_{\mathbb{R}} \frac{\exp\{\delta_{jt} + x'_{2jt}v\}}{1 + \sum_{j'=1}^{J} \exp\{\delta_{j't} + x'_{2j't}v\}} f_{l0}(v|\lambda_{l0}) dv\end{aligned}$$

Reparametrization. The parameter λ associated with the mixture consists of the means, the standard deviation and the probability of each component. As highlighted by Ketz (2019) in the simple Gaussian case, the way we parametrize the model can greatly affect the asymptotic properties of the estimator as well as the quality of the estimation. In particular, he shows that the standard deviations σ should be reparametrized in order to avoid boundaries issues when σ close to 0. We follow this parametrization and perform the minimization with respect to $\{(+/-)\sqrt{\sigma_l}\}_{l=1}^{L}$ instead and $(\sigma_l)_{l=1}^{L}$ directly. An additional difficulty in the case of mixtures concerns the estimation of the probabilities associated to each component. These probabilities must all be between 0 and 1 and their sum must be equal to 1. To smoothly integrate these constraints, we perform the optimization with respect to $\gamma = (\gamma_2, \ldots, \gamma_L)$ with $p = (p_1, p_2, \ldots, p_L) = (\frac{1}{1+\sum_{l=2}^{L}\exp\{\gamma_l\}}, \frac{\exp\{\gamma_2\}}{1+\sum_{l=2}^{L}\exp\{\gamma_l\}}, \ldots, \frac{\exp\{\gamma_L\}}{1+\sum_{l=2}^{L}\exp\{\gamma_l\}})$.

Estimation details. Apart from the modification in the computation of the market shares and the new parametrization of the model, the estimation procedure with a mixture follows closely the traditional one and the parameters of interest are estimated by minimizing a GMM criterion. Let $\mathcal{Q}(\theta)$ the GMM objective function:

$$\mathcal{Q}(\theta) = \hat{\xi}(\theta)' h_E(Z) W h_E(Z)' \hat{\xi}(\theta)$$

We now describe the derivation of the Gradient that we provide to the minimization program.

$$\frac{\partial \mathcal{Q}}{\partial \theta} = 2\left[\frac{\partial \hat{\xi}(\theta)}{\partial \theta}\right]' h_E(Z) W h_E(Z)' \hat{\xi}(\theta)$$

Where $\frac{\partial \hat{\xi}(\theta)}{\partial \beta} = -x_1$ and where by the implicit function theorem we have $\hat{\rho}_j(\delta_t, x_{2t}, \lambda) - s_{jt} = 0 \ \ \forall j, t$ which implies:

$$\frac{\partial \hat{\xi}(\theta)}{\partial \lambda} = \frac{\partial \hat{\delta}(\theta)}{\partial \lambda} = -\left[\frac{\partial \hat{\rho}(\delta, x_2, \lambda)}{\partial \delta}\right]^{-1} \frac{\partial \hat{\rho}(\delta, x_2, \lambda)}{\partial \lambda}$$

- $\frac{\partial \rho}{\partial \delta}$ is a $JT \times JT$ diagonal by block matrix such that:

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \delta_{kt}} = \begin{cases} \sum_l p_l \int \mathcal{T}_{jlt}(v)\left(1 - \mathcal{T}_{klt}(v)\right)\phi_l(v) dv & if \ \ j = k \\ -\sum_l p_l \int \mathcal{T}_{jlt}(v) \mathcal{T}_{klt}(v) \phi_l(v) dv & if \quad j \neq k \end{cases}$$

with $\mathcal{T}_{jlt}(v) \equiv \frac{exp\{\delta_{jt} + x'_{2jt} v_l\}}{1 + \sum_{j'=1}^{J} exp\{\delta_{j't} + x'_{2j't} v_l\}}$

- $\frac{\partial \rho}{\partial \lambda}$ is a $JT \times (3L - 1)$ matrix such that:

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \mu_l} = p_l \int \mathcal{T}_{jlt}\left(x_{2jt} - \sum_{j'} \mathcal{T}_{j'lt} x_{2j't}\right) \phi(v) dv$$

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \sigma_l} = p_l \int \mathcal{T}_{jlt}\left(x_{2jt} - \sum_{j'} \mathcal{T}_{j'lt} x_{2j't}\right) v \phi(v) dv$$

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \gamma_l} = \sum_{l'=1}^{L} \zeta(l, l') \int \mathcal{T}_{jlt}$$

With $\zeta(l, l') = \frac{-\exp\{\gamma_l\}}{1 + \sum_{k \neq 1} \exp\{\gamma_k\}} \times \frac{\exp\{\gamma_{l'}\}}{1 + \sum_{k \neq 1} \exp\{\gamma_k\}} + \mathbf{1}\{l = l'\} \frac{\exp\{\gamma_l\}}{1 + \sum_{k \neq 1} \exp\{\gamma_k\}} = -p_l \times p_{l'} + \mathbf{1}\{l = l'\} p_l$

1.C.7 Properties of the feasible approximations of the MPI

So far, we have studied the properties of the MPI, which is an ideal instrument that cannot be derived in practice. Nevertheless, in light of the previous results, the MPI provides a useful upper bound on the power that can be reached using our specification test. More precisely, the asymptotic slope reached by the MPI can be interpreted as a power envelope on our specification test. Ideally, we want our specification test, with the approximated MPIs as instruments, to achieve slopes close to the ones reached by the MPI. We now distinguish 2 situations.

First, we consider the case where the econometrician tests $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against the a known alternative $over\bar{H}_a : (f, \beta) = (f_a, \beta_a)$. This situation is not interesting in practice as the econometrician usually doesn't know the true alternative and doesn't want to specify an alternative. Nevertheless, it illustrates that in this specific case, we can (in theory) derive a consistent estimator of the MPI. Indeed, in this particular case, we can directly derive an analytical expression for the correction term $\Delta_{0,a}^{\xi_{jt}}$ either using its definition or the expression in 1.4.2. Next, we must to compute the conditional expectation of our the correction term with respect to z_{jt}. This step is quite challenging because the dimension of z_{jt} is large and because the correction term is heavily non-linear and non-separable with respect to the endogenous variables. In theory, a solution is to perform a Sieve non-parametric estimation of the conditional mean and under standard regularity conditions recover a consistent estimator of $\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]$. Unfortunately, the rate of converge will be extremely slow given the dimension of z_{jt} and we don't recommend to do this in practice. Instead, we suggest to use the global approximation and to project the endogenous variables on the space spanned by a relevant subset of z_{jt}. As we show in Appendix 1.C.2, this strategy yields an estimator which converges faster to a first order approximation of the MPI.

Second, we consider the more realistic situation where the econometrician tests $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against an unspecified alternative. In this case, we use the interval instruments that we developed in Section 1.4 as an approximation of the MPI. Due to the different layers of approximations which intervene in the construction of these instruments and the absence of knowledge of f_a, it is quite difficult to establish conditions under which these instruments can reach the optimal slope of the MPI. A thorough analysis of the properties of these instruments is beyond the scope of this paper and may constitute an interesting starting point for future research. In spite of the lack of theoretical analysis, our Monte Carlo exercises show that the interval instruments perform really well in finite sample.

1.D Monte Carlo experiments

1.D.1 Counterfactuals under an alternative distribution

For the simulation exercise presented in Section 1.6.2, we use the following expressions for own- and cross-price elasticities for product $j \in \{1, 2, ..., J\}$. For the sake of simplicity, we drop the market index, t, in the following expressions.

- Own-price elasticity:

$$\eta_j^j = \frac{p_j}{s_j}\frac{\partial s_j}{\partial p_j} = \frac{p_j}{s_j}\int -\alpha \left(1 - \frac{\exp\{\delta_j + x_{cj}v_i\}}{1+\sum_{j'=1}^{J}\exp\{\delta_{j'} + x_{cj'}v_i\}}\right)\underbrace{\frac{\exp\{\delta_j + x_{cj}v_i\}}{1+\sum_{j'=1}^{J}\exp\{\delta_{j'} + x_{cj'}v_i\}}}_{s_{ij}} f_\theta(v)dv$$

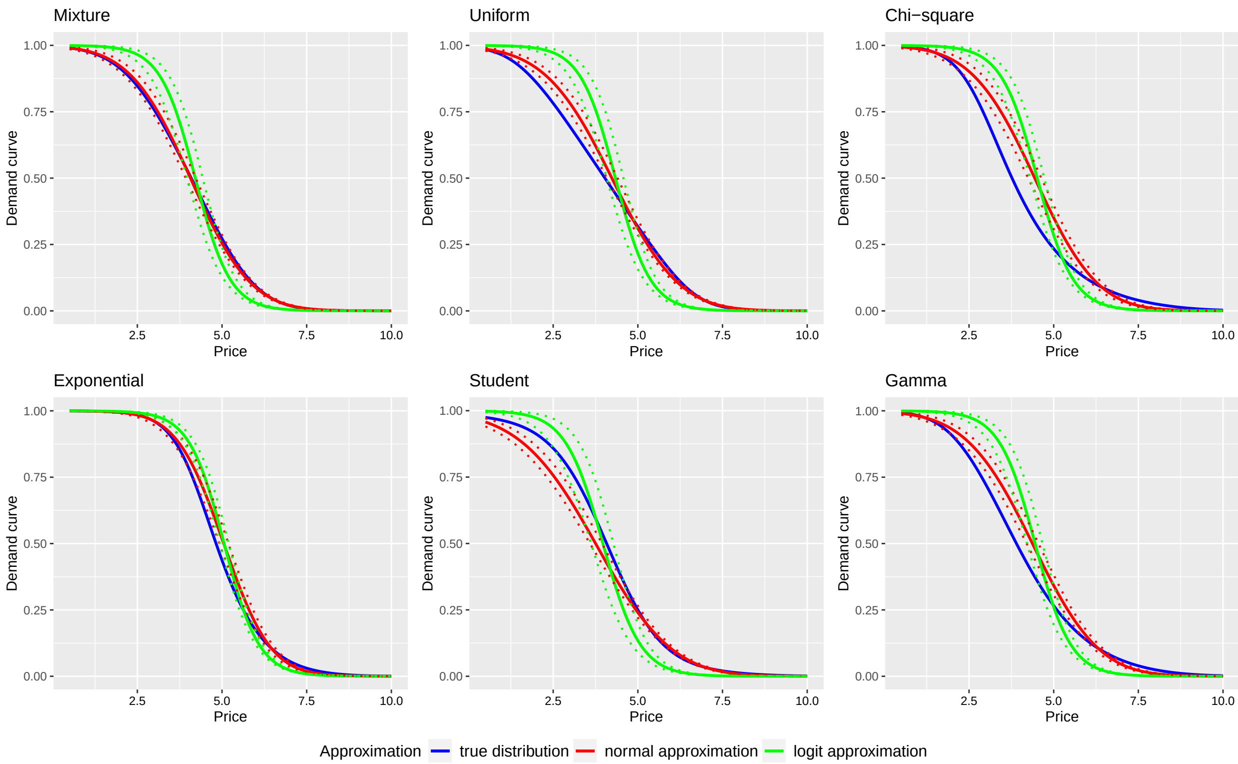
- Cross-price elasticity ($k \neq j$):

$$\eta_j^k = \frac{p_k}{s_j}\frac{\partial s_j}{\partial p_k} = \frac{p_k}{s_j}\int \alpha \left(\frac{\exp\{\delta_j + x_{cj}v_i\}}{1+\sum_{j'=1}^{J}\exp\{\delta_{j'} + x_{cj'}v_i\}}\right)\frac{\exp\{\delta_k + x_{ck}v_i\}}{1+\sum_{j'=1}^{J}\exp\{\delta_{j'} + x_{cj'}v_i\}} f_\theta(v)dv$$

where $\alpha = 2$ and $\delta_j = 2 + x_{aj} + 1.5x_{bj} - 2p_j + \xi_j$ in the DGP.

Demand functions. In Figure 1.9, we plot the demand functions generated under the different specifications (logit and gaussian) of the true densities.

Figure 1.9: Demand function



1.D.2 Finite sample performance of the test

Practical implementation of the test. For each setting, we estimate the model for 1000 replications. Minimization is performed with nloptr (algorithm: NLOPT-LD-LBFGS). We provide an analytical gradient. The Threshold for the outer loop is 1e-9 while the threshold for the inner loop is 1e-13. We use squarem and a C++ implementation for the computation of the market shares to speed up the contraction. We also parallelize the contraction over markets using 7 independent cores. Now we formally describe the instruments included in each test.

Power against local alternatives. We now assess the local power properties of our test by assuming that the random coefficient v_i is distributed according to a local alternative. Namely, we assume $v_i \sim$

$\left(1 - \frac{1}{\sqrt{T}}\right) \mathcal{N}(2, 1) + \frac{1}{\sqrt{T}} Y$ where Y is an alternative distribution including exponential, Chi-square, Student, Uniform. We ensure that Y has mean 2 and variance 1. The results are reported in 1.13. First, we can observe that except for the uniform local alternative, our test appears to have non-trivial power against all the other local alternatives. For the exponential and chi-square distributions, it is clear that our test with interval instruments outperforms the Sargan-J test with traditional instruments. For the student local alternative, the results seem quite unstable for small sample sizes but as T increases, interval instruments also seem to perform better. For the uniform alternative, it appears that we don't have power against this local alternative.

Table 1.13: Empirical power, local alternatives (1000 replications)

Number of markets	T=50				T=100				T=200			
Test type	J	I	J	I Local	J	I Local	J	I	J	I Local	J	I Local
Exponential	0.266	0.704	0.227	0.677	0.222	0.869	0.272	0.868	0.236	0.982	0.394	0.975
Chi-square	0.217	0.219	0.134	0.174	0.13	0.167	0.096	0.151	0.099	0.171	0.086	0.15
Student	0.212	0.139	0.33	0.436	0.115	0.115	0.127	0.093	0.082	0.13	0.134	0.312
Uniform	0.198	0.1	0.126	0.074	0.107	0.062	0.095	0.051	0.073	0.049	0.084	0.044

1.D.3 Finite sample performance of Interval instruments for estimation

Practical implementation of the estimation procedure. To assess the performance of our instruments in estimating the non-linear parameters with a flexible distribution of random coefficients, we simulate data with a distribution of random coefficients following a mixture of gaussians and we estimate the parameters of this mixture. For each setting, we estimate the model for 1050 replications. We select the replications with an objective function below a certain threshold (in order to avoid local minima). Minimization is performed with nloptr (algorithm: NLOPT-LD-LBFGS). We provide an analytical gradient, which we describe subsequently. The Threshold for the outer loop is 1e-9 while the threshold for the inner loop is 1e-13. We use squarem and a C++ implementation for the computation of the market shares to speed up the contraction. We also parallelize the contraction over markets using 7 independent core.

Before we formally define the different sets of instruments, let us present the estimation procedure when the distribution of random coefficients is assumed to be a mixture.

Instruments Now we formally describe the instruments present in each different sets used for estimation.

- Differentiation instruments: differentiation instruments + exogenous characteristics (polynomial terms) + cost shifters (20 instruments)

- Optimal instruments are computed in two stages. The first stage instruments consist of differentiation instruments and exogenous characteristics (polynomial terms). Second stage instruments consist of polynomial terms of exogenous characteristics and the approximation of optimal instruments proposed in Reynaert and Verboven (2014) (approximation of $\mathbb{E}\left[\frac{\partial\rho_j^{-1}(s_t,x_{2t},\lambda)}{\partial\lambda}\middle|z_t\right]$). The set called optimal instruments includes 15 instruments.

- Interval Instruments are computed in two stages. The first stage instruments consist of differentiation instruments and exogenous characteristics (polynomial terms). Second stage instruments are the interval instruments couples with some exogenous characteristics. A total of 23 instruments. The points in the support to compute the interval instruments are chose as follows: we take equally spaced points in the interval $\{\beta_{3L} - 0.5(\beta_{3H} - \beta_{3L}), \beta_{3H} + 0.5(\beta_{3H} - \beta_{3L})\}$.

Comparison of the performance between the different sets of instruments. We now report the mean biases and the empirical $\sqrt{MSE}$ of the estimates for each set of instruments and for different sample sizes. We also plot the distributions of estimates for the non-linear parameters for the different sets of instruments. First, we plot the distribution of estimates obtained when the set of differentiation instruments from Gandhi and Houde (2019) is used with a sample of $T = 200$ markets and $J = 12$ products. We observe that despite a relatively large sample, the differentiation instruments perform rather poorly in estimating the non-linear parameters associated with the mixture of Gaussians. In particular, the estimates of the standard deviation parameters associated to each component are very dispersed and a large portion of the estimates are bunched at zero. Second, we plot the distribution of non-linear estimates obtained with the optimal instruments from Reynaert and Verboven (2014). They tend to perform better than the differentiation instruments as we can see that the estimates are more concentrated around the true value. Yet, it is important to emphasize that the optimal instruments display large failure rates caused by perfect colinearity of the instruments. Finally, we plot the distribution of estimates for the non linear parameters when we use the interval instruments developed in Section 1.4. It appears clearly that the interval instruments yield a more concentrated distribution of estimates than the two other sets of instruments. For the sake of conciseness, we do not report the results with a mixture with 3 components but the observations we make with two components are even more exacerbated.

Table 1.14: Estimation mixture with "differentiation" instruments (1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.12	0.022	-0.016	-0.018	0.214	0.184	-0.022	-0.045	0.027
	$\sqrt{MSE}$	0.308	0.06	0.215	0.215	0.633	0.734	0.281	0.35	0.075
T=50, J=20	bias	-0.064	0.011	-0.01	-0.011	0.189	0.347	0.022	-0.081	0.025
	$\sqrt{MSE}$	0.231	0.044	0.165	0.166	0.566	0.887	0.184	0.291	0.059
T=100, J=12	bias	-0.058	0.01	-0.012	-0.012	0.233	0.226	0.02	-0.066	0.027
	$\sqrt{MSE}$	0.204	0.041	0.147	0.148	0.592	0.703	0.256	0.305	0.072
T=100, J=20	bias	-0.04	0.006	-0.007	-0.007	0.198	0.423	0.047	-0.101	0.025
	$\sqrt{MSE}$	0.165	0.032	0.117	0.116	0.552	0.89	0.164	0.27	0.055
T=200, J=12	bias	-0.038	0.007	-0.003	-0.003	0.184	0.167	0.011	-0.049	0.019
	$\sqrt{MSE}$	0.152	0.03	0.11	0.11	0.466	0.601	0.176	0.262	0.053

Table 1.15: Estimation mixture with "Optimal" instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.09	0.016	-0.012	-0.013	0.076	0.059	0.026	-0.111	0.01
	$\sqrt{MSE}$	0.296	0.057	0.234	0.232	0.361	0.483	0.212	0.281	0.036
T=50, J=20	bias	-0.046	0.007	0	0.001	0.074	0.11	0.028	-0.089	0.01
	$\sqrt{MSE}$	0.225	0.044	0.178	0.176	0.328	0.563	0.163	0.228	0.033
T=100, J=12	bias	-0.041	0.007	-0.004	-0.003	0.054	0.037	0.019	-0.066	0.007
	$\sqrt{MSE}$	0.202	0.039	0.157	0.158	0.279	0.4	0.154	0.211	0.028
T=100, J=20	bias	-0.029	0.004	-0.003	-0.003	0.074	0.107	0.033	-0.074	0.01
	$\sqrt{MSE}$	0.153	0.03	0.126	0.124	0.311	0.52	0.129	0.194	0.034
T=200, J=12	bias	-0.029	0.005	-0.001	-0.001	0.026	0.011	0.021	-0.061	0.004
	$\sqrt{MSE}$	0.136	0.026	0.111	0.111	0.184	0.313	0.113	0.172	0.018

Table 1.16: Estimation mixture with Global Interval instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.154	0.029	-0.043	-0.045	0.017	0	-0.045	0.004	0.005
	$\sqrt{MSE}$	0.341	0.067	0.257	0.258	0.277	0.391	0.227	0.259	0.024
T=50, J=20	bias	-0.092	0.017	-0.02	-0.021	0.013	0.042	-0.018	-0.003	0.004
	$\sqrt{MSE}$	0.245	0.048	0.19	0.19	0.248	0.415	0.166	0.22	0.021
T=100, J=12	bias	-0.07	0.013	-0.017	-0.019	0.004	-0.012	-0.027	0.005	0.002
	$\sqrt{MSE}$	0.2	0.039	0.161	0.161	0.167	0.282	0.157	0.201	0.013
T=100, J=20	bias	-0.047	0.008	-0.006	-0.007	-0.009	-0.005	-0.008	-0.009	0.001
	$\sqrt{MSE}$	0.158	0.031	0.13	0.129	0.115	0.264	0.115	0.169	0.005
T=200, J=12	bias	-0.039	0.007	-0.004	-0.003	-0.006	-0.027	-0.015	-0.001	0.001
	$\sqrt{MSE}$	0.141	0.027	0.109	0.109	0.088	0.219	0.108	0.164	0.003

Table 1.17: Estimation mixture with Local Interval instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.134	0.025	-0.023	-0.024	-0.006	-0.005	-0.039	-0.001	0.003
	$\sqrt{MSE}$	0.307	0.059	0.26	0.259	0.251	0.34	0.214	0.244	0.019
T=50, J=12	bias	-0.084	0.016	-0.024	-0.025	0.019	0.033	-0.023	0.01	0.003
	$\sqrt{MSE}$	0.245	0.047	0.188	0.186	0.228	0.38	0.15	0.184	0.018
T=50, J=12	bias	-0.075	0.015	-0.018	-0.016	0	0	-0.028	0.007	0.001
	$\sqrt{MSE}$	0.199	0.039	0.159	0.16	0.127	0.225	0.143	0.164	0.005
T=50, J=12	bias	-0.039	0.007	-0.011	-0.011	-0.003	0.004	-0.01	0.004	0.001
	$\sqrt{MSE}$	0.162	0.032	0.129	0.129	0.104	0.226	0.103	0.125	0.004
T=50, J=12	bias	-0.037	0.007	-0.008	-0.007	0.002	-0.007	-0.016	0.006	0.001
	$\sqrt{MSE}$	0.136	0.026	0.11	0.109	0.091	0.174	0.099	0.123	0.003

Figure 1.10: Distribution of estimates for non-linear parameters with "Differentiation" instruments $(T = 200, J = 12)$

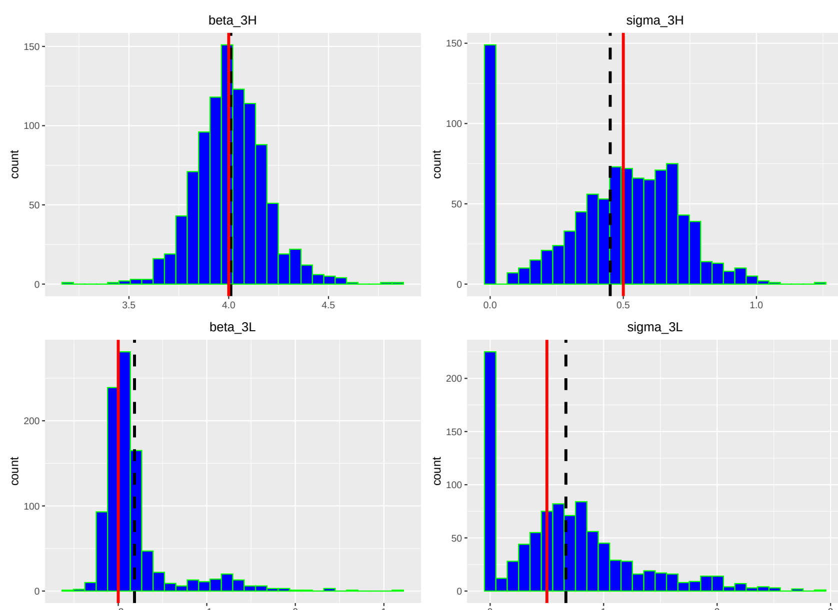


Figure 1.11: Distribution of estimates for non-linear parameters with "Optimal" instruments $(T = 200, J = 12)$

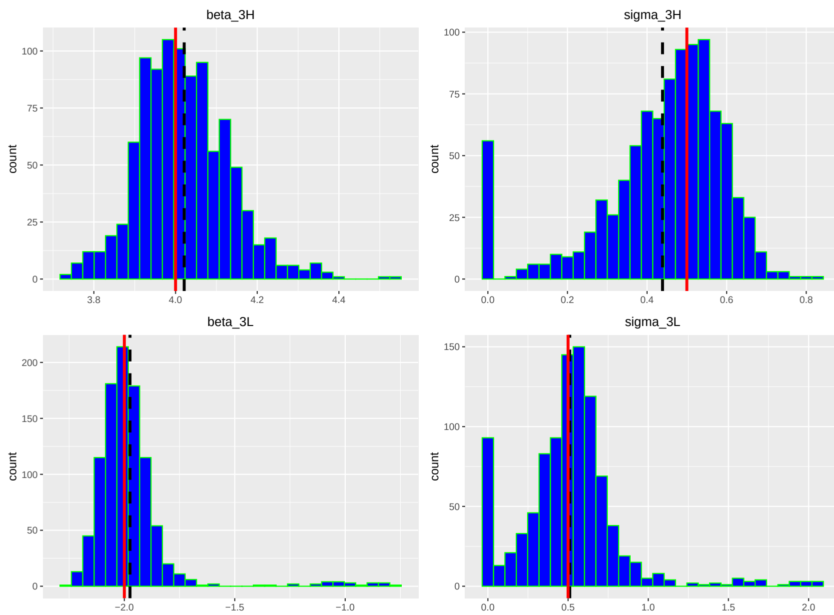


Figure 1.12: Distribution of estimates for non-linear parameters with "Global Interval" instruments $(T = 200, J = 12)$

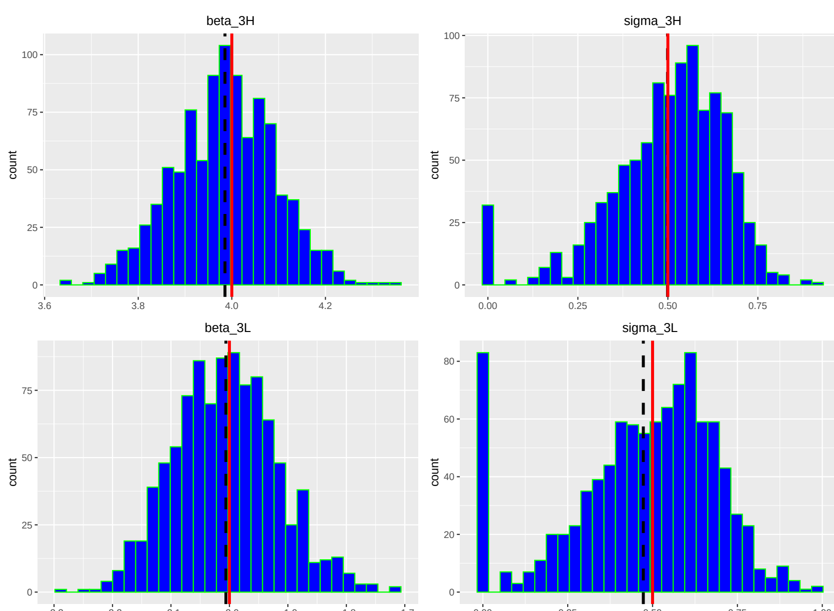
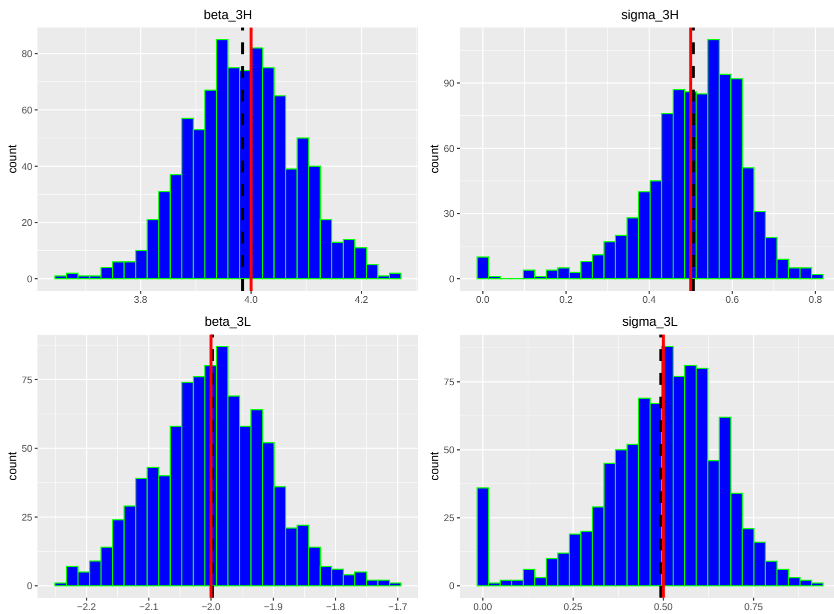


Figure 1.13: Distribution of estimates for non-linear parameters with “Local interval” instruments $(T = 200, J = 12)$



Estimation with a single Gaussian

Table 1.18: Estimation with a single Gaussian (1000 replications)

	Instruments	Differentiation						”Optimal”						Interval Global						Interval Local					
	Parameter	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3
Sample size	true	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5
T=50, J=12	bias	-0.16	0.032	-0.031	-0.028	-0.032	-0.004	-0.09	0.018	-0.016	-0.014	-0.018	-0.003	-0.15	0.03	-0.028	-0.026	-0.03	-0.004	-0.15	0.03	-0.028	-0.026	-0.03	-0.001
	$\sqrt{MSE}$	0.292	0.057	0.212	0.209	0.138	0.069	0.27	0.053	0.214	0.211	0.138	0.067	0.288	0.056	0.212	0.209	0.138	0.066	0.286	0.056	0.212	0.209	0.138	0.064
T=50, J=20	bias	-0.091	0.018	-0.022	-0.022	-0.015	0.001	-0.047	0.009	-0.013	-0.013	-0.006	0.001	-0.084	0.017	-0.021	-0.021	-0.013	0	-0.086	0.017	-0.021	-0.021	-0.014	0.002
	$\sqrt{MSE}$	0.209	0.041	0.159	0.16	0.106	0.05	0.199	0.039	0.16	0.161	0.106	0.05	0.206	0.041	0.16	0.16	0.106	0.052	0.208	0.041	0.159	0.16	0.106	0.052
T=100, J=12	bias	-0.088	0.017	-0.001	0	-0.027	0.001	-0.052	0.01	0.007	0.007	-0.02	0.001	-0.082	0.016	0	0.001	-0.026	0.001	-0.074	0.014	-0.016	-0.016	-0.013	0.001
	$\sqrt{MSE}$	0.199	0.039	0.146	0.145	0.1	0.045	0.189	0.037	0.148	0.147	0.099	0.047	0.197	0.039	0.146	0.146	0.1	0.044	0.185	0.036	0.151	0.152	0.099	0.044
T=100, J=20	bias	-0.043	0.009	-0.011	-0.012	-0.006	-0.001	-0.021	0.004	-0.007	-0.008	-0.002	-0.001	-0.04	0.008	-0.011	-0.012	-0.006	-0.001	-0.035	0.007	-0.01	-0.009	-0.004	0
	$\sqrt{MSE}$	0.145	0.028	0.115	0.114	0.075	0.035	0.141	0.028	0.115	0.114	0.075	0.035	0.145	0.028	0.115	0.114	0.076	0.035	0.14	0.027	0.116	0.115	0.076	0.035
T=100, J=20	bias	-0.038	0.007	-0.012	-0.012	-0.004	0.001	-0.017	0.003	-0.006	-0.007	-0.001	0	-0.032	0.006	-0.009	-0.01	-0.004	0	-0.033	0.006	-0.009	-0.01	-0.004	0.001
	$\sqrt{MSE}$	0.132	0.026	0.11	0.11	0.073	0.032	0.127	0.025	0.109	0.109	0.069	0.032	0.129	0.026	0.109	0.109	0.069	0.032	0.129	0.026	0.109	0.109	0.069	0.031

1.E Empirical application

1.E.1 First stage regression: instruments on price

In Table 1.19 we present the first stage regression for the endogenous variable, price. The explanatory variables include exogenous characteristics as well as the excluded instruments we presented in 1.7. We find that the excluded instruments are jointly significant with an F-stat of 467.41. As expected, we see that the steel futures price and its interaction with the weight of the car correlate positively with the price. We also see that the higher the exchange rate between the Euro and the local currency at the place of assembly, the lower the price of the car as the cost of production decreases. Moreover, we also see that if the location places a role as the European (country of assembly) dummy is negatively correlated with price. This could point to shipping expenses that are reflected in the price. Note that the effect of the labor costs is not as strong and not of the expected sign. Although we control for missing labor costs in the data with a missing dummy variable, it could still be causing a bias for the coefficient. Finally, the correlation between the competition-related instruments and the price shows that the degree of competition across cars of the same class matters.

Table 1.19: First stage regression for price

	Price/income
Labor costs (hourly)	-0.0002^{**}
	(0.0001)
Steel (futures) price	-0.0001^{***}
	(0.00001)
Steel (futures) price * Weight	0.00003^{***}
	(0.00000)
# Cars by engine-type	0.001^{***}
	(0.0001)
# Cars by engine-type and class	-0.002^{***}
	(0.00004)
Exchange rate (non European)	-0.0002^{***}
	(0.00002)
Europe dummy	-0.018^{***}
	(0.003)
Horsepower	0.527^{***}
	(0.003)
Gasoline	-0.057^{***}
	(0.003)
Fuel cost	-0.003^{***}
	(0.001)
Size	0.037^{***}
	(0.002)
Foreign	-0.008^{***}
	(0.003)
Height	0.062^{***}
	(0.009)
Observations	38,999
R^2	0.896
Adjusted R^2	0.896
F Statistic	$3,167.373^{***}$ (df = 65; 38933)

Note:Brand, Year and State FE's are included.

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

1.E.2 Baseline specifications: logit and nested logit

Table 1.20 shows the results from the logit and nested logit specifications. We define the nests by the class of the car, therefore limiting the substitution between the cars that belong to the same class within a nest.[43] We observe that the estimates are stable across specifications.

[43]Car classes in the data are: Mini, small, lower-middle, middle, upper-middle, luxury.

Table 1.20: Estimation results - Logit and Nested Logit

	OLS	*IV*			
	(1)	(2)	(3)	(4)	(5)
Price/income	-0.354^{***}	-2.907^{***}	-2.356^{***}	-2.729^{***}	-2.615^{***}
	(0.041)	(0.133)	(0.124)	(0.053)	(0.052)
log(within market shares)				0.420^{***}	0.407^{***}
				(0.006)	(0.006)
Fuel Cost	-0.210^{***}	-0.138^{***}	-0.247^{***}	-0.074^{***}	-0.126^{***}
	(0.008)	(0.006)	(0.009)	(0.004)	(0.006)
Size(m^2)	0.031	0.001	0.158^{***}	-0.001	0.104^{***}
	(0.038)	(0.040)	(0.041)	(0.025)	(0.026)
Horsepower(KW/100)	0.136	3.151^{***}	2.511^{***}	2.586^{***}	2.431^{***}
	(0.089)	(0.183)	(0.172)	(0.080)	(0.078)
Foreign	0.351^{***}	0.083	0.120^{*}	-0.106^{**}	-0.101^{**}
	(0.064)	(0.073)	(0.070)	(0.046)	(0.044)
Height(m)	0.870^{***}	1.505^{***}	3.487^{***}	1.121^{***}	2.270^{***}
	(0.216)	(0.197)	(0.228)	(0.125)	(0.145)
Gasoline	1.399^{***}	0.625^{***}	1.118^{***}	0.190^{***}	0.422^{***}
	(0.055)	(0.061)	(0.063)	(0.039)	(0.041)
Fuel cost × income	0.020^{***}	-0.002^{**}	0.014^{***}	-0.002^{***}	0.007^{***}
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Size × income	-0.005^{***}	-0.002^{***}	-0.006^{***}	0.0003	-0.002^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Horsepower × income	0.009^{***}	-0.026^{***}	-0.017^{***}	-0.027^{***}	-0.024^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Horsepower × time	-0.084^{***}	-0.068^{***}	-0.083^{***}	-0.038^{***}	-0.045^{***}
	(0.006)	(0.007)	(0.007)	(0.004)	(0.004)
Foreign × income	-0.019^{***}	-0.015^{***}	-0.016^{***}	-0.008^{***}	-0.008^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Height × income	-0.006	0.032^{***}	-0.002	0.016^{***}	-0.003
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)
Height × density	-0.037^{***}	-0.003^{***}	-0.037^{***}	-0.001^{***}	-0.021^{***}
	(0.004)	(0.0003)	(0.004)	(0.0002)	(0.003)
Gasoline × income	-0.016^{***}	-0.003^{***}	-0.010^{***}	0.0004	-0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gasoline × Post 2015		-0.024		-0.019	
		(0.019)		(0.012)	
Constant	-7.937^{***}	-12.482^{***}	-11.171^{***}	-9.144^{***}	-8.506^{***}
	(0.167)	(0.149)	(0.167)	(0.092)	(0.102)
State FE/ Year FE	*Yes*	*No*	*Yes*	*No*	*Yes*
Observations	39,888	39,888	39,888	39,888	39,888
R^2	0.385	0.217	0.272	0.686	0.709
Adjusted R^2	0.384	0.216	0.271	0.686	0.709

*p<0.1; **p<0.05; ***p<0.01

Note: Brand FE's are included.

1.E.3 Counterfactual quantities under different specifications

We define quantities of interest and derive them under the different specifications considered previously. For exposition purposes, we omit the dependence of the market shares in δ_t, x_{2t} and f, and simply write $s_j(\mathbf{p})$ instead of $\rho_j(\delta_t, x_{2t}; f)$, where $\mathbf{p}$ is the price vector.

Price elasticities. For the calculation of the price elasticities one can refer to 1.D.1 that writes the quantities for the simulation exercise.

Demand curvature. The demand curvature is defined using second derivative of demand as follows: $\eta_j^2(\mathbf{p}) = s_j(\mathbf{p}) \frac{\partial^2 s_j(\mathbf{p})}{\partial p_j^2} \left(\frac{\partial s_j(\mathbf{p})}{\partial p_j} \right)^{-2}$.

Marginal costs and mark-ups. To recover the marginal costs and the implied mark-ups, we need to make additional assumptions on the supply side. Following the literature, we consider that each multi-product firm $f \in F$ sets prices for its own products in accordance with a Bertrand-Nash equilibrium. The profit of each firm writes:

$$\Pi_f(\mathbf{p}) = \sum_t \sum_{j \in J_f} (p_j - c_j) M_t s_{jt}(\mathbf{p})$$

where J_f is the set of goods produced by firm f, c_j is the marginal cost for good j, M_t is the market size and $s_j(p)$ is the market share of product j. The first-order condition with respect to price p_j writes:

$$\sum_t M_t \, s_{jt}(\mathbf{p}) + \sum_t M_t \sum_{j' \in J_f} \left(p_{j'} - c_{j'} \right) \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} = 0.$$

We gather all the FOCs and rewrite them in matricial form:

$$\mathbf{s}(\mathbf{p}) + (\boldsymbol{\Delta}(\mathbf{p}))(\mathbf{p} - \mathbf{c}) = 0.$$

where $\boldsymbol{\Delta}(\mathbf{p}) = \sum_t M_t \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j}$ if j' and j are produced by the same firm and equals to zero otherwise. $\boldsymbol{\Delta}(\mathbf{p})$ is known as the ownership matrix. Assuming that the prices are in equilibrium, one can recover the marginal costs using the following equation:

$$\mathbf{c} = \mathbf{p} - (\Delta(\mathbf{p}))^{-1} \mathbf{s}(\mathbf{p}). \tag{1.E.17}$$

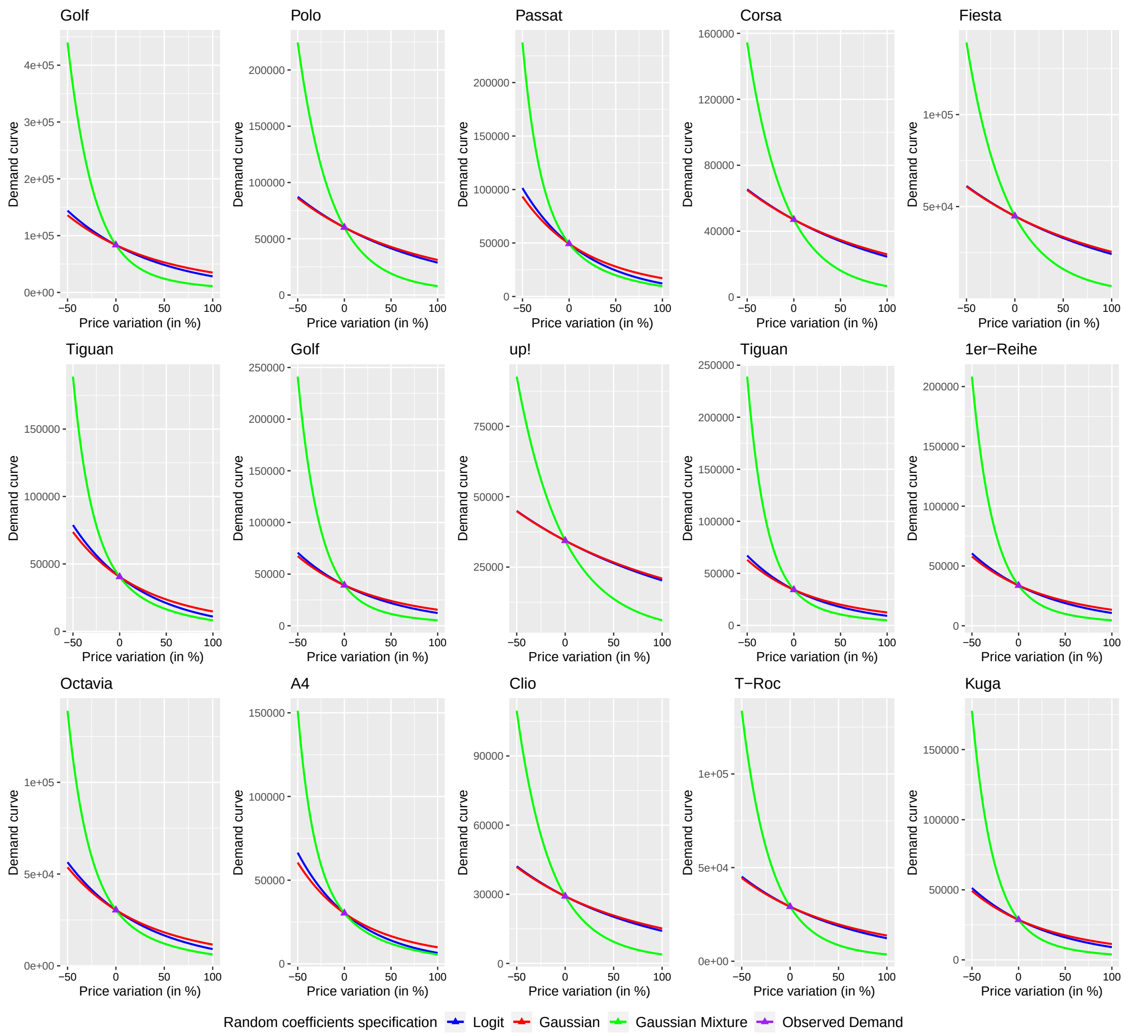
The mark-up for product j simply writes: $p_j - c_j$.

Pass-through The pass-through of cost is defined as follows. Let us assume that the marginal cost for product j goes from c_j to c'_j (with $c'_j > c_j$), then the cost pass-through equals $\alpha_j = \frac{p'_j - p_j}{c'_j - c_j}$, where p'_j is the new equilibrium price. We calculate the new equilibrium price using Eqn. 1.E.17 using fixed point iteration. The pass-through corresponds to the proportion of the cost increase that is transmitted to the price.

Table 1.21: Counterfactual quantities under different specifications on RCs (20 most popular cars)

Counterfactual quantity		Price elasticity			Curvature			Marginal cost			Mark-up			Pass-through		
Car	Manufacturer	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture
Golf	Volkswagen	-1.09	-0.95	-3.03	1.00	1.14	1.21	1260	-9670	15436	24098	35028	9922	0.92	-	1.30
Polo	Volkswagen	-0.74	-0.70	-2.50	1.00	1.15	1.09	-6643	-14366	9073	23819	31542	8103	1.05	-	1.09
Passat	Volkswagen	-1.43	-1.21	-2.27	1.00	1.17	1.57	9488	-1033	17826	24631	35153	16294	1.02	-	2.65
Corsa	PSA	-0.66	-0.63	-2.28	1.00	1.14	1.07	-8432	-11246	8410	24088	26902	7246	1.02	-	1.12
Fiesta	Ford	-0.62	-0.60	-2.18	1.00	1.15	1.07	-8983	-10806	7657	23487	25310	6847	1.03	-	1.10
Tiguan	Volkswagen	-1.32	-1.14	-2.28	1.00	1.17	1.55	6831	-2919	16211	24118	33868	14738	1.01	-	2.62
Golf	Volkswagen	-1.17	-1.03	-3.12	1.00	1.18	1.27	3128	-7932	16582	23828	34888	10374	0.99	-	1.41
up!	Volkswagen	-0.53	-0.52	-1.92	1.00	1.14	1.05	-11231	-17703	4594	23278	29749	7453	1.04	-	0.96
Tiguan	Volkswagen	-1.34	-1.15	-3.09	1.00	1.19	1.38	7051	-4117	19186	23842	35009	11706	1.01	-	1.66
1er-Reihe	BMW	-1.16	-1.03	-3.09	1.00	1.18	1.28	3845	-769	19179	25138	29753	9805	0.99	-	1.39
Octavia	Volkswagen	-1.23	-1.08	-2.33	1.00	1.17	1.50	4629	-4504	15464	24211	33345	13377	1.01	-	2.34
A4	Volkswagen	-1.56	-1.30	-2.26	1.00	1.19	1.56	13209	1995	20260	25865	37079	18814	1.01	-	2.66
Clio	Renault	-0.73	-0.70	-2.49	1.00	1.16	1.10	-6240	-8684	9817	23120	25563	7063	1.03	-	1.17
T-Roc	Volkswagen	-0.87	-0.81	-2.80	1.00	1.17	1.14	-3645	-12275	11578	23798	32427	8575	1.06	-	1.16
Kuga	Ford	-1.16	-1.03	-3.09	1.00	1.18	1.28	3654	-518	18214	23684	27856	9124	1.03	-	1.39
Golf	Volkswagen	-1.10	-0.99	-2.34	1.00	1.16	1.44	1548	-7284	13678	23929	32762	11799	0.96	-	2.13
A-Klasse	Daimler	-1.28	-1.10	-3.07	1.00	1.19	1.35	6608	562	20662	25066	31112	11013	1.01	-	1.56
Golf	Volkswagen	-1.05	-0.94	-2.33	1.00	1.16	1.42	417	-8115	13135	24177	32710	11460	0.72	-	2.11
Golf	Volkswagen	-1.18	-1.05	-3.15	1.00	1.18	1.27	3202	-8230	16705	23921	35353	10418	0.98	-	1.40
Octavia	Volkswagen	-1.05	-0.95	-3.02	1.00	1.17	1.21	380	-8835	14808	23862	33077	9433	0.78	-	1.30

Figure 1.14: Estimated demand functions under different specifications



Chapter 2

Moment Inequalities for Entry Games with Heterogeneous Types

co-authored with Christian Bontemps and Rohit Kumar

Abstract

Following Bresnahan and Reiss (1991a), Bresnahan and Reiss (1991b) and Berry (1992), entry games have become a popular model in the empirical industrial organization literature. They enable researchers to study different features of an industry with easy-to-obtain data on entry. In this paper, we provide new tools to simplify the estimation of entry games when the equilibrium selection mechanism is unrestricted. In particular, we develop an algorithm that allows us to recursively select a relevant subset of inequalities and compute the theoretical upper bounds on the probability of each outcome (without having to simulate them). We also propose a new testing procedure that is asymptotically pivotal by smoothing the set defined by the moment inequalities. We show that this new estimation procedure can seamlessly accommodate covariates, including continuous ones. We conduct full-scale Monte Carlo simulations to assess the performance of our new estimation procedure.

Keywords: empirical entry games, moment inequalities, core determining class, smoothing.

2.1 Introduction

In the wake of the seminal contributions by Bresnahan and Reiss (1991a), Bresnahan and Reiss (1991b) and Berry (1992), entry games have become a popular model in the empirical industrial organization literature (IO). They allow researchers to study the determinants of firm profitability and the degree of competition from data on entry, which is usually easy to collect. Entry games can also serve as a building block to multi-stage games involving for instance price competition (Eizenberg (2014), Ciliberto, Murry, and Tamer (2021)). Among the most influential applications, one can mention the analysis of competition, market structure and regulation in various industries including airlines (Ciliberto and Tamer (2009), Berry (1992)), retailers (Cleeren, Verboven, Dekimpe, and Gielens (2010), Aradillas-Lopez and Rosen (2022), Andrews, Berry, and Jia (2004), Grieco (2014)...), motels (Mazzeo (2002)) and fast food restaurants (Toivanen and Waterson (2005)).

This paper provides a novel estimation strategy for static entry games of complete information, which significantly simplifies the estimation procedure when no restrictions are imposed on the equilibrium selection mechanism. It is a well-known difficulty that in the absence of a known equilibrium selection mechanism, the estimation is substantially complicated by the presence of multiple equilibria. Namely, there are regions in the space of unobservable shocks in which the entry game predicts multiple equilibria and yet the econometrician only observes one outcome. Without additional information on the equilibrium selection mechanism, the model is said to be incomplete and the econometrician cannot rely on standard estimation techniques. To tackle this problem, the literature proposes various solutions, which can be divided in two distinct categories. The first approach consists in restricting the equilibrium selection mechanism so as to complete the model. There are various ways of proceeding. The econometrician can impose an order of entry (Mazzeo (2002), Cleeren, Verboven, Dekimpe, and Gielens (2010)...). Bjorn and Vuong (1984) suggests to randomly draw an equilibrium out of the multiple potential equilibria. Grieco (2014) and Bajari, Hong, and Ryan (2010) explicitly model the equilibrium selection mechanism. Building on the vast literature related to set identification initiated by Manski (1995), the second generic solution consists in characterizing the set of parameters which can generate the observed data without restricting the equilibrium selection mechanism (among prominent examples of this approach in the context of games, see Ciliberto and Tamer (2009), Beresteanu, Molchanvb, and Molinari (2012), Galichon and Henry (2011), Bontemps and Kumar (2020), Aradillas-Lopez and Rosen (2022), Chesher and Rosen (2019), Magnolfi and Roncoroni (2022)). Despite the risk of misspecification implied by incorrect restrictions on the equilibrium selection mechanism, heretofore, the empirical literature has largely favored the first approach due to the relative simplicity of its implementation. The second family of solutions is less restrictive but its implementation faces major theoretical and practical challenges. First, even in seemingly harmless games, characterizing the sharp identified set (the set of admissible parameters, which satisfy all the inequalities implied by the model) can be a grueling task. (i) The number of inequalities generated by the model increases exponentially with the number of players and can quickly become overwhelming[1]. (ii) Except for a few toy models, there are no closed form expressions for the theoretical bounds implied by the model and one must resort to simulation methods, which mechanically induce biases. Additionally, the estimation of the identified set also poses many challenges. (i) The exact asymptotic distribution of the test statistic under the null depends on the set of binding moments, which is unknown to the econometrician. This seriously complicates the derivation of the critical value. The methods proposed in the literature either rely on simulation methods, which are computationally intensive[2], or upper bounds, which are conservative. Moreover, the finite sample performance of these methods is known to decrease steadily with the inclusion of many moment inequalities, which mechanically inflate the critical value. (ii) The presence of exogenous covariates (and in particular, continuous ones) complicates even more the estimation as the identified set is now characterized by conditional moment inequalities that must be converted into unconditional ones. (iii) Finally, the esti-

[1] For classical entry games, with N players, the total number of inequalities is 2^{2^N}

[2] sub-sampling, bootstrap or simulation of the asymptotic distribution and these methods even if they mitigate the inferential loss, still yield conservative critical values

mation of confidence region for the structural parameters is based on a test inversion over a grid, which can quickly become very large if the dimension of the structural parameter θ increases. Therefore, the objective of this paper is to remove or mitigate most of the difficulties exposed above, and thus, facilitate and encourage the estimation of static entry games in empirical work, while remaining agnostic about the equilibrium selection mechanism.

We now briefly summarize the main methodological improvements that we initiate in the paper. The baseline model we study is a generic static entry game with types, which corresponds to a generalization of entry games where some players are pooled together according to their characteristics. By regrouping some of the competitors in a same category, we can substantially increase the number of potential entrants, while keeping the number of parameters to estimate low[3]. The first part of the paper addresses the challenges related to identification. Each candidate parameter induces a graph on the set of outcomes, which is such that there is a link between two outcomes if their equilibrium regions overlap. To reduce the number of inequalities that sharply characterize the identified set, we leverage this graph over the set of outcomes induced by each candidate parameter θ. The novelty in this paper is to provide a systematic way of deriving the graph and inferring the subset of relevant inequalities. As for the computation of the theoretical bounds, we show how to derive them by exploiting the inclusion-exclusion formula and observing that intersection regions are cubes, for which the bounds can be easily derived[4]. In the second part of the paper, we tackle the issues related to estimation. To mitigate the inferential loss due to the inability to recover the exact asymptotic distribution in the context of moment inequalities, we develop an alternative approach which consists in smoothing the identified set in order to recover a test statistic with a known and pivotal asymptotic distribution. The smoothed set that we estimate is an outer set of the sharp identification set, which we make converge to the sharp identified set by letting the amount of smoothing decrease with the sample size. The general philosophy of this approach can be linked to

[3] Thus, the introduction of types help reduce the size of grid that we need to explore in the estimation

[4] allowing the theoretical probabilities to be easily derived by integrating over cubes

the common bias variance trade off which appears in most econometric problems. We provide a general guideline on how to optimally choose the smoothing parameter. Last but not least, we show that this smoothing procedure facilitates the inclusion of covariates into the model, which represents a major improvement with respect to the rest of the literature.

Related literature While this paper focuses essentially on the estimation of entry games, some of the tools we develop in this paper apply more broadly to the estimation of models characterized by moment inequalities. In this sense, this paper contributes to the rich literature on conditional and unconditional moment inequalities, which includes, among others, contributions by Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Beresteanu and Molinari (2008), Andrews and Soares (2010), Bontemps, Magnac, and Maurin (2012), Romano, Shaikh, and Wolf (2014), Chernozhukov, Chetverikov, and Kato (2018b), Andrews and Shi (2013), Armstrong and Chan (2016), Armstrong (2014), Molchanov and Molinari (2014), Bugni, Canay, and Shi (2017),Cox and Shi (2022), Chen, Christensen, and Tamer (2018), Kitamura and Stoye (2018), Kaido, Molinari, and Stoye (2019), Andrews, Roth, and Pakes (Forthcoming), Cho and Russell (2018), Gafarov (2019), Berry and Compiani (2022).

Structure of the paper. The remainder of the paper is organized as follows. Section 2 describes the general set-up as well as the standard assumptions we impose on the model. In section 3, we characterize the identified set and we present a practical approach to the selection of relevant inequalities and the derivation of the theoretical bounds. In section 4, we present our novel estimation strategy which builds on smoothing the identified and we compare our approach with more conventional procedures. In section 5, we provide some Monte Carlo simulations to assess the performance of our estimation procedure in comparison to alternative strategies proposed in the literature. The proofs are given in the Appendix.

2.2 The model

We consider a flexible entry game model in the spirit of the models already developed in the literature (Berry (1992), Aradillas-Lopez and Rosen (2022) and Cleeren, Verboven, Dekimpe, and Gielens (2010) among others). We pool similar competitors in categories of players (which we refer to as a types or formats). In this setup, profit functions are heterogeneous across types and homogeneous within each type.[5] As we will see later, there is a trade-off between the accuracy of the inference procedure and the flexibility of the model. Pooling the different competitors in types results in a substantial decrease in the number of parameters to estimate while keeping a large number of potential entrants. For example, in the airline industry, marginal Low Cost Airlines are not present in all the markets and it seems reasonable to consider that the market structure depends more on how Low Cost Airlines are exploiting direct flights between two cities rather than which Low cost airline is exploiting these lines. Similarly, in the retail industry, firms of the same format have the same business model and people go shopping at the local hypermarket or one of the closest hard-discounters, whatever their specific brand. Also, models with discrete outcomes like the one of Aradillas-Lopez and Rosen (2022) (known as ordered response games) have a similar structure to the general setup we study in this model and the methods below can be adapted to this case. Finally, we assume that the types of firms are predetermined and not endogenously chosen. If there is free-entry, this assumption is not restrictive and results like the ones in Mazzeo (2002) can be derived similarly. We now describe our model in greater detail and we study its equilibrium structure.

2.2.1 Payoff for entering firms

In a given market m, the profit of an entering firm of type t depends on the number of entering firms $N_{t,m}$ of each type, $t = 1, \ldots, T$, gathered into one vector $\mathcal{N}_m = (N_{1,m}, N_{2,m}, \ldots, N_{T,m})$. It also depends on a

[5] We want to emphasize that traditional entry games in which all the players have different profit functions are simply a special case of this model in which all types can have at most 1 player (hence, in this specific case, each type represents a single player).

vector of dimension d market and type characteristics $X_{t,m}$ and a firm profit shock $\varepsilon_{t,m}$, which is market and type specific, drawn from a parametric distribution $F_\eta(\cdot)$, $\eta \in \Lambda \subset \mathbb{R}^q$ and independent from the characteristics. Formally, we have for each market m:

$\forall t \in \{1, \ldots, T\}$,

$$\Pi_{t,m} = \pi_t(X_{t,m}, N_{t,m}, \mathbf{N}_{-t,m}; \omega) + \varepsilon_{t,m},$$

in which the function π_t is parametrized by parameter $\omega \in \mathbb{R}^{q'}$ and $\mathbf{N}_{-t,m}$ denotes the vector of the number of entering firms of type $t' \neq t$ in market m. Observe that we keep the possibility to get heterogeneous reactions on a type t firm's profit with respect to the potential entry of different types of firms. In the following parameter θ denotes the $q + q'$ vector of parameters gathering ω and η. θ_0 is the unknown true value.

Now, we impose some restrictions on the profit function that are consistent with economic theory.

Assumption 2 *The profit is decreasing with respect to the number of competitors, i.e.,* $\forall t \in \mathcal{T}$, $\pi_t(X_m, N_{t,m}, \mathbf{N}_{-t,m}; \omega)$ *is strictly decreasing in* $N_{t,m}$ *and weekly decreasing in any* $N_{t',m}$, $t' \neq t$.

Assumption 2 uses the fact that more competitors are worse for economic profitability of a firm of a given type t. Also, firms enter if their long run profit is weakly positive, otherwise receive a zero payoff. We assume that firms have complete information and thus observe all the profit shocks of their competitors when they decide to enter or not, contrary to the econometrician.[6] Also, when making their decisions, they do not observe the decisions of the other firms, and thus all make simultaneous moves. We focus on pure strategy Nash Equilibria (NE hereafter), like most of the literature (Berry (1992), Ciliberto and Tamer (2009), Bontemps and Kumar (2020) or Aradillas-Lopez and Rosen (2022)). As it is well known, different equilibria concepts can be considered (mixed strategy or correlated equilibria),

[6]Grieco (2014) and Bajari, Hong, Krainer, and Nekipelov (2006) provide identification and estimation strategies to tackle games of incomplete information.

and the solutions proposed in this paper can be adapted to these settings by modifying the set of moment inequalities which are derived (see, in particular, Beresteanu, Molchanov, and Molinari (2011) or Magnolfi and Roncoroni (2022)).

Additionally, we assume the following boundary conditions:

Assumption 3

- $\forall t \in \mathcal{T}$, $\pi_t(X_{t,m}, 0, \mathbf{N}_{-t,m}; \omega) = +\infty$.
- $\forall t \in \mathcal{T}$, $\lim_{N_t \to +\infty} \pi_t(X_{t,m}, N_t, \mathbf{N}_{-t,m}; \omega) = -\infty$.
- *The distribution of the profit shocks ε, $F_\eta(\cdot)$, is absolutely continuous on $\mathbb{R}^T$ with full support and mean 0.*

Assumption 3 is standard. The first two ones are only a normalization to calculate for the first one the probability of no entry for a given type given the number of entrants of the other types and to ensure the finiteness of $\mathcal{N}$ for the second one. The third one ensures that for any X the probability to observe no entry is always strictly positive. Less restrictive assumptions can be made to ensure the same requirements like in Aradillas-Lopez and Rosen (2022) but ours is not very restrictive. In the remaining of the paper, we refer to the following model to illustrate our results. As a pedagogical example of our general model, we consider this simple 2-type game with linear profit functions and no covariates.

Example 1 *Profit functions of firms of type 1 and 2 write as follows, omitting the subscript m for the ease of the exposition:*

$$\Pi_1 = \beta_1 - \delta_{1,1} N_1 - \delta_{2,1} N_2 + \varepsilon_1$$
$$\Pi_2 = \beta_2 - \delta_{1,2} N_1 - \delta_{2,2} N_2 + \varepsilon_2$$

where

- *N_t is the number of firms of type $t = 1, 2$, active in the markets*
- *ε_t unobserved heterogeneity for types $t = 1, 2$,*

- $\theta = (\beta_1, \beta_2, \delta_{1,1}, \delta_{2,1}, \delta_{1,2}, \delta_{2,2})$ *is the parameter of interest which we seek to identify.* δ*'s capture competition effects within each type and between types.*

Such a model satisfy Assumptions 2 and 3.

We allow a maximum of 3 potential entrants of each type $t = 1, 2$*. Unobserved shocks are normally distributed and uncorrelated.*

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

The competitive effect is driven by the vector of parameters δ_t *which ensure that intra-format and inter-format competition can differ and that the competitive effect may differ across type (a testable assumption).*[7]

2.2.2 Equilibrium Structure

To lighten the notations, we also drop the market index m from now on. An outcome $y = (N_1, ..., N_T)$ is a NE if each type t number of entrants N_t is a best response to the other types' number of entrants, $\mathbf{N}_{-t}$. We recall that we assume free entry without loss of generality. Therefore, for each type t, we have:

- First, it is profitable for any type t firm which is entering to operate in such a market structure, i.e.,

$$\pi_t(X, N_t, \mathbf{N}_{-t}; \omega) + \varepsilon_t \geq 0.$$

- Second, an additional entrant of the same type would lead to negative profit for any type t firm, i.e.,

$$\pi_t(X, N_t + 1, \mathbf{N}_{-t}; \omega) + \varepsilon_t < 0.$$

As a result, we have the following necessary condition for $\mathcal{N} = (N_1, ..., N_T)$ to be a Pure Strategy Nash Equilibrium.

[7] In the pictures drawn across the text, the values chosen for the parameters are $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$

Proposition 2.1 *An outcome $Y = (N_1, ..., N_T)$ is a NE of our game if and only if: $\forall t \in \mathcal{T}$,*

$$-\pi_t(X, N_t, \mathbf{N}_{-t}; \omega) \leq \varepsilon_t \leq -\pi_t(X, N_t + 1, \mathbf{N}_{-t}; \omega). \tag{2.2.1}$$

In the following, we denote this region $\mathcal{R}_\omega(X, Y)$.

It is possible to find instances of the generic game we study for which there exists one or more regions in the space of unobserved heterogeneity which cannot sustain any pure strategy NE[8]. If this is the case, the model is said to be incoherent. Chesher and Rosen (2019) provides various ways to tackle this issue. In our case, we abstract away from this issue by assuming that the data can only be generated by a coherent model (meaning that if a given θ yields an incoherent model, then it cannot belong to the identified set).

It is well-known that for a given shock ε_t, the game may generate multiple equilibria. Equivalently, the equilibria regions $\mathcal{R}_\omega(X, Y)$ of different outcomes Y may overlap.

Figure 2.1 displays the equilibrium regions ($\mathcal{R}(N_1, N_2)$) for our leading example.[9]. We can see that $\mathcal{R}_\omega(1, 0)$ overlaps with $\mathcal{R}_\omega(0, 1)$, i.e., for such a draw of profit shocks, either $(1, 0)$ or $(0, 1)$ is an outcome but we can't say which one is the realized one. In other words, we have multiplicity of equilibria, that is, there are regions of realizations of ε which do not predict a single outcome. We call them, with an abuse of language, multiple equilibria regions. In the absence of a known equilibrium selection mechanism, there is no longer a one-to-one mapping between the set of observed outcomes and the regions of profit shocks, which prevents the econometrician from using the usual identification and estimation procedures. The model is said to be *incomplete*.

The most straightforward way to circumvent the multiplicity issue is to impose restrictions on the equilibrium selection mechanism so as to ensure equilibrium uniqueness in each region of the space of unobserved heterogeneity. This is by far the approach that has gained the most traction in the empirical

[8] Berry (1992) proves the existence of a PSNE for every value of ε, X, θ in entry games with homogeneous competition where the profit functions are only affected by the total number of competitors entering the market and not their identity

[9] In the graph, we omit the X in the notations as there are no covariates in this example.

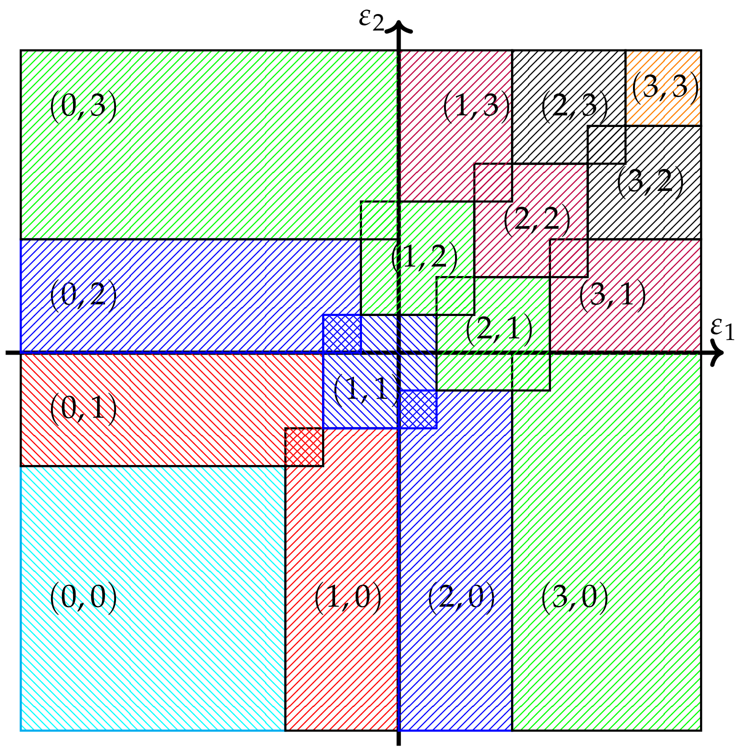


Figure 2.1: Equilibrium structure for $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$

literature. There are various ways of restricting the equilibrium selection mechanism: Mazzeo (2002) and Cleeren, Verboven, Dekimpe, and Gielens (2010) impose an order of entry over types, Bajari, Hong, and Ryan (2010) explicitly models the equilibrium selection mechanism as a parametric function that can be estimated[10]. By constraining the equilibrium selection mechanism, the econometrician forces each region of the space of unobserved heterogeneity to yield a unique equilibrium. We say that imposing an equilibrium selection completes the model. The econometrician is then able to associate a well-defined probability to each observed outcome.

In our example, if we impose that firms of type 1 always decide first (i.e. before firms of type 2), then the predictions of the model are unique as illustrated in Figure 2.2. As a result, a likelihood can be derived and we are back to the standard procedure. However, this strategy suffers from a huge specification risk. Another alternative, exploited in Berry (1992) in particular, is to look for a combination of outcomes that are invariant in the regions of multiple equilibria. Berry (1992) shows that the number of active firms is constant at the equilibrium. Furthermore, Cleeren, Verboven, Dekimpe, and Gielens (2010) shows that, with two types and additional mild restrictions on the profit function, this remains valid. However, it is linked to particular interaction structures and it is impossible to generalize this property to more than two types unless imposing strong restrictions on the horizontal positioning of the different types.

Another important strand of the literature leverages recent developments in the moment inequality literature to characterize the set of parameters that can generate the observed data without restricting the equilibrium selection. The idea is to exploit the inequalities implied by the model while abstaining from making assumptions about the equilibrium selection mechanism. Furthermore, this method allows the selection mechanism to differ from one market to the other. This strategy has been extensively studied by econometricians (Andrews, Berry, and Jia (2004),Ciliberto and Tamer (2009), Beresteanu, Molchanvb, and Molinari (2012), Galichon and Henry (2011), Bontemps and Kumar (2020)). Andrews,

[10]their setup is slightly different to ours: mixed strategies are allowed but no types are considered and stochastic shocks are action dependent

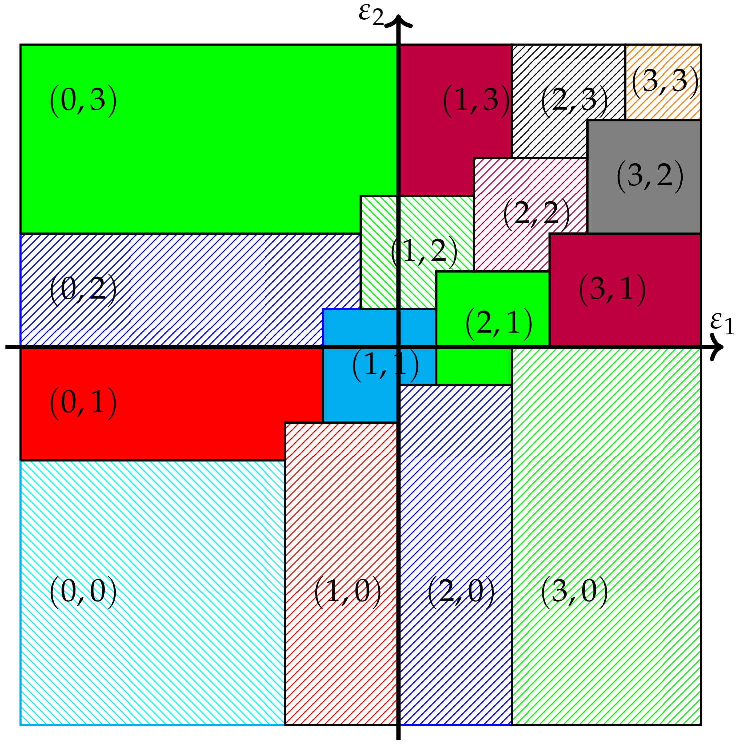


Figure 2.2: Equilibrium structure for $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$ when firms of type 1 enter first.

Berry, and Jia (2004) suggests deriving an upper bound on the probability of each individual outcome. Ciliberto and Tamer (2009) improves upon Andrews, Berry, and Jia (2004) by computing lower bounds on the probability of each outcome. Galichon and Henry (2011) proposes a sharp characterization of the identified set which exhausts all the moments implied by the model by deriving a theoretical upper bound on each subset of the set of outcomes $\mathcal{Y}$. Beresteanu, Molchanvb, and Molinari (2012) generates the set of moment inequalities required to sharply characterize the identified set but, as we show below, it becomes quickly numerically intractable. Bontemps and Kumar (2020) proposes, in a entry model á la Berry, a selection among this set of inequalities by looking for the adjacent vertex of a convex set and using only the inequalities related to this vertex. In the next section, we show how we select our inequalities by exploiting our specific structure.

2.3 Deriving the smallest set of moment inequalities

First, let us introduce a few additional notations. $\mathcal{Y}$ is the ordered set of possible outcomes of the game.[11] and $\mathcal{X}$ the support of the exogenous covariates X. For any subset A of $\mathcal{Y}$ ($A \in \mathcal{P}(\mathcal{Y})$), $P_\theta(A|X)$ denotes the conditional probability of $y \in A$ given X for a value θ and we denote it $P_0(A|X)$ when $\theta = \theta_0$, the true unknown value. We define the identified set for θ, Θ_I, as the collection of parameters which are observationally equivalent to the true value θ_0, i.e.,

$$\Theta_I = \left\{ \theta \in \mathbb{R}^{q+q'}, \forall y \in \mathcal{Y}, P_\theta(Y = y|X) = P_0(Y = y|X),\ Xa.s. \right\}.$$

Observe that Θ_I might be a true set or a point. Our procedure does not depend on the true nature of the set. Now, we characterize the identified set by a collection of inequalities.

[11] In our example, $\mathcal{Y} = \{(0,0),(1,0),\ldots,(3,3)\}$, i.e., 16 possible outcomes.

2.3.1 A sharp characterization of the identified set

Following Proposition 2.1, for any $x \in \mathcal{X}$, an outcome $y \in \mathcal{Y}$ is a possible equilibrium of the game if and only if the unobserved shock lies in $\mathcal{R}_\omega(x, y)$. Given that there exist multiple equilibria regions, the probability of ε to be in $\mathcal{R}_\omega(x, y)$ is an upper bound on the probability to observe the outcome y. Therefore, we get that for any $\theta \in \Theta_I$, the following inequality holds:

$$\forall x \in \mathcal{X},\ \forall y \in \mathcal{Y},\ P_0(Y = y|X = x) \leq \int_{\mathcal{R}_\omega(x,y)} dF_\eta(\varepsilon). \tag{2.3.2}$$

The inequality in (2.3.2) can be extended to any subset $A \subset \mathcal{Y}$ and following the propositions in Beresteanu, Molchanvb, and Molinari (2012) and Galichon and Henry (2011) in particular, we can characterize the identified set by a countable number of conditional moment inequalities:

$$\Theta_I = \{\theta \in \Theta \mid \forall A \in \mathcal{Y}, P_0(Y \in A|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A))\ X\ a.s.\}. \tag{2.3.3}$$

We call $P_0(Y \in A|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A))$ the (conditional moment) inequality generated by A. Unfortunately, this characterization of the identified set is hardly exploitable as such by the econometrician. Even for the simplest games, the number of inequalities characterizing the sharp identified set can already be overwhelming. For example, when $\mathcal{X}$ is finite, the number of inequalities is $\text{card}(\mathcal{P}(\mathcal{Y})) \times \text{card}(\mathcal{X})$. In the pedagogical example that we consider and if we assume that the exogenous cost shifter is degenerate ($\text{card}(\mathcal{X}) = 1$), the number of inequalities is already $\text{card}(\mathcal{P}(\mathcal{Y})) = 2^{16} = 65536$ which is tractable. With 5 players and three types, a case compatible with the retail or the airline industry, the number of inequalities increases to 2^{125} which is untractable.

Furthermore, with the exception of simple cases, it is not clear how to compute the theoretical bounds implied by the model without making use of simulation methods that necessarily induce biases. For the sake of clarity, let us abstract from the additional layer of difficulty induced by the exogenous covariates X and perform the identification analysis as if we condition on a given realization $X = x$. We postpone the discussion on the inclusion of the covariates to the section on inference. Accordingly, to lighten the notations, we omit the X in the rest of this section and the inequalities have to be interpreted as X almost surely. We now propose simple solutions to substantially simplify the characterization of the identified set.

2.3.2 Selection of the Inequalities

The set of inequalities in (2.3.3) characterizes sharply the identified set. Any parameter in the identified set satisfies these inequalities and reversely. However, in order to be implementable, one needs to derive the regions $\mathcal{R}_\omega(A)$ for any set of outcomes $(y_1, \ldots, y_k)$. Though Proposition 2.1 gives closed form expressions when A is a single outcome, calculating the regions $R_\omega(A)$ for subsets A composed by several outcomes requires to know whether the different regions $R_\omega(y)$ for $y \in A$ are multiple equilibria regions and how they overlap with other multiple equilibria regions. In other words, characterizing the set of inequalities requires to know the structure of the multiple equilibria regions. One could bypass this issue by considering outer sets (that is, to exploit only a subset of the inequalities implied by the model). Andrews, Berry, and Jia (2004) compute the outer-set defined by upper bounds on single outcomes while Ciliberto and Tamer (2009) study the outer-set characterized by upper and lower bounds on single outcomes. However, deriving these lower bounds require to know the multiple equilibria structure or to simulate them.

As exhibited previously, the number of inequalities in (2.3.3) can be very large. Even in cases where brute force would be possible, the inference procedure would be challenging, in particular to get competitive critical values. Luckily, a lot of these inequalities are redundant in the sense that they are implied by the knowledge of other inequalities. Notice that it is a different notion than the redundancy used in the GMM literature.[12] To be more specific, we provide a definition to be self-contained.

Definition 2 (Redundancy of a moment inequality) *Let $A \in \mathcal{P}(\mathcal{Y})$, we say that A yields a redundant*

[12] See, for example, Breusch, Qian, Schmidt, and Wyhowski (1999).

inequality if there exist $A_1 \in \mathcal{P}(\mathcal{Y})$ *and* $A_2 \in \mathcal{P}(\mathcal{Y})$ *not empty such that*

$$P_0(Y \in A_i) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(A_i)), i = 1,2 \Rightarrow P_0(Y \in A) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(A)).$$

For instance, in our leading example, $A = \{(1,0),(2,1)\}$ yields a redundant inequality because these two outcomes do not have regions $R_\omega(y)$ that overlap, so

$$R_\omega((1,0)) \cap R_\omega((2,1)) = \emptyset.$$

Therefore if $P_0(Y = (1,0)) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(1,0))$ and $P_0(Y = (2,1)) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(2,1))$,

$$P_0(Y \in A) = P_0(Y = (1,0)) + P_0(Y = (2,1)) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(1,0)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(2,1)) = P_\eta(\varepsilon \in \mathcal{R}_\omega(A)).$$

Here, we propose an algorithm that determines the multiple equilibria structure for a given parameter θ. This algorithm allows us in the same procedure to eliminate the redundant inequalities and to calculate (or bound) the probabilities of the different regions $R_\omega(A),\ A \in \mathcal{P}(\mathcal{Y})$. Before detailing it, we explain how we eliminate redundant inequalities by sufficient conditions. It has been explained in the literature (Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), Chesher and Rosen (2017), Bontemps and Kumar (2020), Luo and Wang (2017)) that eliminating redundant inequalities, or, equivalently, determining a core determining class is linked to the graph structure of the model. The graph $\Gamma(\theta)$ generated by the model is defined as a graph linking the outcomes $y_i \in \mathcal{Y}$ such that there exists an edge between two elements y_1 and y_2 if their equilibrium regions $\mathcal{R}_\omega(y_1)$ and $\mathcal{R}_\omega(y_2)$ overlap. Following the terminology used in graph theory, a subset A is connected in the graph $\Gamma(\theta)$ if and only if there exists a path between every pair of elements in A.

Proposition 3.1 (Sufficient condition for redundancy) *If a subset* $A \subset \mathcal{Y}$ *is not connected in the graph* $\Gamma(\theta)$*, then* A *yields a redundant inequality.*

The proof is simple and therefore omitted.

In our trade-off between feasibility and efficiency, we decide to eliminate as many redundant inequalities as possible to get the smallest set of inequalities to test. Observe, however, that adding redundant inequalities might improve the small sample properties of the estimated identified set (in terms of volume for example). For a given parameter θ, the core determining class is a subset of $\mathcal{P}(\mathcal{Y})$ that exhausts all the inequalities used to define Θ_I in (2.3.3). Let us underline that the core determining class is not unique and can vary in size depending on the conditions that are used to eliminate inequalities. Luo and Wang (2017) provides conditions to find the smallest core determining class in the context of entry games. The novelty in our paper is to provide a simple way to derive a core determining class of inequalities by directly deriving the graph $\Gamma(\theta)$.

2.3.3 Our algorithm to determine a core determining class

In this part, we show how our algorithm allows us to eliminate moment inequalities and to calculate the upper bound of $P(Y \in A)$ for each selected element A of $\mathcal{P}(\mathcal{Y})$. Remark that, in some special case like in Mazzeo (2002) and Cleeren, Verboven, Dekimpe, and Gielens (2010), it is possible to predetermine the graph $\Gamma(\theta)$. However, the restrictions imposed on the two-type model of Cleeren, Verboven, Dekimpe, and Gielens (2010) are difficult to generalize to more types without assuming much stronger restrictions, especially on the horizontal differentiation between the different types. We now state a necessary and sufficient condition for two equilibrium regions $\mathcal{R}_\omega(y_1)$ and $\mathcal{R}_\omega(y_2)$ to overlap.

Proposition 3.2 (Overlapping equilibrium regions) *Two outcomes $y_1 = (N_1, ..., N_T)$ and $y_2 = (\bar{N}_1, ..., \bar{N}_T)$ have their equilibrium regions $\mathcal{R}_\omega(y_1)$ and $\mathcal{R}_\omega(y_2)$ which overlap if and only if $\forall t \in \mathcal{T}$,*[13]

$$\max\left(-\pi_t(N_t, \mathbf{N}_{-t}; \omega), -\pi_t(\bar{N}_t, \bar{\mathbf{N}}_{-t}; \omega)\right) < \min\left(-\pi_t(N_t + 1, \mathbf{N}_{-t}; \omega), -\pi_t(\bar{N}_t + 1, \bar{\mathbf{N}}_{-t}; \omega)\right) \tag{2.3.4}$$

The proof is straightforward given the fact that the regions $\mathcal{R}_\omega(y_1)$ are cubes in $\mathbb{R}^T$. It can be generalized to any set of outcomes.

[13] Remember that, by convention, we have that $\forall t \in \mathcal{T}$, $\pi_t(0, \mathbf{N}_{-t}; \omega) = -\infty$ and $\lim_{N_t \to +\infty} \pi_t(N_t, \mathbf{N}_{-t}; \omega) = +\infty$.

Proposition 3.3 (Characterization of the intersection regions) *For any element A of $\mathcal{P}(\mathcal{Y})$, when it is non-empty, the intersection region $\bigcap_{y_k \in A} \mathcal{R}_\omega(y_k)$ is defined in each dimension t in $\mathcal{T}$ as follows:*

$$\max_{y_k \in A} -\pi_t(y_k;\omega) \leq \varepsilon_t < \min_{y_k \in A} -\pi_t(y_k^+;\omega),$$

with $y_k = (N_{t,k}, \mathbf{N}_{-t,k})$ and $y_k^+ = (N_{t,k}+1, \mathbf{N}_{-t,k})$.

Our algorithm is executed as follows.

Stage 1: Compute the regions $\mathcal{R}_\omega(y)$ for all single outcomes of $\mathcal{Y}$. Collect the moment inequalities generated from (2.3.2).

Stage 2: Check all the pairs ($K = 2$) to see if the two outcomes y_1 and y_2 of each pair have their equilibrium regions $\mathcal{R}_\omega(y_1)$ and $\mathcal{R}_\omega(y_2)$ that overlap using condition (2.3.4). Draw an edge between these outcomes (for the graph $\Gamma(\theta)$) if it is the case and eliminate all moment inequalities generated from pairs for which this is not the case.

For the remaining pairs $A = (y_1, y_2)$, compute the sharp upper bound of $P(Y \in A)$:

$$P(Y \in A) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(y_2)) - P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1) \cap \mathcal{R}_\omega(y_2)).$$

Add these moment inequalities to the set of inequalities generated by the single outcomes (2.3.2).

Stage 3: $K = 3$. We now check all triplets (y_1, y_2, y_3) given that their equilibrium regions might overlap if and only if they overlap two by two. In other words, we focus on the remaining pairs to select our "triplet candidates". Again, if the three regions overlap we have a connected subset of three elements, otherwise we do not keep the triplet and do not consider the moment inequality generated by an eliminated triplet.

For the remaining connected subsets of four elements $A = (y_1, y_2, y_3)$, compute the sharp upper bound of $P(Y \in A)$:

$$
\begin{aligned}
P(Y \in A) \leq\ & P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(y_2)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(y_3)) \\
& - P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1) \cap \mathcal{R}_\omega(y_2)) - P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1) \cap \mathcal{R}_\omega(y_3)) - P_\eta(\varepsilon \in \mathcal{R}_\omega(y_2) \cap \mathcal{R}_\omega(y_3) \\
& + P_\eta(\varepsilon \in \mathcal{R}_\omega(y_1) \cap \mathcal{R}_\omega(y_2) \cap \mathcal{R}_\omega(y_3)).
\end{aligned}
$$

Stage 4: $K = 4$. Check now for all connected subsets of four elements given that each subset of three elements must be in the remaining connected subsets of three elements and so forth

The algorithm stops when there are no longer connected subsets of $K = l + 1$ elements. Remark that it allows us to derive all the components H_i of $\Gamma(\theta)$.[14] For each component, we have a moment equality. Keeping it as an inequality is however sufficient to have a sharp characterization of the set. Given that the region corresponding to the outcome $(0, 0)$ does not overlap with any other region $\mathcal{R}_\omega(y)$, we have at least two components.

The algorithm stops within a finite number of steps, each of them being polynomial in the number of types, given the fact that profit of a type t firm is strictly decreasing in the number of active firms of this type and tends to $-\infty$ when N_t tends to infinity. When a subset $A = (y_1, \ldots, y_K)$ of K elements is selected by the algorithm, the sharp upper bound for the probability of $Y \in A$ can be derived using the expression:

$$
\begin{aligned}
P_\eta(\varepsilon \in \mathcal{R}_\omega(A)) &= P_\eta(\varepsilon \in \bigcup_{y_i \in A} \mathcal{R}_\omega(y_i)) \\
&= \sum_{k=1}^{K} P_\eta(\varepsilon \in \mathcal{R}_\omega(y_k)) + \sum_{i=2}^{K} (-1)^{i-1} \sum_{A_i \subset A\ \&\, \text{card}(A_i) = i} P_\eta(\varepsilon \in \bigcap_{y_k \in A_i} \mathcal{R}_\omega(y_k)).
\end{aligned}
$$

If this subset is a component of the graph $\Gamma(\theta)$, it generates an equality. An important remark here is that these intersection regions are also T-cubes and thus integrating over these regions is straightforward.

[14] The components of a graph are subgraphs $\{H_i\}_{i=1}^k$ such that each H_i is connected and H_i is not connected to H_j for $i \neq j$.

Application to our example Figure 2.3 displays the resulting graph for our example. In this case there are no connected subsets of three elements. The identified set can therefore be characterized sharply by 14 inequalities related to the single outcomes which are not $(0,0)$ and $(3,3)$, 9 inequalities generated by the selected pairs and 7 equalities generated by the 7 components (including $(0,0)$ and $(3,3)$).

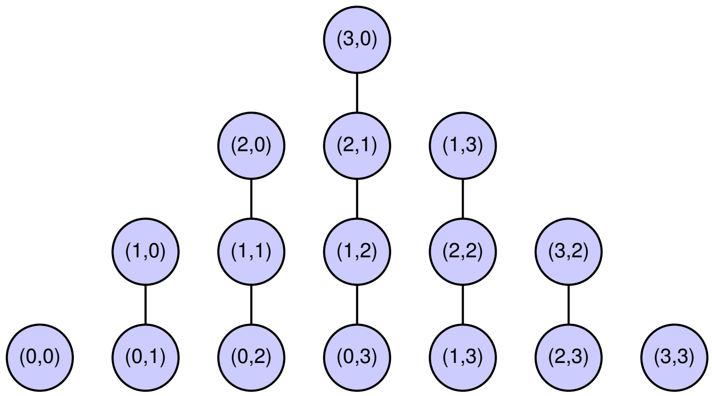


Figure 2.3: $\Gamma(\theta)$ for $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$

Remarks on the algorithm

- The cost to check whether a subset of p elements is connected is low because it is just comparing the maximum of p quantities with the minimum of p other quantities following Proposition 3.3.
- The collection of all connected subsets constitutes a core determining class $\mathcal{C}(\theta)$, i.e. the core determinig class generates moment inequalities that sharply characterizes the identified set Θ_I. However, it might not be the smallest class (see the Monte Carlo section).
- One advantage of our sequential algorithm is that it allows the researcher to stop the procedure after a few iterations. This can prove useful in practice if the set of connected subsets is too large and it also allows to measure the effect on the identified set of going one iteration further in the collection of connected subsets.

- If one wants to compute the minimum probability of any subset, one needs to execute the algorithm until the end to determine all the possible multiplicities. This is the reason why Ciliberto and Tamer (2009) derives the lower bound for each probability of a single outcome using simulation methods. However, simulating these bounds induces simulation noise that is often (wrongly) ignored in the inference procedure.

2.4 Inference on the full vector

Given the knowledge of $P_0(Y = y|X)$ (for simplicity, in this section, we denote this vector $P_0(X)$), the last section allows us to test that parameter θ is in the identified set Θ_I by exploiting the conditional moment inequalities generated by a subset of $\mathcal{P}(\mathcal{Y})$, called $\mathcal{C}(\theta, X)$. In general, the core determining class that we exhibited previously depends on θ and on X. Calling $p_{\theta,X}$ the number of elements in $\mathcal{C}(\theta, X)$, the moment inequality generated by any member A_j of $\mathcal{C}(\theta, X)$ can be rewritten in a simple form:

$$q_j^\top P_0(X) \leq C_{\theta,j}(X),$$

in which q_j is a $\mathbf{card}(\mathcal{Y}) \times 1$ vector of zeros and ones, the ones corresponding to the outcomes present in A_j and $C_{\theta,j}(X) = P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_j))$. Calling $D_{\theta,X}$ the stacked version of the $q_j^\top$s and $C_\theta(X)$ the stacked version of the $C_{\theta,j}(X)$s, we get:

$$\begin{aligned} \theta \in \Theta_I &\iff D_{\theta,X} P_0(X) \leq C_\theta(X),\ X\ a.s. \\ &\iff \mathbb{E}\left(C_{\theta,j}(X) - q_j^\top \mathbf{1}(Y = y)|X\right) \geq 0, \forall\, j = 1, \ldots, p_{\theta,X}\ X\ a.s. \end{aligned} \tag{2.4.5}$$

We have a collection of conditional moment inequalities $m_j(X, \theta)$ which are linear in $\mathbb{E}(\mathbf{1}(Y = y)|X)$. Moreover, for a given $X = x$, the condition $D_{\theta,x} P \leq C_\theta(x)$ in (2.4.5) defines a convex set $A(\theta, x)$ in which $P_0(X = x)$ should lie for θ to be in the identified set.

When we turn to inference, $P_0(X)$ is no longer observed but must be estimated in a first stage. The objective of this section is to provide an inferential procedure that is asymptotically valid (each point in

Θ_I should be in the confidence region with a probability that is asymptotically greater than one minus the nominal level of the test) and consistent (each point outside θ_I should be rejected with probability that goes to one). In contrast to the classical setup in which the true parameter is uniquely defined by conditional moment equalities, here we must repeat a test over a grid of candidates θ. Therefore, it is critical for the testing procedure to remain simple to implement while being sufficiently powerful to avoid estimating unnecessarily large confidence regions. In this section, we propose a new estimation method that seems to satisfy these requirements. Furthermore, explanatory variables create additional issues when they are taken into account (curse of dimensionality). In most empirical applications, researchers have suggested to discretize the covariates at the cost of changing the original model. We will show that one key advantage of our method is that it can smoothly handle the introduction of covariates with a slight correction of the test statistic.

In the following, we assume we observe an i.i.d. sample of n outcomes $(X_1, Y_1), \ldots, (X_n, Y_n)$ in independent markets. We first study the inference procedure in the absence of covariates and we show how to adapt the procedures to include covariates.

2.4.1 Inference without covariates

First, we focus on the case where there are no covariates. In the case, the identified set Θ_I is characterized by the following collection of moment inequalities:

$$\theta \in \Theta_I \iff \mathbb{E}\left(C_{\theta,j} - q_j^\top \mathbf{1}(Y = y)\right) \geq 0, \ \forall j = 1, \ldots, p_\theta \tag{2.4.6}$$

Let P_n be the empirical frequency vector of outcomes:

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(Y_i = y),$$

in which the inequality $\mathbf{1}(Y_i = y)$ should be interpreted term by term, i.e.,

$$\mathbf{1}(Y_i = y) = \left[\mathbf{1}(Y_i = y_1), \mathbf{1}(Y_i = y_2), \ldots, \mathbf{1}(Y_i = y_{card(\mathcal{Y})})\right]^\top.$$

Our empirical analogue of the moment inequalities is:

$$m_{n,j}(Y,\theta) = \frac{1}{n}\sum_{i=1}^{n} m_j(Y_i,\theta) = C_{\theta,j} - q_j^\top P_n \quad j = 1,\ldots,p_\theta$$

which should be the estimates of positive quantities. Also, we denote $\Sigma_n = diag(P_n) - P_n^\top P_i$, i.e., a consistent estimator of $Var(\mathbf{1}(Y = y))$.

Existing procedures

Various test statistics have been proposed in the literature to test moment inequality restrictions similar to the ones in Equation (2.4.6). Andrews and Soares (2010) proposes different test statistics and the GMS procedure to calculate accurate critical values. Andrews and Barwick (2012) study the performance of these different values and provide guidance about the tuning parameters involved in the procedure. In a more recent work, Chernozhukov, Chetverikov, and Kato (2018b) proposes a test statistic easy to compute with a critical value which is valid whatever the correlation structure of the moments involved.[15]. Their approach is particularly suited for cases such as games, which display a very large number of inequalities. Contrary to alternatives such as subsampling or general moment selection, the critical value, which is based on a moderate deviation inequality for self-normalized sums, is straightforward to compute and increases (in absolute value) slowly in the number of moments. This property is particularly attractive as we need to repeat the testing procedure for each point in the grid but still be competitive.

Minimum test statistic To be more specific, let $\xi_n(\theta)$ be defined as the minimum over the studentized moments:

$$\xi_n(\theta) = \min_{j=1,\ldots,p_\theta} \frac{\sqrt{n}(C_{\theta,j} - q_j^\top P_n)}{\sqrt{q_j^\top \Sigma_n q_j}} = \min_{j=1,\ldots,p_\theta} \frac{\sqrt{n} m_{n,j}(Y,\theta)}{\sqrt{V_{n,j}}},$$

[15] For papers applying this procedure, see for instance Bontemps and Kumar (2020),Chesher and Rosen (2019)

When θ does not belong to the identified set, the quantity above should diverge to $-\infty$. When $\theta \in \Theta_I$, the asymptotic distribution of $\xi_n(\theta)$ can be derived [16] and it is equal to

$$\min_{j \in \mathcal{J}(\theta)} \frac{q_j^\top Z}{\sqrt{q_j^\top \Sigma_0 q_j}},$$

in which Z follows a normal distribution with variance Σ_0 and $\mathcal{J}(\theta)$ is the collection of indices j corresponding to the binding moments. This asymptotic distribution depends on the number and the identity of the binding moments, as expected. In the following, p^* denotes the number of binding moments, i.e., the cardinal of $\mathcal{J}(\theta)$.

A critical value can be computed after a first step estimation of the set of binding moments $\mathcal{J}(\theta)$ like in the GMS procedure of Andrews and Soares (2010). Simulation methods (bootstrap and/or subsampling techniques) can be also considered to improve the accuracy of the critical value. Chernozhukov, Chetverikov, and Kato (2018b) proposes the following one:

$$c^*(\alpha) = \frac{\Phi^{-1}(\alpha/p)}{\sqrt{1-\Phi^{-1}(\alpha/p)^2/n}} \tag{2.4.7}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution ($\Phi^{-1}(\cdot)$ is its inverse). The advantage of this critical value is that it is easy to compute, quite competitive and it decreases at a rate of the order $-\sqrt{\log(p/\alpha)}$, i.e., does not diverge too quickly when the number of moments is high. Under some mild regularity assumptions, satisfied in our framework, Chernozhukov, Chetverikov, and Kato (2018b) shows that the confidence set $\mathrm{CR}_n(1-\alpha)$ induced by $c^*(\alpha)$ is asymptotically valid,

$$\mathrm{CR}_n(1-\alpha) = \{\theta \in \Theta \mid \xi_n(\theta) \geq c^*(\alpha)\}.$$

Proposition 4.1 *The confidence set generated by $c^*(\alpha)$ is asymptotically valid.*

$$\liminf_{n\to\infty} \inf_{\theta \in \Theta_I} \Pr(\theta \in CR_n) \geq 1-\alpha.$$

Proof: See Theorem 4.1. of Chernozhukov, Chetverikov, and Kato (2018b). In our framework, all moment inequalities are trivially bounded. ■

[16] See Bontemps and Kumar (2020), Proposition 9.

Applying methods in the convex set theory. Alternatively, one can exploit the equivalence between checking the moment inequalities and testing that P_0 belongs to the convex set $A(\theta)$. Convexity is an attractive feature that has been exploited in the set identification literature by Beresteanu, Molchanov, and Molinari (2011) or by Bontemps and Kumar (2020) for games with multiple equilibria. In particular, the authors use the support function

$$\delta^*(q; A(\theta)) = \sup_{P \in A(\theta)} q^\top P.$$

Following Rockafellar (1997),

$$P_0 \in A(\theta) \Longleftrightarrow \min_q \delta^*(q; A(\theta)) - q^\top P_0 \geq 0.$$

In other words, the support function embeds all the moment inequalities.

Classical reformulation of the problem above shows that the program is strictly equivalent to testing that the (euclidian) distance between P_0 and $A(\theta)$ is equal to 0. Therefore, we can also consider test statistic based on generalized distance:

$$d_\Omega(P_0, A(\theta)) = \min_{P, MP \leq C_\theta} (P_0 - P)^\top \Omega^{-1} (P_0 - P). \tag{2.4.8}$$

As $A(\theta)$ is a convex set, the distance can be easily computed from quadratic solvers under linear constraints. However, the asymptotic distribution of $d_\Omega(P_n, A(\theta))$ is still a complicated distribution because it depends on whether the true probability vector P_0 is inside the set or lies on an exposed face or an edge of the convex set. One can work with conservative critical values that are easy to compute like the critical value proposed by Cox and Shi (2022) for a conditional version of the test or use the old literature on inequality testing and the upper bound proposed by some of the authors (see **?** for a review of the existing procedures). Kitamura and Stoye (2018) propose simulation methods that are valid in our framework.

The smoothed-min approach

As illustrated above, the difficulty in the traditional moment inequality approach is to recover the exact asymptotic distribution of the test statistic, which depends on the set of binding moments. In the case of games, the set of binding moments depends on whether the true vector P_0 lies inside, on an exposed face of a given order or is a vertex. While computationally intensive methods [17] often display better approximation properties of the exact asymptotic distribution, the implementation difficulties make these methods unappealing for the estimation of games. On the other hand, the usage of upper bounds on the asymptotic or exact distribution of the test statistic results in conservative confidence regions.

Here, we propose a new statistical procedure where the asymptotic distribution of the test statistic is a standard normal. The general idea is to estimate a smooth outer set of Θ_I that asymptotically converges to the true identified set Θ_I. The advantage of manipulating a smooth outerset is that we can replace a set of moment inequalities by one moment binding moment inequality for which we know the asymptotic distribution. This approach completely removes the inferential loss caused by the inability to derive the exact asymptotic distribution by introducing a small and manageable identification loss through smoothing, which we make vanish asymptotically. The resulting confidence regions have the right size asymptotically. Last but not least, we will see in the next section how this smoothing procedure facilitates the introduction of covariates in the model.

The Boltzmann operator Let us now introduce our new estimation strategy. An important remark is that this strategy is valid for the vast majority of models defined by moment inequalities (modulo some mild regularity assumptions that we make explicit) beyond the context of entry games that we study in this paper. From what precedes, the sharp identified is such that:

$$\begin{aligned} \theta \in \Theta_I &\iff \min_{j=1,\ldots,p_\theta} \mathbb{E}m_j(Y,\theta) \geq 0. \\ &\iff \min\{0, \min_{j=1,\ldots,p_\theta} \mathbb{E}m_j(Y,\theta)\} = 0 \end{aligned} \tag{2.4.9}$$

The second equivalence transforms p moment inequalities into one equality. However, we cannot exploit directly an empirical counterpart of $\min\{0, \min_{j=1,\ldots,p_\theta} \mathbb{E}m_j(Y,\theta)\}$ because deriving the asymptotic

[17] Even if they don't eliminate the inferential loss.

distribution of such a statistic is difficult. What we propose is to replace the minimum by a smooth approximation, what allows us to recover asymptotic normality through a standard Taylor expansion. For $z = (z_1, z_2, ..., z_p) \in \mathbb{R}^p$, we have that a smooth approximation of the minimum between the elements of z and 0 writes:

$$g_\rho(z) = \frac{\sum_{j=1}^p z_j \exp(-\rho z_j)}{1 + \sum_{j=1}^p \exp(-\rho z_j)},$$

in which ρ, the smoothing parameter, controls the level of approximation. $g_\rho(\cdot)$ is known as the Boltzmann operator. This function is also used in machine-learning and numerical optimization. In the appendix, we show that other smooth approximations of the minimum such as the LogSumExp could also be used. A nice property of this approximation is that it is possible to control for the difference between the minimum and its approximation. Following Chernozhukov, Kocatulum, and Menzel (2015), we have:

$$\left| min(0, z_1, z_2, \ldots, z_p) - g_\rho(z) \right| \leq \frac{1}{\rho} W\left(\frac{p-1}{e}\right),$$

where $W(\cdot)$ is the Lambert function. In particular, $W(x)$ is bounded by $\log x$ when $x > e$. The minimum is simply the limit of $g_\rho(z)$ when $\rho \to +\infty$. For simplicity, in what follows, we define $m_\theta \equiv \mathbb{E}[m(Y, \theta)]$ and we omit the dependence of p in θ.

A smooth outer set. For exposition, it is useful to first study the effect of replacing the minimum function by a smooth approximation. We define an outer set $\Theta_I^o(\rho)$ as follows:

$$\Theta_I^o(\rho) = \left\{ \theta \in \mathbb{R}^{\dim(\theta)} | \; g_\rho(m_\theta) = \frac{\sum_{j=1}^p m_{\theta,j} e^{-\rho m_{\theta,j}}}{1 + \sum_{j=1}^p e^{-\rho m_{\theta,j}}} \geq 0 \right\}. \tag{2.4.10}$$

The next proposition shows that $\Theta_I^o(\rho)$ contains the true identified set Θ_I.

Proposition 4.2 *The following statements hold*

(i) For any $\rho > 0$, $\Theta_I \subset \Theta_I^o(\rho)$

(ii) $\lim_{\rho \to +\infty} d_H(\Theta_I, \Theta_I^o(\rho)) = 0$, *where* d_H *is the Hausdorff distance used in set theory.*

See the proof in the Appendix. These properties indicate that the smoothing mechanically induces an identification loss. If we were to use an empirical counterpart of $g_\rho(m_\theta)$ while holding ρ fixed, we would estimate an outer set of Θ. Now the idea is to decrease the level of smoothing at a certain speed in order to estimate the true identified set while keeping asymptotic normality.

A pivotal test statistic. We define our smooth test statistic as follows:

$$\xi_n(\theta) = \sqrt{n}\frac{g_{\rho_n}(m_{\theta,n})}{\sqrt{\nabla g_{\rho_n}(m_{\theta,n})^T \Sigma_n \nabla g_{\rho_n}(m_{\theta,n})}}$$

with $\nabla g_{\rho_n}(\cdot)$ the gradient of $g_{\rho_n}(\cdot)$ that we define in the appendix and Σ_n a consistent estimator of Σ_0. Our confidence region of confidence level $1-\alpha$ is defined as follows:

$$CR_n(1-\alpha) = \{\xi_n(\theta) \geq z_\alpha\}$$

in which z_α is the α-quantile of the standard normal distribution.

In order to show asymptotic validity of our test statistic, we are going to further assume that the moments we consider are asymptotically normal.

Assumption 4 (Asymptotic normality of the moments) *Let us assume that:*

$$\sqrt{n}(m_{\theta,n} - m_\theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$$

with Σ_0 *definite positive.*

This last assumption is almost always satisfied from an application of the usual CLT. In the case of games, Assumption 4 is satisfied:

$$\sqrt{n}(m_{\theta,n} - m_\theta) = Q_\theta^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(Y_i = y) - \mathbb{E}(\mathbf{1}\{Y = y\}) \right) \xrightarrow{d} Q_\theta^T \mathcal{N}(0, \Sigma_0)$$

with $\Sigma_0 = diag(P_0) - P_0 P_0^\top$.

The next proposition describes the asymptotic behavior of the test statistic.

Proposition 4.3 *Let ρ_n a divergent sequence of positive numbers such that $\rho_n = O(n^{\alpha})$, $0 < a < 1/2$, then there are 3 different cases:*

- $\theta \in int(\Theta_I)$ *($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0):*

$$Pr(\xi_n(\theta) > z_\alpha) \underset{n\to+\infty}{\to} 1$$

- $\theta \in \partial\Theta_I$ *($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0). Then our test statistic is asymptotically distributed as a standard normal:*

$$\xi_n(\theta) \xrightarrow{d} \mathcal{N}(0,1)$$

- $\theta \notin \Theta_I$: $\qquad Pr(\xi_n(\theta) > z_\alpha) \underset{n\to+\infty}{\to} 0$

The proof of this proposition is in the appendix. The Proposition shows that our procedure is consistent: the econometrician rejects with probability that goes to 1 when $\theta \notin \Theta_I$. Second, it has asymptotically the exact size and it is therefore not conservative when at least one moment is binding. Finally, in constrast with many other procedures, the test statistic and the critical value are straightforward to derive. Proposition 4.3 builds on the following asymptotic expansion of the smooth test statistic.

Proposition 4.4 (Asymptotic expansion of the test statistic) *Let ρ_n a divergent sequence of positive number such that $\rho_n = O(n^{\alpha})$, $0 < a < 1/2$, then*

$$\sqrt{n}g_{\rho_n}(m_{\theta,n}) = \sqrt{n}g_{\rho_n}(m_\theta) + \Gamma_0(\theta)\sqrt{n}(m_{\theta,n} - m_\theta) + o_p(1)$$

with $\Gamma_0(\theta) \equiv \lim_{\rho\to\infty} \nabla g_\rho(m_\theta)$.

The proof is in the appendix. The constraint on the speed of ρ implies that the smoothing cannot be faster than the standard square-root speed of convergence.

Standardization of the moments ex-ante. As an alternative to $\xi_n(\theta)$ we can consider a modified test statistic in which we standardize the moment m_θ by its variance. Namely, let $\tilde{m}_{\theta,n} = diag(\Sigma_n)^{-1/2} m_{\theta,n}$. The new test statistic writes:

$$\tilde{\xi}_n(\theta) = \sqrt{n} \frac{g_{\rho_n}(\tilde{m}_{\theta,n})}{\nabla g_{\rho_n}(\tilde{m}_{\theta,n})^T \Omega_n \nabla g_{\rho_n}(\tilde{m}_{\theta,n})}$$

with Ω_n a consistent estimator of the correlation matrix $\Omega_0 = diag(\Sigma_0)^{-1/2} \Sigma_0 diag(\Sigma_0)^{-1/2}$. In the appendix, we show that $\tilde{\xi}_n(\theta)$ has the same asymptotic behavior as $\xi_n(\theta)$. The advantage of $\tilde{\xi}_n(\theta)$ over $\xi_n(\theta)$ is to have a procedure that is robust to the unit of measurement of all the moments $m_{\theta,j}$.

Choice of ρ_n. Let us now discuss the choice of the smoothing parameter ρ_n, which is a critical choice in practice. In order for our strategy to work, we require the smoothing parameter ρ_n to diverge at a certain speed, namely, we want $\rho_n = O(n^\alpha)$ with $0 < \alpha < 1/2$. Remember that if we keep ρ fixed, $g_\rho(m_{\theta,n})$ would a consistent estimator of $g_\rho(m_\theta)$ and not $min\{0, m_{\theta,1}, ..., m_{\theta,p}\}$ anymore. In turn, this implies that we are estimating consistently the outer set $\Theta_I^o(\rho)$ defined in equation (2.4.10), whereas our goal is to estimate consistently the sharp identified set Θ_I. On the other hand, we cannot let the amount of smoothing decrease too fast (here we require ρ_n to increase at a lower speed than the parametric convergence rate) otherwise we lose the asymptotic normality of our estimator. Intuitively, when ρ_n increases too rapidly, g_{ρ_n} gets closer to the minimum function and thus the higher order terms in the Taylor expansion don't vanish asymptotically. The bounds we provide on the speed of divergence ρ_n are not very informative on the choice of ρ_n in practice. Our goal is now to provide more precise guidelines on ρ_n.

Interestingly, these two forces appear in the mean value Taylor expansion exhibited in the proof of Proposition 4.4. $\forall n \in \mathbb{N}^*$, $\exists \tilde{m}_{\theta,n} \in [m_{\theta,n}, m_\theta]$ such that the following expansion holds:

$$\sqrt{n} g_{\rho_n}(m_{\theta,n}) = \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{\text{Identification bias}} + \underbrace{\nabla g_{\rho_n}(m_\theta) \sqrt{n}(m_{\theta,n} - m_0)}_{\text{First order approximation}} + \underbrace{\frac{\rho_n}{\sqrt{n}} \sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n} \sqrt{n}(m_{\theta,n} - m_\theta)}_{\text{Rest in Taylor's expansion}}$$

The smooth test statistics consists of three components. The identification bias term corresponds to the loss implied by the use of a smooth approximation of the minimum instead of the minimum directly (to be precise, the identification bias is equal to $\sqrt{n}g_{\rho_n}(m_\theta) - \min_j m_{\theta,j}$ with the second term equal to 0 under $H_0 : \theta \in \Theta_I$). The first-order approximation is the component that is normally distributed and from which we derive the asymptotic distribution of the test statistic. The rest of the Taylor expansion corresponds to the discrepancy between the first-order approximation and the smoothing function. It can be interpreted as a measure of the non-normality of the estimator. The approach that we propose to chose ρ_n consists in quantifying the biases induced by the two opposite forces in finite sample and chose ρ_n to minimize the bias implied by these two forces. One can easily check that the bias of our estimator is equal to:

$$\mathbb{E}\left[\sqrt{n}g_{\rho_n}(m_{\theta,n})\right] = \underbrace{\sqrt{n}g_{\rho_n}(m_\theta)}_{\text{Identification bias}} + \underbrace{\frac{\rho_n}{\sqrt{n}}\mathbb{E}\left[\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n} - m_\theta)\right]}_{\text{Bias implied by the rest in Taylor's expansion}}$$

The next proposition provides an upper bound on the small sample bias on the absolute value of these two terms and a choice of ρ_n to minimize this upper bound.

Proposition 4.5 (Choice of ρ_n)

$$|\mathbb{E}\left[\sqrt{n}g_{\rho_n}(m_{\theta,n})\right]| \leq (p - J_0)\frac{\frac{1}{\rho_n}e^{-1}}{1 + J_0} + \frac{\rho_n}{\sqrt{n}}K_0$$

with $K_0 > 0$ a constant that we make explicit in the appendix and that increases with the number of binding moments and the variance of the moments. Thus, the choice ρ_n^* that minimizes this upper bound is equal to:

$$\rho_n^* = n^{1/4}\sqrt{\frac{(p - J_0)e^{-1}}{(1 + J_0)K_0}}$$

We see that the "optimal" choice of ρ_n increases with the number of non-binding moments and decreases with the number of binding moments and the variance of these moments. Let us observe that the optimal speed of divergence $\alpha^* = \frac{1}{4}$ is also contained in $(0, 1/2)$

2.4.2 Inference with covariates

We now move to the case with covariates. Following Equation 2.4.5, in the presence of covariates, the identified is characterized by a collection of moment inequalities.

$$\begin{aligned} \theta \in \Theta_I &\Longleftrightarrow \mathbb{E}\left(C_{\theta,j}(X) - q_j^\top \mathbf{1}(Y = y)|X\right) \geq 0, \forall\, j = 1, \ldots, p_{\theta,X} \; X\; a.s. \\ &\Longleftrightarrow m_j(X,\theta) \equiv \mathbb{E}\left(m_j(Y, X, \theta)|X\right) \geq 0, \forall\, j = 1, \ldots, p_{\theta,X} \; X\; a.s. \end{aligned}$$

Θ_I is defined by a collection of conditional moment inequalities $m_j(X, \theta)$ that are linear in $\mathbb{E}(\mathbf{1}(Y = y)|X)$. The presence of covariates poses several theoretical and practical issues. First, the algorithm to select the inequalities and the approach to compute the theoretical bounds developed in section 2.3.3 remain the same but have to be performed for each element $x \in \mathcal{X}$ (and for each θ), which can significantly increase the computational burden especially if the dimension of X is large and/or if X is continuous. More fundamentally, if X is continuous, the set is sharp identified set is characterized by an infinite number of inequalities. Second, conditional moments are non-parametric objects that are harder to estimate than unconditional moments and which display non-standard asymptotic properties. In particular, we lose the asymptotic normality and the parametric rate of convergence of the estimator of $m_j(X, \theta)$. We now propose different methods to alleviate the two difficulties mentioned above and more generally to facilitate the estimation of models defined by conditional moment inequalities.

First, we tackle the problem of deriving the core determining class for each $x \in \mathcal{X}$ as we want to avoid repeating the algorithm introduced in section 2.3.3 for each $x \in \mathcal{X}$ (we already must do it for each θ). To keep things tractable numerically, we suggest to make the following separability assumption for our profit functions:

Assumption 5 (Additively separable profit shifters)

$$\pi_t(X, N_t, \mathbf{N}_{-t}; \omega) = \kappa_t(X; \omega_1) + \phi_t(N_t, \mathbf{N}_{-t}; \omega_2), \quad \forall t = 1, \ldots, T.$$

The next proposition states that, with the profit structure assumed in Assumption 5, the core determining class remains the same for all $x \in \mathcal{X}$.

Proposition 4.6 (Invariant core determining class) *Under Assumptions 2, 3, 5, we have:*

$$\forall x \in \mathcal{X}, \quad \mathcal{C}(\theta, x) = \mathcal{C}(\theta).$$

It implies that under assumption 5, the core determining class only needs to be computed once for each candidate θ (as opposed to deriving the core determining for every combination of x and θ). The intuition for this result is that under additive separability of the covariates in the profit functions, the covariates only translate the equilibrium structure in the space of unobserved shocks and the graph $\Gamma(\theta)$ remains the same.

From now on, we assume Assumption 5 and we denote the core determining class $\mathcal{C}(\theta)$ since it does not depend on X. p_θ is its cardinal. In what follows, we omit the dependence of p in θ.

Existing procedure

Following (2.4.5), Θ_I is characterized by p conditional moment inequalities that should be positive almost surely in X. As we established previously, conditional moment inequalities are much more difficult to tackle than unconditional ones. Various methods have been proposed to estimate confidence sets in models defined by conditional moment inequalities (Andrews and Shi (2013), Armstrong and Chan (2016), Armstrong (2014) among others) and theoretical econometricians have suggested to directly apply these methods to the case of entry games.

The leading method (proposed in Andrews and Shi (2013)) consists in transforming the conditional moment inequalities into unconditional ones. More precisely, Andrews and Shi (2013) considers a collection $\mathcal{G}$ of non-negative functions of X, denoted $g(X)$.

$$\theta \in \Theta_I \implies m_{j,g}(X,\theta) = \Big(C_{\theta,j}(X) - q_j^\top \mathbb{E}(\mathbf{1}(Y=y)|X)\Big)\, g(X) \geq 0 \; a.s.\,, j = 1, ...p, \; \forall g \in \mathcal{G}$$

These functions allow us to transform the conditional moment inequalities into unconditional ones as follows:

$$\bar{\Theta}_I = \{\theta \in \Theta \mid \mathbb{E}\left[m_{j,g}(Y,X,\theta)\right] \geq 0, \; \forall j \in \{1,\ldots,p\}, \; \forall g \in \mathcal{G}\}.$$

Under high level conditions on $\mathcal{G}$, the outer set defined above coincides with the sharp identified set. The choice of $\mathcal{G}$ is critical: as an example, they suggest to use a countable family of hypercubes. For the estimation, they integrate these unconditional moments into either a Cramer Von Mises (CvM) or a Kolmogorov Smirnov (KS) type of statistic. Finally, they adapt the GMS procedure to derive the critical value.

The inference strategy in Andrews and Shi (2013) is extremely challenging numerically. The econometrician must first choose the family $\mathcal{G}$. For each $g \in \mathcal{G}$, the econometrician must compute a test statisitc estimate the set of binding moments following the GMS procedure. Next, the econometrician is required to integrate the family of moments over a certain measure μ and to simulate the asymptotic distribution under the null. Moreover, this procedure involves many tuning parameters (choice of integration functions, choice of the measure, choice of the test statistic, choice of the parameter to select the binding moments ...). To our knowledge, a few papers only used it for real empirical applications.

Researchers often favor an approach based on discretizing the support of continuous variables (even if it results in a modification of the initial model and thus of the identified set) or picking particular $g(\cdot)$ functions in $\mathcal{G}$ above and exploiting the critical value in Chernozhukov, Kocatulum, and Menzel (2015). For example, Aradillas-Lopez and Rosen (2022) use the density of X as weighting function.

The smooth-min approach

Let us now introduce our new estimation strategy. An important remark is that this strategy is valid for the vast majority of models defined by moment inequalities (modulo some mild regularity assumptions that we make explicit) beyond the context of entry games that we study in this paper. We build our approach on the following characterization of the sharp identified set:

Proposition 4.7

$$\begin{aligned}\theta \in \Theta_I &\iff m_j(X,\theta) \geq 0, \;\; \forall \;\; j = 1,\ldots,p \; X \; a.s \\ &\iff \min\{0, \min_{j=1,\ldots,p} m_j(X,\theta)\} = 0 \; X \; a.s. \\ &\iff \mathbb{E}\left[\min\{0, \min_{j=1,\ldots,p} m_j(X,\theta)\}\right] = 0 \\ &\iff \mathbb{E}\left[\min\{0, \min_{j=1,\ldots,p} m_j(X,\theta)\} g(X)\right] = 0\end{aligned}$$

for any choice of weighting function $g(\cdot)$ that is positive, smooth, and that does not vanish on the support of X

The characterization of Θ_I in Proposition 4.7 is extremly useful as it transforms p conditional moment inequalities into one unconditional moment equality without losing any identification power.[18] If we knew $m_j(X,\theta)$, we could directly use a CLT and use a standard one-sided t-test. However, $m_j(X,\theta) = \mathbb{E}\left(m_j(Y,X,\theta)|X\right)$ is unknown and must be estimated non-parametrically in a first stage. Given that we must take into account the first stage estimation in the derivation of the asymptotic distribution, it becomes convenient to once again use a smooth approximation of the minimum function. Again, our smooth-min approach allows us to recover an asymptotically normal estimator for $\mathbb{E}\left[\min\{0, \min_{j=1,\ldots,p} m_j(X,\theta)\}\right]$. We will see that our test will yield a consistent test under mild conditions that is easy to implement. Our testing approach is very reminiscent of Zheng (1996) which

[18] In particular, there is no need to consider a countable family of weighting function. In practice, we choose $g = 1$ but we could choose g equal to the density of X

proposes a consistent specification test on conditional moment equalities by also taking the expectation of the conditional expectation of the residuals.

In what follows, we denote $m_\theta(X)$ the true conditional moment inequality and $\hat{m}_\theta(X)$ a consistent non-parametric estimator. Moreover, let $W = (Y, X)$ and we denote $m(W, \theta)$ the moment function. An important simplification that occurs in practice is that, in many cases, the conditional moment is a function of $\mathbb{E}(Y|X)$, which implies that the econometrician only needs to estimate $\mathbb{E}(Y|X)$ once for every candidate θ. For instance, this simplification occurs in the case of the entry game we study in this paper:

$$m_{\theta,j}(X) = C_{\theta,j}(X) - q_j^\top \mathbb{E}(\mathbf{1}(Y = y)|X) \quad \forall j = 1, ..., p$$

A smooth outer set. As previously, we consider the Boltzmann operator as our smoothing operator. Before defining our formal test, it is useful to first study the effect of replacing the min by a smooth approximation on the identified set. We define an outer set $\Theta_I^o(\rho)$ as follows:

$$\Theta_I^o(\rho) = \{\theta \in \mathbb{R}^{dim(\theta)} \, |\mathbb{E}[g_\rho(m_\theta(X))] \geq 0\}$$

It is straightforward to show that $\Theta_I^o(\rho)$ contains the true identified set.

$$\theta \in \Theta_I \implies g_\rho(m_\theta(X)) \geq 0, \;\; X\text{a.s} \implies \mathbb{E}\left[g_\rho(m_\theta(X))\right] \geq 0.$$

A pivotal test statistic. We can now define our smooth test statistic as follows.

$$\xi_n(\theta) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n g_{\rho_n}(\hat{m}_\theta(X_i))}{\sqrt{V_n}}$$

with V_n a consistent estimator of the variance of $g_{\rho_n}(\hat{m}_\theta(X_i))$ that we will make explicit later on. As for the case without covariates, we let the smoothing decrease with n. The confidence region of level $1 - \alpha$ is simply define as follows.:

$$CR_n(1-\alpha) = \{\xi_n(\theta) \geq z_\alpha\}$$

in which z_α is the α-quantile of the standard normal distribution. Therefore, in contrast with most of the existing methods, here the critical value is straightforward to derive, which is a key consideration in practice.

If the general idea is the same as in the case without covariates, the main difficulty here comes from the fact that we plugg-in a non-parametric estimator into the test statistic (instead of a parametric estimator). Here, the non-parametric rate affects the derivation of the asymptotic distribution. This type of problem has been extensively studied in the literature on semi-parametric estimation (Newey (1994), Ai and Chen (2003), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018a), Ackerberg, Chen, Hahn, and Liao (2014)) and a key regularity condition to recover a square-root asymptotic distribution is for g_{ρ_n} to be "differentiable" with respect to the first stage non-parametric estimator. We now derive the asymptotic distribution of $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n g_{\rho_n}(\hat{m}_\theta(X_i)) - \mathbb{E}\left[g_{\rho_n}(m_\theta(X))\right]\right)$. In particular, we exploit some existing results in the literature on semi-parametric estimation (Ackerberg, Chen, Hahn, and Liao (2014) and Newey (1994)). Before we state the result, we list some regularity conditions on the non-parametric estimator that we require to derive the asymptotic distribution.

In this paper, we consider a kernel based non-parametric estimator for $m_\theta(\cdot)$ and we make following assumptions on the smoothness of kernel as well as the density and the moment functions. We could have considered other non-parametric estimators as well (for instance a series estimator). We now state additional regularity assumptions on the kernel estimator that allows us to prove mean square differentiability and stochastic equicontinuity (see the appendix) that are sufficient conditions to recover asymptotic normality of our estimator.

Assumption 6 *$Supp(X) = \mathcal{X}$ is compact and $f_0(\cdot)$, the p.d.f. of X is bounded away from zero as well as bounded above.*

1. *The kernel $K(\cdot)$ is differentiable of order β with bounded derivatives, $K(\cdot)$ is zero outside the bounded set, $\int K(u)du = 1$ and there is positive integer m such that for $j < m$,*

$$\int u^j K(u)du = 0.$$

2. *The density $f_0(\cdot)$ and regression function $m_\theta(\cdot)$ both are continuously differentiable of order d with bounded derivatives in an open set containing $\mathcal{X}$.*

Under assumptions 6, we are able to derive the following linear expansion.

Proposition 4.8 (Asymptotic expansion with covariates) *Let $W_i = (Y_i, X_i)$. Under assumption 6, there exists $\tilde{m}_\theta(\cdot)$, such that*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g_{\rho_n}(\hat{m}_\theta(X_i) - \mathbb{E}[g_{\rho_n}(m_\theta(X_i))]\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g_{\rho_n}(m_\theta(X_i)) + \alpha(W_i) - \mathbb{E}[g_{\rho_n}(m_\theta(X_i))]\right)$$

$$+ \frac{\rho}{\sqrt{n}}\sum_{i=1}^{n}(\hat{m}_\theta(X_i) - m_\theta(X_i))^\top J_{\rho_n}(\tilde{m}_\theta(X_i))(\hat{m}_\theta(X_i) - m_\theta(X_i)) + o_p(1),$$

with $\alpha(W_i) = \frac{\partial g_{\rho_n}(m_\theta(X_i)}{\partial m}^\top (m_\theta(W_i) - m_\theta(X_i))$

Several remarks are in order. First, $\alpha(\cdot)$ is the adjustment term that arises because of the first stage estimation of $m_\theta(\cdot)$. Second, the 2nd order term in the expansion is of order $O_p(\rho n^{1/2-2\gamma})$ and thus, constrains the rate of divergence of ρ toward $+\infty$ in order this term to go to 0 asymptotically. It is linked to the non-parametric convergence rate of $\hat{m}_\theta(\cdot)$, $\gamma > 1/4$.

Building on proposition 4.8, we can show that the confidence region is asymmptotically valid and consistent.

Proposition 4.9 *Let $\rho_n = cn^\alpha$ with $\alpha < 2\gamma - \frac{1}{2}$ and $c > 0$ a constant, let V_n an estimator of the asymptotic variance of $g_\rho(m_\theta(X_i)) + \alpha(W_i)$, let $\hat{m}_\theta$ a non-parametric estimator satisfying Assumption 6. Then, $\mathrm{CR}_n(1-\alpha)$ is asymptotically valid and consistent, i.e.,*

- *Asymptotic validity:* $\liminf_{n\to\infty} \inf_{\theta\in\Theta_I} \Pr(\theta \in \mathrm{CR}_n(1-\alpha)) \geq 1-\alpha$.
- *Consistency:* $\forall\theta \notin \Theta_I$, $\Pr(\theta \in \mathrm{CR}_n(1-\alpha)) \to 0$.

See the proof in the appendix.

2.5 Monte Carlo simulations

We perform Monte Carlo simulations to evaluate the relative performance of the different procedures proposed in this paper of the identified set as well as the different estimation procedures.

2.5.1 Simulations without covariates

Simulation Design The simulation design directly follows the example introduced in this paper for different sample sizes. Profit functions of firms of type 1 and 2 write as follows:

$$\Pi_1 = \beta_1 - \delta_{1,1}N_1 - \delta_{2,1}N_2 + \varepsilon_1$$

$$\Pi_2 = \beta_2 - \delta_{1,2}N_1 - \delta_{2,2}N_2 + \varepsilon_2,$$

in which

- N_i is the number of firms of type i=1,2
- ε_i for $i = 1, 2$ is the profit shocks. We assume that they are i.i.d., each of them drawn from a standard distribution

In our DGP, we have $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$. To make the exposition easier, we assume that the econometrician knows β_1 and β_2 and that $\delta_{11} = \delta_{22}$ as well as $\delta_{12} = \delta_{21}$. This is of course, unrealistic but allows us to simplify the exposition of the results. The general case does present the same type of conclusions. In the multiple equilibria regions, we assume that a firm of type 1 always decides first and, therefore, we pick the equilibrium with the highest number of type 1 firms. Observe that following Cleeren, Verboven, Dekimpe, and Gielens (2010), we know that in the multiple equilibria regions, $N_1 + N_2$ is invariant.

The graph related to the equilibrium structure is given in Figure 2.4. Therefore, we can sharply characterize the identified set from 9 inequalities derived from each single outcome, completed by 4 inequalities derived from pairs of outcomes (the links in the graph) and one inequality related to the component of outcomes with $N_1 + N_2 = 3$, i.e., a total of 14 inequalities. Among these fourteen inequalities, 5 are equalities (we have five components).

The sample size is $n = 1,000$. The number of Monte Carlo replications is 1,000. For each sample, we compute the decision to reject or not $\theta \in \Theta_I$ for a 5% level of significance. The grid tested is composed by values from 1 to 2 with a tick of 0.02 for δ_{ii} and values from 0.4 to 1.4 with a tick of 0.02 for δ_{ij}, i.e., 2601 points tested in total. We report the mean across simulations of the lowest value and highest value for the two parameters tested. We also report the mean number of points not rejected as well as the coverage rate of the true value θ_0. Given the DGP, we can calculate P_0 which is

$$P_0 = [0.021, 0.074, 0.256, 0.047, 0.131, 0.421, 0.012, 0.034, 0.0004]^\top.$$

Different procedures evaluated Following Section 2.3, we can derive different sets of inequalities to test our candidates θ. We now detail them.

- *core*. Here, we select the set of inequalities which corresponds to the core determining class $\mathcal{C}(\theta)$ that we derived by gathering all the connected subsets as shown in section 2.3.3, applying our algorithm.

- *core*$^+$. Here, we add to the previous set of inequalities, the five equalities satisfied by the model. It is not necessary. In fact, if $P_1 \leq Q_1$ and $P_2 \leq Q_2$ with $P_1 + P_2 = 1$ and $Q_1 + Q_2 = 1$, then $P_1 = Q_1$ and $P_2 = Q_2$. Nevertheless, we decide to add these five inequalities to reinforce the equality requirement.

- *core**. Here, we add to the previous set of inequalities, four out of the five equalities satisfied by the model. The fifth one be redundant, we wonder what would be the impact of ignoring one of them (we drop the last one related to the outcome $(2, 2)$.

- $\min - \max$ set of inequalities which corresponds to an upper bound and a lower bound on the probability of each individual outcome

- $\min - \max^+$. Here we add the three equalities related to the non-single components to the min max inequalities.

Let us emphasize that for all these sets of inequalities, we derive the exact bounds implied by the model as opposed to the rest of the literature, which except in simple cases, simulates the bounds. Thus, the min-max strategy, which has been exploited in earlier work (like in Ciliberto and Tamer (2009) for example) can already be understood as an enhanced version of the min-max strategy.

Concerning the critical value, we used Equation (2.4.7) with the number of inequalities tested (labeled $CCK, p^* =$). In some cases, we were able to refine this maximum number of binding moment inequalities exploiting the geometry of the set $A(\theta)$. We also use Liu and Xie (2020) (label "cauchy") who propose to aggregate the p-values of the different inequalities/equalities tested into one single test statistic. Interestingly, the critical value they propose is valid for any correlation structure.

Finally, we also smooth from each set of inequalities following our methodology and we present the results for different values of ρ.

We display the results in Table 2.1 and 2.2. Let us first focus on the first set of results. Except for the case $\min - \max^+$ with the smallest amount of degrees of freedom, the size is controlled for the true value. Following the specific geometry, we know that 10 out of 14 inequalities are binding at the maximum. Given that firm of type 1 always enter first, P_0 is a kink of the convex set $A(\theta)$ and, therefore, these 10 equalities are indeed biding with $\theta = \theta_0$. Interestingly, despite the fact that testing these fourteen inequalities is equivalent to testing these fourteen inequalities completed by the 5 equalities (i.e. adding the "opposite" inequalities for five of them), the small sample properties are much better in the latter case. One of the message is to include these equalities explicitly in the testing procedure. Deleting one of the redundant one ($core^*$ $CCK,\ p^* =$13) does not help.

The min max procedure seems competitive. We recall that computing the minimum of the probability of a single outcome requires to run our algorithm to the end, which could require a lot of numerical evaluations in more complicated DGPs, or to simulate it (like in Ciliberto and Tamer, 2009) and incorporate an additional noise in the testing procedure. Adding equalities to the min-max procedure seems to work well.

Finally, the aggregation of all test into one dilutes power and does not provide a competitive alternative.

In Table 2.2, we evaluate the same set of equalities/inequalities with our smooth min function $g_\rho(\cdot_n)$ with three values of ρ, 1, 5 and 10. Again, for these procedures, the test statistic is asymptotically pivotal. We consider two versions, one in which we compte the moments $C_{\theta,j} - q_j^\top P_n$ for each j, one in which we standardize each of these moments before incorporating them in the calculation of the test statistic. First, as expected, higher values of ρ lead to smaller but valid confidence regions, which are competitive with respect to the best procedures of Table 2.1. Second, it is better to normalize before aggregating, because it avoids the results to be driven by the moment of the highest variance.

We change the DGP for the results displayed in Table 2.3 and 2.4 by considering uniform profit shocks instead of normally distributed ones. The impact on the data is to get smaller multiple equilibria regions.[19] The results remain qualitatively the same. And so, for the same DGP for which we allocate randomly the outcome in the multiple equilibria regions (Table 2.5 and 2.6).

2.5.2 Simulations with covariates

We know consider the case with covariates, in a similar setting than above. Here, X follows uniform distribution on $[0,1]$. Given X the profit of both types is equal to:

$$\Pi_1 = \beta_1 + \beta_X X - \delta_{1,1} N_1 - \delta_{2,1} N_2 + \varepsilon_1$$
$$\Pi_2 = \beta_2 + \beta_X X - \delta_{1,2} N_1 - \delta_{2,2} N_2 + \varepsilon_2,$$

[19] For the values chosen, the total area of multiplicity is around 6% against 15% for the DGP with normal shocks. Here the new equilibrium probabilities are

$$P_0 = [0.104, 0.078, 0.135, 0.151, 0.047, 0.333, 0.036, 0.094, 0.021]^\top.$$

In our DGP, we have $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.5$, like before, and $\beta_X = 1$. In a first step, the shocks are normally distributed like before.

We compare two strategies. Remember that for our specification, the core determining class does not depend on the realization of X. As a result, one necessary condition for θ to be in the identified set is that the unconditional probability vector (estimated at the standard rate) should belong to a convex set which is the Aumann expectation of $A(\theta, X)$ (see Beresteanu and Molinari (2008)). Therefore, after having calculated the expectation of the quantities involved in the case without covariate (with respect to X), we can perform the same procedure than before in Table 2.1. Results are displayed in Table 2.7. Alternatively, we can apply our smooth procedure and the results are displayed in Table 2.8.

First, the size properties are less good than in the case without covariates. When ρ is too large, the smooth "approximation" is not competitive. The same hierarchy than without the covariates still holds. The results seem to depend on the "size" of the multiple equilibria regions. They are better with the uniform shocks, see Table 2.9. Alternative estimators for the variance calculation of the quantity V_n in 4.9 should be perhaps considered. It is left for future research.

2.6 Conclusion

In this paper, we develop a new method to estimate entry games with multiple equilibria, which may or may be not point identified. First, we propose an algorithm which allows us to characterize the equilibrium structure in polynomial time (in the number of types). This algorithm permits to derive a competitive core determining class and to calculate the bounds used for the moment inequalities generated by this class.

Then, we propose to circumvent the problem of deriving a competitive critical value but easy-to-derive by smoothing the minimum function. The smooth min or max functions have been used in applied mathematics. In our case, it allows us to obtain a pivotal test statistic which automatically eliminates "numerically" the non binding moments. Obviously, there is a trade-off between high values of the smoothing parameter to get close to the true identified set and accuracy of the normal approximation. Values of $\rho = 5$ seem to be competitive though they should be confirmed with more simulations.

Interestingly, our procedure can easily be adapted to the case with covariates, either by testing the unconditional probability vector or by adapting our asymptotic distribution to a non parametric estimator of the conditional probability vector. The square-root speed of convergence of our statistic is recovered. Monte Carlo simulations, study the properties of our procedure and underline the fact that adding redundant moments in the procedure may improve the small sample properties and that size properties of our test are very sensitive to the plug-in of the empirical variance.

Many pending questions remain. First, we are currently working on generalizing this testing procedure to any model defined by conditional moment inequalities. Then, inference methods on subvectors have been proposed in very general settings. It would be worth investigating whether the specific structure of entry games would allow an improvement in the application of these techniques. Finally, we plan to evaluate our strategy on real data.

Tables of results

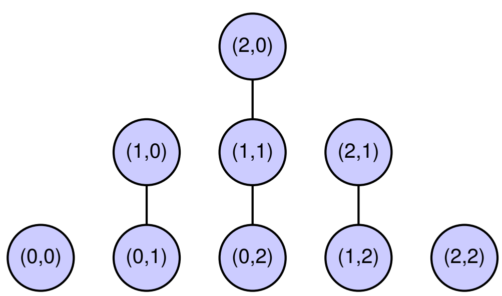


Figure 2.4: $\Gamma(\theta)$ for $\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.1: Coverage rate and confidence region - normal shocks

	Coverage	$\min \delta_{11}$	$\max \delta_{11}$	$\min \delta_{12}$	$\max \delta_{12}$	Nb. points
core $CCK,\ p^* =$10	0.98	1.39	1.58	0.55	1.00	133
core $CCK,\ p^* =$14	0.99	1.38	1.58	0.54	1.02	148
core$^+$ $CCK,\ p^* =$15	0.95	1.41	1.57	0.63	0.99	105
core* $CCK,\ p^* =$13	0.95	1.41	1.57	0.62	1.04	123
core Cauchy	1.00	1.16	1.75	0.41	1.40	976
$\min - \max$ $CCK,\ p^* =$11	0.90	1.38	1.60	0.63	1.02	139
$\min - \max$ $CCK,\ p^* =$18	0.94	1.38	1.61	0.62	1.05	163
$\min - \max$ Cauchy	1.00	1.00	2.00	0.40	1.40	2257
$\min - \max^+$ $CCK,\ p^* =$17	0.89	1.41	1.57	0.63	0.97	97
$\min - \max^+$ $CCK,\ p^* =$24	0.93	1.41	1.58	0.63	0.99	108

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.2: Coverage rate and confidence region - normal shocks - Pivotal tests

	Coverage	min δ_{11}	max δ_{11}	min δ_{12}	max δ_{12}	Nb. points
Smooth-*core* NS ρ =1	1.00	1.00	1.73	0.40	1.40	1716
Smooth-*core* NS ρ =5	1.00	1.08	1.65	0.43	1.40	1170
Smooth-*core* NS ρ =10	1.00	1.25	1.62	0.50	1.40	719
Smooth-*core* STD ρ =1	1.00	1.00	1.82	0.40	1.40	1811
Smooth-*core* STD ρ =5	1.00	1.24	1.64	0.41	1.31	667
Smooth-*core* STD ρ =10	1.00	1.34	1.60	0.51	1.07	273
Smooth-*core*$^+$ NS ρ =1	1.00	1.00	1.72	0.40	1.40	1659
Smooth-*core*$^+$ NS ρ =5	1.00	1.19	1.64	0.44	1.40	912
Smooth-*core*$^+$ NS ρ =10	1.00	1.31	1.61	0.52	1.40	561
Smooth-*core*$^+$ STD ρ =1	1.00	1.00	1.75	0.40	1.40	1440
Smooth-*core*$^+$ STD ρ =5	1.00	1.30	1.61	0.54	1.32	421
Smooth-*core*$^+$ STD ρ =10	1.00	1.37	1.58	0.61	1.06	171
Smooth-min − max NS ρ =1	1.00	1.00	2.00	0.40	1.40	2547
Smooth-min − max NS ρ =5	1.00	1.01	1.85	0.45	1.40	1749
Smooth-min − max NS ρ =10	1.00	1.19	1.75	0.51	1.40	1114
Smooth-min − max STD ρ =1	1.00	1.00	1.99	0.43	1.40	2014
Smooth-min − max STD ρ =5	1.00	1.23	1.69	0.57	1.40	825
Smooth-min − max STD ρ =10	0.99	1.33	1.62	0.62	1.26	380
Smooth-min − max$^+$ NS ρ =1	1.00	1.00	1.97	0.40	1.40	2237
Smooth-min − max$^+$ NS ρ =5	1.00	1.20	1.68	0.48	1.40	1031
Smooth-min − max$^+$ NS ρ =10	1.00	1.31	1.63	0.54	1.40	647
Smooth-min − max$^+$ STD ρ =1	1.00	1.00	1.82	0.45	1.40	1645
Smooth-min − max$^+$ STD ρ =5	1.00	1.31	1.63	0.58	1.39	500
Smooth-min − max$^+$ STD ρ =10	0.98	1.38	1.59	0.63	1.10	188

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.3: Coverage rate and confidence region - unif. shocks

	Coverage	$\min \delta_{11}$	$\max \delta_{11}$	$\min \delta_{12}$	$\max \delta_{12}$	Nb. points
core $CCK,\ p^* = 10$	0.98	1.29	1.72	0.47	1.08	398
core $CCK,\ p^* = 14$	0.99	1.28	1.74	0.45	1.10	448
core$^+$ $CCK,\ p^* = 15$	0.95	1.31	1.68	0.56	0.99	231
core* $CCK,\ p^* = 13$	0.96	1.23	1.68	0.46	1.13	517
core Cauchy	0.98	1.01	1.94	0.45	1.40	1273
$\min - \max$ $CCK,\ p^* = 11$	0.94	1.29	1.68	0.56	1.04	248
$\min - \max$ $CCK,\ p^* = 18$	0.95	1.27	1.70	0.54	1.07	303
$\min - \max$ Cauchy	0.99	1.00	2.00	0.40	1.40	2028
$\min - \max^+$ $CCK,\ p^* = 17$	0.93	1.32	1.67	0.56	0.99	221
$\min - \max^+$ $CCK,\ p^* = 24$	0.94	1.30	1.68	0.55	1.01	251

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.4: Coverage rate and confidence region - unif. shocks - Pivotal tests

	Coverage	min δ_{11}	max δ_{11}	min δ_{12}	max δ_{12}	Nb. points
Smooth-*core* NS ρ =1	1.00	1.00	2.00	0.54	1.40	2063
Smooth-*core* NS ρ =5	1.00	1.00	1.90	0.56	1.40	1753
Smooth-*core* NS ρ =10	0.99	1.03	1.79	0.58	1.40	1418
Smooth-*core* STD ρ =1	1.00	1.00	1.93	0.40	1.40	2369
Smooth-*core* STD ρ =5	1.00	1.12	1.73	0.47	1.30	1026
Smooth-*core* STD ρ =10	1.00	1.25	1.67	0.54	1.08	459
Smooth-$core^+$ NS ρ =1	1.00	1.00	1.99	0.55	1.40	2012
Smooth-$core^+$ NS ρ =5	1.00	1.01	1.83	0.57	1.40	1554
Smooth-$core^+$ NS ρ =10	0.99	1.16	1.74	0.59	1.37	954
Smooth-$core^+$ STD ρ =1	1.00	1.00	1.91	0.45	1.40	1943
Smooth-$core^+$ STD ρ =5	1.00	1.20	1.71	0.56	1.26	681
Smooth-$core^+$ STD ρ =10	0.96	1.30	1.64	0.61	1.05	291
Smooth-min $-$ max NS ρ =1	1.00	1.00	2.00	0.40	1.40	2597
Smooth-min $-$ max NS ρ =5	1.00	1.00	2.00	0.42	1.40	2269
Smooth-min $-$ max NS ρ =10	1.00	1.05	1.98	0.47	1.40	1788
Smooth-min $-$ max STD ρ =1	1.00	1.00	2.00	0.40	1.40	2293
Smooth-min $-$ max STD ρ =5	1.00	1.10	1.80	0.53	1.40	1253
Smooth-min $-$ max STD ρ =10	0.99	1.25	1.69	0.59	1.23	551
Smooth-min $-$ max^+ NS ρ =1	1.00	1.00	2.00	0.40	1.40	2563
Smooth-min $-$ max^+ NS ρ =5	1.00	1.05	1.93	0.46	1.40	1793
Smooth-min $-$ max^+ NS ρ =10	1.00	1.18	1.78	0.51	1.37	1077
Smooth-min $-$ max^+ STD ρ =1	1.00	1.00	2.00	0.42	1.40	2189
Smooth-min $-$ max^+ STD ρ =5	1.00	1.20	1.73	0.56	1.30	781
Smooth-min $-$ max^+ STD ρ =10	0.97	1.31	1.65	0.61	1.07	321

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.5: Coverage rate and confidence region - unif. shocks -interior point

	Coverage	min δ_{11}	max δ_{11}	min δ_{12}	max δ_{12}	Nb. points
core $CCK,\ p^* = 10$	0.99	1.29	1.73	0.46	1.08	408
core $CCK,\ p^* = 14$	0.99	1.28	1.74	0.45	1.10	458
core$^+$ $CCK,\ p^* = 15$	0.96	1.31	1.68	0.55	0.99	234
core* $CCK,\ p^* = 13$	0.97	1.21	1.68	0.45	1.16	551
core Cauchy	0.98	1.01	1.92	0.46	1.40	1150
min − max $CCK,\ p^* = 11$	0.96	1.28	1.69	0.54	1.05	277
min − max $CCK,\ p^* = 18$	0.97	1.26	1.71	0.52	1.08	335
min − max Cauchy	0.99	1.00	2.00	0.40	1.40	2002
min − max$^+$ $CCK,\ p^* = 17$	0.96	1.31	1.67	0.55	0.99	235
min − max$^+$ $CCK,\ p^* = 24$	0.97	1.30	1.68	0.54	1.01	267
Smooth-*core* NS $\rho = 1$	1.00	1.00	2.00	0.54	1.40	2062
Smooth-*core* NS $\rho = 5$	1.00	1.00	1.90	0.56	1.40	1758
Smooth-*core* NS $\rho = 10$	1.00	1.03	1.79	0.57	1.40	1433
Smooth-*core* STD $\rho = 1$	1.00	1.00	1.92	0.40	1.40	2318
Smooth-*core* STD $\rho = 5$	1.00	1.12	1.72	0.47	1.30	992
Smooth-*core* STD $\rho = 10$	1.00	1.25	1.66	0.54	1.08	460
Smooth-*core*$^+$ NS $\rho = 1$	1.00	1.00	1.99	0.55	1.40	2010
Smooth-*core*$^+$ NS $\rho = 5$	1.00	1.01	1.83	0.57	1.40	1559
Smooth-*core*$^+$ NS $\rho = 10$	0.99	1.16	1.74	0.58	1.37	967
Smooth-*core*$^+$ STD $\rho = 1$	1.00	1.00	1.90	0.47	1.40	1879
Smooth-*core*$^+$ STD $\rho = 5$	0.99	1.20	1.70	0.57	1.25	658
Smooth-*core*$^+$ STD $\rho = 10$	0.96	1.30	1.64	0.61	1.05	289
Smooth-min − max NS $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2598
Smooth-min − max NS $\rho = 5$	1.00	1.00	2.00	0.41	1.40	2283
Smooth-min − max NS $\rho = 10$	1.00	1.05	1.98	0.46	1.40	1816
Smooth-min − max STD $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2298
Smooth-min − max STD $\rho = 5$	1.00	1.09	1.80	0.52	1.40	1284
Smooth-min − max STD $\rho = 10$	1.00	1.23	1.69	0.58	1.24	597

Table 2.6: Coverage rate and confidence region - normal. shocks -interior point

	Coverage	$\min \delta_{11}$	$\max \delta_{11}$	$\min \delta_{12}$	$\max \delta_{12}$	Nb. points
$core\ CCK,\ p^* = 10$	0.99	1.39	1.59	0.49	1.01	162
$core\ CCK,\ p^* = 14$	1.00	1.38	1.59	0.49	1.02	178
$core^+\ CCK,\ p^* = 15$	0.95	1.41	1.59	0.60	0.99	115
$core^*\ CCK,\ p^* = 13$	0.96	1.40	1.59	0.59	1.07	145
core Cauchy	0.99	1.16	1.78	0.42	1.40	974
$\min - \max\ CCK,\ p^* = 11$	0.94	1.35	1.61	0.60	1.06	201
$\min - \max\ CCK,\ p^* = 18$	0.96	1.35	1.62	0.59	1.09	230
$\min - \max$ Cauchy	1.00	1.00	2.00	0.40	1.40	2193
$\min - \max^+\ CCK,\ p^* = 17$	0.93	1.41	1.59	0.60	0.99	115
$\min - \max^+\ CCK,\ p^* = 24$	0.96	1.41	1.59	0.59	1.00	126
Smooth-*core* NS $\rho = 1$	1.00	1.00	1.73	0.40	1.40	1689
Smooth-*core* NS $\rho = 5$	1.00	1.09	1.66	0.41	1.40	1165
Smooth-*core* NS $\rho = 10$	1.00	1.25	1.63	0.46	1.40	736
Smooth-*core* STD $\rho = 1$	1.00	1.00	1.80	0.40	1.40	1736
Smooth-*core* STD $\rho = 5$	1.00	1.24	1.65	0.40	1.25	625
Smooth-*core* STD $\rho = 10$	1.00	1.34	1.61	0.49	1.05	290
Smooth-$core^+$ NS $\rho = 1$	1.00	1.00	1.72	0.40	1.40	1631
Smooth-$core^+$ NS $\rho = 5$	1.00	1.20	1.64	0.43	1.40	905
Smooth-$core^+$ NS $\rho = 10$	1.00	1.31	1.62	0.49	1.40	566
Smooth-$core^+$ STD $\rho = 1$	1.00	1.00	1.72	0.40	1.40	1349
Smooth-$core^+$ STD $\rho = 5$	1.00	1.30	1.61	0.54	1.26	376
Smooth-$core^+$ STD $\rho = 10$	1.00	1.37	1.59	0.60	1.04	171
Smooth-$\min - \max$ NS $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2561
Smooth-$\min - \max$ NS $\rho = 5$	1.00	1.02	1.86	0.42	1.40	1789
Smooth-$\min - \max$ NS $\rho = 10$	1.00	1.20	1.75	0.47	1.40	1150
Smooth-$\min - \max$ STD $\rho = 1$	1.00	1.00	1.99	0.42	1.40	2010
Smooth-$\min - \max$ STD $\rho = 5$	1.00	1.22	1.69	0.55	1.40	873
Smooth-$\min - \max$ STD $\rho = 10$	1.00	1.31	1.64	0.60	1.27	438

Table 2.7: Coverage rate and confidence region - normal. shocks - with covariates

	Coverage	$\min \delta_{11}$	$\max \delta_{11}$	$\min \delta_{12}$	$\max \delta_{12}$	Nb. points
core CCK, $p^* = 10$	0.98	1.37	1.63	0.65	0.95	113.00
$core^+$ *CCK*, $p^* = 15$	0.88	1.37	1.59	0.66	0.90	78.00
$core^*$ *CCK*, $p^* = 13$	0.93	1.35	1.59	0.65	0.91	97.00
$\min - \max^+$ *CCK*, $p^* = 17$	0.83	1.38	1.59	0.66	0.89	73.00
$\min - \max^+$ *CCK*, $p^* = 24$	0.85	1.37	1.59	0.65	0.90	81.00
Smooth-*core* P_0 $\rho = 1$	1.00	1.00	1.99	0.40	1.40	2196.00
Smooth-*core* P_0 $\rho = 5$	1.00	1.18	1.70	0.54	1.14	563.00
Smooth-*core* P_0 $\rho = 10$	1.00	1.30	1.63	0.61	0.99	233.00
Smooth-$core^+$ P_0 $\rho = 1$	1.00	1.00	1.84	0.42	1.40	1626.00
Smooth-$core^+$ P_0 $\rho = 5$	0.90	1.23	1.65	0.60	1.08	343.00
Smooth-$core^+$ P_0 $\rho = 10$	0.89	1.33	1.60	0.64	0.96	141.00
Smooth-$\min - \max$ P_0 $\rho = 1$	1.00	1.00	1.98	0.44	1.40	2019.00
Smooth-$\min - \max$ P_0 $\rho = 5$	0.90	1.15	1.68	0.59	1.34	758.00
Smooth-$\min - \max$ P_0 $\rho = 10$	0.90	1.28	1.63	0.64	1.05	264.00
Smooth-$\min - \max^+$ P_0 $\rho = 1$	1.00	1.00	1.85	0.47	1.40	1711.00
Smooth-$\min - \max^+$ P_0 $\rho = 5$	0.90	1.23	1.64	0.61	1.10	364.00
Smooth-$\min - \max^+$ P_0 $\rho = 10$	0.90	1.33	1.60	0.65	0.95	144.00

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.8: Coverage rate and confidence region - normal. shocks -interior point

	Coverage	min δ_{11}	max δ_{11}	min δ_{12}	max δ_{12}	Nb. points
Smooth-*core* h_0 ρ =1	1.00	1.00	2.00	0.40	1.40	2204.00
Smooth-*core* h_0 ρ =5	0.98	1.35	1.76	0.59	1.35	599.00
Smooth-*core*$^+$ h_0 ρ =1	0.95	1.00	1.89	0.44	1.40	1655.00
Smooth-*core*$^+$ h_0 ρ =5	0.63	1.38	1.70	0.66	1.26	332.00
Smooth-min $-$ max h_0 ρ =1	1.00	1.00	1.98	0.48	1.40	1975.00
Smooth-min $-$ max h_0 ρ =5	0.89	1.26	1.80	0.64	1.40	830.00
Smooth-min $-$ max$^+$ h_0 ρ =1	0.99	1.00	1.92	0.51	1.40	1726.00
Smooth-min $-$ max$^+$ h_0 ρ =5	0.67	1.37	1.70	0.68	1.33	406.00

$\beta_1 = 3$, $\beta_2 = 2$, $\delta_{11} = \delta_{22} = 1.5$ and $\delta_{12} = \delta_{21} = 0.75$

Table 2.9: Coverage rate and confidence region - unif. shocks -interior point

	Coverage	$\min \delta_{11}$	$\max \delta_{11}$	$\min \delta_{12}$	$\max \delta_{12}$	Nb. points
$core\ CCK,\ p^* = 15$	0.97	1.28	1.70	0.55	0.96	256.00
Minmax $CCK,\ p^* = 11$	0.93	1.25	1.68	0.55	1.01	263.00
Minmax $CCK,\ p^* = 18$	0.96	1.23	1.70	0.53	1.03	308.00
$\min - \max^+ CCK,\ p^* = 17$	0.94	1.28	1.68	0.56	0.96	237.00
$\min - \max^+ CCK,\ p^* = 24$	0.96	1.27	1.69	0.55	0.97	264.00
Smooth-$core$ P_0 $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2300.00
Smooth-$core$ P_0 $\rho = 5$	1.00	1.15	1.84	0.50	1.33	1124.00
Smooth-$core$ P_0 $\rho = 10$	0.99	1.26	1.72	0.56	1.08	460.00
Smooth-$core^+$ P_0 $\rho = 1$	1.00	1.00	2.00	0.43	1.40	2122.00
Smooth-$core^+$ P_0 $\rho = 5$	0.99	1.20	1.76	0.56	1.24	744.00
Smooth-$core^+$ P0y $\rho = 10$	0.98	1.30	1.67	0.61	1.03	304.00
Smooth-$\min - \max$ P_0 $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2366.00
Smooth-$\min - \max$ P_0 $\rho = 5$	1.00	1.07	1.81	0.52	1.39	1308.00
Smooth-$\min - \max$ P_0 $\rho = 10$	0.99	1.23	1.69	0.59	1.12	490.00
Smooth-$\min - \max^+$ P_0-rho=y1	1.00	1.00	2.00	0.42	1.40	2288.00
Smooth-$\min - \max^+$ P_0 $\rho = 5$	1.00	1.17	1.74	0.55	1.24	792.00
Smooth-$\min - \max^+$ P_0 $\rho = 10$	0.98	1.29	1.66	0.61	1.02	317.00
Smooth-$core$ h_0 $\rho = 1$	1.00	1.00	2.00	0.40	1.40	2290.00
Smooth-$core$ h_0 $\rho = 5$	1.00	1.19	1.95	0.48	1.39	1434.00
Smooth-$core$ h_0 $\rho = 10$	0.90	1.33	1.84	0.56	1.21	673.00
Smooth-$core^+$ h_0 $\rho = 1$	1.00	1.00	2.00	0.41	1.40	2134.00
Smooth-$core^+$ h0y $\rho = 5$	0.97	1.23	1.88	0.54	1.36	1055.00
Smooth-$core^+$ h0y $\rho = 10$	0.71	1.37	1.78	0.61	1.15	445.00
Smooth-$\min - \max$ h_0 $\rho = 1$	1.00	1.00	2.00	0.41	1.40	2325.00
Smooth-$\min - \max$ h_0 $\rho = 5$	1.00	1.12	1.93	0.55	1.40	1441.00
Smooth-$\min - \max$ h_0 $\rho = 10$	0.85	1.31	1.80	0.63	1.30	691.00
Smooth-$\min - \max^+$ h_0 $\rho = 1$	1.00	1.00	2.00	0.42	1.40	2259.00
Smooth-$\min - \max^+$ h_0 $\rho = 5$	0.98	1.22	1.85	0.58	1.36	1024.00

Acknowledgements

We gratefully acknowledge financial support from the Agence Nationale de Recherche ANR-05-BLAN-0024-01. This research has also received financial support from the European Research Council under the European Community Seventh Framework Program FP7/2007-2013 grant agreement N_295298. We would like to thank participants at various conferences and seminars for helpful discussions.

Bibliography

ACKERBERG, D., X. CHEN, J. HAHN, AND Z. LIAO (2014): "Asymptotic efficiency of semiparametric two-step GMM," *Review of Economic Studies*, 81, 919–943.

AI, C. AND X. CHEN (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1843.

ANDREWS, D. W., S. BERRY, AND P. JIA (2004): "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location," *WP Cowles*.

ANDREWS, D. W. AND X. SHI (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666.

ANDREWS, D. W. AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

ANDREWS, D. W. K. AND P. J. BARWICK (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

ANDREWS, I., J. ROTH, AND A. PAKES (Forthcoming): "Inference for Linear Conditional Moment Inequalities," *Review of Economic Studies*.

ARADILLAS-LOPEZ, A. AND A. M. ROSEN (2022): "Inference in ordered response games with complete information," *Journal of Econometrics*, 226, 451–476.

Armstrong, T. B. (2014): "Weighted KS statistics for inference on conditional moment inequalities," *Journal of Econometrics*, 181, 92–116.

Armstrong, T. B. and H. P. Chan (2016): "Multiscale adaptive inference on conditional moment inequalities," *Journal of Econometrics*, 194, 24–43.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006): "Estimating static models of strategic interaction," Tech. rep., National Bureau of Economic Research.

Bajari, P., H. Hong, and S. P. Ryan (2010): "Identification and estimation of a discrete game of complete information," *Econometrica*, 78, 1529–1568.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79, 1785–1821.

Beresteanu, A., I. Molchanvb, and F. Molinari (2012): "Partial identification using random set theory," *Journal of Econometrics*, 166, 17–32.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, 889–917.

Berry, S. T. and G. Compiani (2022): "An Instrumental Variable Approach to Dynamic Models," *The Review of Economic Studies*, rdac061.

Bjorn, P. A. and Q. H. Vuong (1984): "Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation," .

BONTEMPS, C. AND R. KUMAR (2020): "A geometric approach to inference in set-identified entry games," *Journal of Econometrics*, 218, 373–389.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): "Set identified linear models," *Econometrica*, 80, 1129–1155.

BRESNAHAN, T. F. AND P. C. REISS (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48, 57–81.

——— (1991b): "Entry and competition in concentrated markets," *Journal of political economy*, 99, 977–1009.

BREUSCH, T., H. QIAN, P. SCHMIDT, AND D. WYHOWSKI (1999): "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

BUGNI, F. A., I. A. CANAY, AND X. SHI (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8, 1–38.

CHEN, X., T. M. CHRISTENSEN, AND E. TAMER (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965–2018.

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018a): "Double/debiased machine learning for treatment and structural parameters," .

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2018b): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86, 1867–1900.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 75, 1243–1284.

CHERNOZHUKOV, V., E. KOCATULUM, AND K. MENZEL (2015): "Inference on sets in finance," *Quantitative Economics*, 6, 309–358.

CHESHER, A. AND A. M. ROSEN (2017): "Generalized instrumental variable models," *Econometrica*, 85, 959–989.

——— (2019): "Structural modeling of simultaneous discrete choice." *Cemmap WP*.

CHO, J. AND T. M. RUSSELL (2018): "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments," .

CILIBERTO, F., C. MURRY, AND E. TAMER (2021): "Market structure and competition in airline markets," *Journal of Political Economy*, 129, 2995–3038.

CILIBERTO, F. AND E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77, 1791–1828.

CLEEREN, K., F. VERBOVEN, M. G. DEKIMPE, AND K. GIELENS (2010): "Intra-and interformat competition among discounters and supermarkets," *Marketing science*, 29, 456–473.

COX, G. AND X. SHI (2022): "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *The Review of Economic Studies*, rdac015.

EIZENBERG, A. (2014): "Upstream innovation and product variety in the us home pc market," *Review of Economic Studies*, 81, 1003–1045.

GAFAROV, B. (2019): "Inference in high-dimensional set-identified affine models," .

GALICHON, A. AND M. HENRY (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, 78, 1264–1298.

GRIECO, P. L. (2014): "Discrete games with flexible information structures: An application to local grocery markets," *The RAND Journal of Economics*, 45, 303–340.

KAIDO, H., F. MOLINARI, AND J. STOYE (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432.

KITAMURA, Y. AND J. STOYE (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86, 1883–1909.

LIU, Y. AND J. XIE (2020): "Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures," *Journal of the American Statistical Association*, 115, 393–402, pMID: 33012899.

LUO, Y. AND H. WANG (2017): "Core Determining Class and Inequality Selection," *American Economic Review*, 107, 274–77.

MAGNOLFI, L. AND C. RONCORONI (2022): "Estimation of Discrete Games with Weak Assumptions on Information," *The Review of Economic Studies*, rdac058.

MANSKI, C. F. (1995): *Identification problems in the social sciences*, Harvard University Press.

MAZZEO, M. J. (2002): "Product choice and oligopoly market structure," *RAND Journal of Economics*, 221–242.

MOLCHANOV, I. AND F. MOLINARI (2014): "Applications of random set theory in econometrics," *Annu. Rev. Econ.*, 6, 229–251.

NEWEY, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, 1349–1382.

NEWEY, W. K. AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

ROCKAFELLAR, R. (1997): *Convex Analysis*, Convex Analysis, Princeton University Press.

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): "A practical two-step method for testing moment inequalities," *Econometrica*, 82, 1979–2002.

ROSEN, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146, 107–117.

TOIVANEN, O. AND M. WATERSON (2005): "Market structure and entry: Where's the beef?" *RAND Journal of Economics*, 680–699.

ZHENG, J. X. (1996): "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics*, 75, 263–289.

2.A Extension

2.A.1 Inference with the LSE smoothing function

Instead of using the Boltzmann operator as a smoothing function, we could have used the LogSumExp (LSE) function. This function is also used in machine-learning and numerical optimization. In this section, we keep the same notations as previously: $m_\theta \equiv \mathbb{E}[m(Y,\theta)]$ and we omit the dependence of p in θ.

For $z = (z_1, z_2, ..., z_p) \in \mathbb{R}^p$, the LSE smooth approximation $g_\rho(\cdot)$ of the minimum function is as follows:

$$g_\rho(z) = -\rho^{-1} \log\Big(\sum_{j=1}^{p} \exp(-\rho z_j)\Big),$$

in which ρ, the smoothing parameter, controls the level of approximation. A nice property of this approximation is that it is possible to control for the difference between the minimum and its approximation through the following inequality:

$$0 \leq \min_{1 \leq j \leq p} z_j - g_\rho(z) \leq \rho^{-1} \log(p). \tag{2.A.11}$$

This inequality is straightforward to derive and the upper bound is reached when all the elements in z are equal. The minimum is simply the limit of $g_\rho(z)$ when $\rho \to +\infty$.

As we did with the Boltzmann operator, we can show that a smooth outer set $\Theta_I^o(\rho)$ of the identified set Θ_I writes as follows:

We know collect the set $\Theta_I^o(\rho)$ of parameters θ such that

$$\Theta_I^o(\rho) = \left\{ \theta \in \mathbb{R}^{q+q'} \mid g_\rho(m_\theta) = -\rho^{-1} \log \left(\frac{1 + \sum_{j=1}^{p} \exp(-\rho m_{\theta,j})}{p+1} \right) \geq 0 \right\}. \tag{2.A.12}$$

Uncentered test statistic. Similarly to what we did with the Boltzmann operator, let us define our smooth test statistic as follows:

$$\xi_n(\theta) = \sqrt{n} \frac{g_{\rho_n}(m_{\theta,n})}{\sqrt{\nabla g_{\rho_n}(m_{\theta,n})^T \Sigma_n \nabla g_{\rho_n}(m_{\theta,n})}}$$

with $\nabla g_{\rho_n}(\cdot)$ the gradient of $g_{\rho_n}(\cdot)$ that we define in the appendix and Σ_n a consistent estimator of Σ_0. Under the assumption that the moments are normally distributed, we can show that the uncentered statistic has the following asymptotic expansion

Proposition 1.1 (Asymptotic expansion of the test statistic.) *Let ρ_n a divergent sequence of positive number such that $\rho_n = O(n^\alpha)$, $0 < a < 1/2$, then*

$$\sqrt{n} g_{\rho_n}(m_{\theta,n}) = \sqrt{n} g_{\rho_n}(m_\theta) + \Gamma_0(\theta) \sqrt{n}(m_{\theta,n} - m_\theta) + o_p(1)$$

with $\Gamma_0(\theta) \equiv \lim_{\rho \to \infty} \nabla g_\rho(m_\theta)$.

The proof is in the appendix. The constraint on the speed of ρ_n implies that the smoothing cannot be faster than the standard square-root speed of convergence. Contrary to the Boltzmann case, here the term $\sqrt{n} g_{\rho_n}(m_\theta)$ does not vanish asymptotically and thus we need to recenter our test statistic. This is why we favor the Boltzmann operator over the LSE smothing function.

Let p^* the number of binding moments and $\widehat{p^*}_n$ a consistent estimator of p^*. Namely, for $\tau_n = n^{-\beta}$ and $0 < \beta < \frac{1}{2}$,

$$\widehat{p^*}_n = \sum_{j=1}^{p} \mathbf{1} \left\{ \frac{m_{\theta,j}}{\sqrt{\sigma_{n,j}^2}} < \tau_n \right\}. \tag{2.A.13}$$

$\widehat{p^*}_n$ is obtained collecting the number of empirical moments which are "close to 0", with $\sigma_{n,j}$ a consistent estimator of $Var(m_{\theta,j})$. Our re-centered test statistic writes:

$$\tilde{\xi}_n(\theta) = \sqrt{n} \frac{g_{\rho_n}(m_{\theta,n}) - \frac{1}{\rho_n} \log\left(\frac{1+p}{1+\widehat{p^*}_n} \right)}{\sqrt{\nabla g_{\rho_n}(m_{\theta,n})^T \Sigma_n \nabla g_{\rho_n}(m_{\theta,n})}}$$

The confidence region associated with the re-centered test statistic simply writes:

$$\mathrm{CR}_n(1-\alpha) = \{\theta \in \mathbb{R}^{q+q'}, \tilde{\xi}_n(\theta) \geq z_\alpha\}, \tag{2.A.14}$$

Next, we derive the asymptotic distribution of the re-centered test statistic and we show that our procedure yields valid confidence region with asymptotically the exact size when at least one moment is binding. Moreover, our procedure is consistent: the econometrician rejects with probability that goes to 1 when $\theta \notin \Theta_I$. The next proposition describes the asymptotic behavior of the test statistic.

Proposition 1.2 *Let ρ_n a divergent sequence of positive numbers such that $\rho_n = O(n^\alpha)$, $0 < a < 1/2$ and let $\widehat{p^*}_n$ an estimator of the number of binding moments, p^*, defined in Equation (2.A.13). Then there are 3 different cases:*

- *$\theta \in int(\Theta_I)$ ($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0):*

$$Pr(\tilde{\xi}_n(\theta) > z_\alpha) \underset{n \to +\infty}{\to} 1$$

- $\theta \in \partial\Theta_I$ *($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0). Then our test statistic is asymptotically distributed as a standard normal:*

$$\tilde{\tilde{\xi}}_n(\theta) \xrightarrow{d} \mathcal{N}(0,1)$$

- $\theta \notin \Theta_I$: $\qquad Pr(\tilde{\tilde{\xi}}_n(\theta) > z_\alpha) \underset{n\to+\infty}{\to} 0$

The proof of this proposition is in the appendix. With the LSE smoothing function, our procedure is also very simple (even if unlike with the Boltzmann smoother, we need to estimate the number of non-binding moments). The test statistic and the critical value are straightforward to derive.

Remarks. We achieve this subsection with two remarks.

- As we did with the Boltzmann operator, it is possible to modify the test statistic by standardizing the empirical moments by the estimated variance.
- Second, we can still compute confidence regions based on $\xi_n(\theta)$ instead of the re-centered version. This would avoid to estimate the number of binding moments but leads to conservative procedure.

2.A.2 Alternative solutions for the choice of ρ_n

As an alternative approach, we can take ρ_n large to "kill" the identification bias and take into account the second-order term in the asymptotic distribution. One of the drawbacks is that here, the asymptotic distribution is no longer pivotal and we need to resort to simulation methods.

2.B Proof of Propositions

Selection of the moment inequalities

Proof of proposition 3.1

We want to prove the following statement: if A is not connected in $\Gamma(\theta)$, then A generates a redundant moment.

To show this, we are going to use the equivalent definition of connectedness. A subset $A \subset \mathcal{Y}$ is connected in $\Gamma(\theta)$ if and only if for every partition in 2 subsets A_1 and A_2 of A, there exists at least one element $y_1 \in A_1$ and $y_2 \in A_2$ that have overlapping equilibrium regions $\mathcal{R}_\omega(X, y_1)$ and $\mathcal{R}_\omega(X, y_2)$.

Assume that A is not connected in $\Gamma(\theta)$, then there exists A_1 and A_2 such that $A = A_1 \cup A_2$ and A_1 and A_2 are such that $\mathcal{R}_\omega(X, A_1) \cap \mathcal{R}_\omega(X, A_2) = \emptyset$. First, let us consider the moment inequalities generated by A_1 and A_2 separately. We have

$$P_0(Y \in A_1|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_1)) \text{ and } P_0(Y \in A_2|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_2)).$$

By combining these 2 inequalities, we have:

$$\begin{aligned} & P_0(Y \in A_1|X) + P_0(Y \in A_2|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_1)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_2)) \\ \iff & P_0(Y \in A_1 \cup A_2|X) \leq P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_1)) + P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_2)) - P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A_1) \cap \mathcal{R}_\omega(X, A_2)) \\ \iff & P_0(Y \in A|X) = P_0(Y \in A_1 \cup A_2|X) \leq \mathcal{R}_\omega(X, A_1) \cup \mathcal{R}_\omega(X, A_2)) = P_\eta(\varepsilon \in \mathcal{R}_\omega(X, A). \end{aligned}$$

The second line in the expression above comes from the fact that A_1 and A_2 are disjoint and thus $P_0(Y \in A_1 \cup A_2) = P_0(Y \in A_1) + P_0(Y \in A_2)$. What's more, by assumption $\mathcal{R}_\omega(A_1) \cap \mathcal{R}_\omega(A_2) = \emptyset$ and as a consequence, $P_\eta(\varepsilon \in \mathcal{R}_\omega(A_1) \cap \mathcal{R}_\omega(A_2)) = 0$. The last line stems from the inclusion-exclusion formula. This proves our result. ■

Proof of proposition 3.2

From proposition 2.1, we know that an outcome $y_1 = (N_1, ..., N_T)$ is a NE of this game if and only if: $\forall t \in \mathcal{T}$,

$$-\pi_t(X, N_t, \mathbf{N}_{-t}; \omega) \leq \varepsilon_t \leq -\pi_t(X, N_t + 1, \mathbf{N}_{-t}; \omega).$$

This system of $2T$ inequalities defines a region $\mathcal{R}_\omega(X, y_1)$ in the space of unobserved heterogeneity. Analogously, we can define a region $\mathcal{R}_\omega(X, y_2)$ in the space of unobserved heterogeneity where $y_2 = (\bar{N}_1, ..., \bar{N}_\tau)$ is a NE of this game. $\mathcal{R}_\omega(X, y_1)$ and $\mathcal{R}_\omega(X, y_2)$ are T-cubes in $\mathbb{R}^T$. Hence, these regions have a non-empty intersection if and only if for each dimension (which here corresponds to a type), the projections of these cubes have a non-empty intersections. Formally,

$$\mathcal{R}_\omega(X, y_1) \cap \mathcal{R}_\omega(X, y_2) \neq \emptyset \iff \forall t \in \mathcal{T}, \text{Proj}(\mathcal{R}_\omega(X, y_1)|e_t) \cap \text{Proj}(\mathcal{R}_\omega(X, y_2)|e_t) \neq \emptyset$$

Once again, by definition of $\mathcal{R}_\omega(X, y_1)$, $\text{Proj}(\mathcal{R}_\omega(X, y_1)|e_t) = [-\pi_t(X, N_t, \mathbf{N}_{-t}; \omega) \leq \varepsilon_t \leq -\pi_t(X, N_t + 1, \mathbf{N}_{-t}; \omega)[$. Likewise, $\text{Proj}(\mathcal{R}_\omega(X, y_2)|e_t) = [-\pi_t(X, \bar{N}_t, \bar{\mathbf{N}}_{-t}; \omega) \leq \varepsilon_t \leq -\pi_t(X, \bar{N}_t + 1, \bar{\mathbf{N}}_{-t}; \omega)[$

Furthermore, from basic analysis results on sets in $\mathbb{R}$, $[a, b[\cap [c, d[\neq \emptyset \iff a < d$ and $c < b$, which applied to our two regions, proves the result. ■

Inference without covariates

Boltzmann operator as smoothing function

For $z = (z_1, z_2, ..., z_p) \in \mathbb{R}^p$, we have a smooth approximation of the minimum function:

$$g_\rho(z) = \frac{\sum_{j=1}^p z_j \exp(-\rho z_j)}{1 + \sum_{j=1}^p \exp(-\rho z_j)},$$

in which ρ, the smoothing parameter, controls the level of approximation. We consider a set-up in which we have p moment inequalities. The identified set is defined as follows

$$\Theta_I = \{\theta \in \Theta \mid \text{for } j = 1, ..., p,\ \mathbb{E}[m_j(W, \theta)] \geq 0\}$$

For simplicity, we define $m_\theta \equiv \mathbb{E}[m(W,\theta)]$. A special case of moment inequalities is the case of games in which the j^{th} is defined as follows:

$$m_{\theta,j} = h_{\theta,j}(P_0) = C_{\theta,j} - q_j^\top P_0.$$

In matricial form:

$$h_{\theta,j}(P_0) = C_\theta - QP_0.$$

We denote $G_{\rho,\theta}(P_0) = g_\rho(h_\theta(P_0))$

Preliminary results:

- **Gradient.** A closed-form expression for the gradient is:

$$\nabla g_\rho(m_\theta) = w \tag{2.B.15}$$

  where for each element $j = 1, ..., p$:

$$w_j = \frac{e^{-\rho m_{\theta,j}}\left(1 - \rho m_{\theta,j} + \rho g_\rho(m_\theta)\right)}{1 + \sum_{j=1}^{p} e^{-\rho m_{\theta,j}}},$$

  In the case of games, we have:

$$\begin{aligned}\nabla G_{\theta,\rho}(P_0) &= \frac{\partial h_\theta(P_0)}{\partial P} \nabla g_\rho(h_\theta(P_0)) \\ &= -Q^\top \nabla g_\rho(h_\theta(P_0)) \\ &= -\sum_{j=1}^{p} w_j q_j\end{aligned}$$

- **Hessian.** Let $H_\rho(m_\theta) = \nabla^\top \nabla g_\rho(m_\theta)$ the Hessian. We have for any element $h_{j,k}$ of H_ρ:

$$h_{j,k} = \rho\left(\frac{e^{-\rho m_{\theta,j}} w_k + e^{-\rho m_{\theta,k}} w_j}{1 + \sum_{j'=1}^{p} e^{-\rho m_{\theta,j'}}} - \mathbf{1}\{j = k\}\left(w_j + \frac{e^{-\rho m_{\theta,k}}}{1 + \sum_{j'=1}^{p} e^{-\rho m_{\theta,j'}}}\right)\right) \tag{2.B.16}$$

In matricial form:

$$H_\rho(m_\theta) = \rho\frac{1}{1+\sum_{j'=1}^{p} e^{-\rho m_{\theta,j'}}}\left(e^{-\rho m_\theta}w^T + we^{-\rho m_\theta T} - \left(\text{diag}(w)(1+\sum_{j'=1}^{p} e^{-\rho m_{\theta,j'}}) + \text{diag}(e^{-\rho m_\theta})\right)\right) \tag{2.B.17}$$

One can easily check that the Hessian is symmetric. In the case of games, we have $H_{\rho,\theta}(P_0) = \nabla^\top \nabla G_{\theta,\rho}(P_0)$ that writes as follows:

$$\begin{aligned} H_{\rho,\theta}(P_0) &= \frac{\partial}{\partial P} G_{\theta,\rho}(P_0)^\top \\ &= -\frac{\partial h_\theta(P_0)}{\partial P}\nabla^\top \nabla g_\rho(m_\theta)Q \\ &= Q^\top H_\rho(m_\theta)Q \end{aligned}$$

- **Limits population moments**: we now derive useful convergence results with population moments that we will use later on. First, let us define $\gamma = min\{m_{\theta,1}, ..., m_{\theta,2}\}$. We define

$$\mathcal{J}(\theta) = \{j \in \{1, \ldots, p\} \, | m_{\theta,j} = min\{0, \gamma\}\}$$

and $J_0 = card(\mathcal{J}(\theta))$. Finally, let $\mathbf{1_p} = (1, ..., p)'$ the vector of indices. Then, we have the following results:

$$\lim_{\rho\to\infty} \nabla g_\rho(m_\theta) \equiv \Gamma_0(\theta) = \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{1\{\gamma \geq 0\} + J_0}.$$

In the case of games:

$$\lim_{\rho\to\infty} \nabla G_{\theta,\rho}(P_0) = -Q^\top \Gamma_0(\theta).$$

Likewise, we can show that:

$$\lim_{\rho\to\infty} \frac{H_\rho(m_\theta)}{\rho} = H_0(\theta) \equiv 2\Gamma_0(\theta)\Gamma_0(\theta)^\top - 2diag(\Gamma_0(\theta))$$

This matrix is diagonal dominant with all of its diagonal term negative and thus it is negative definite. Finally, we have:

$$\lim_{\rho\to\infty} \frac{H_{\theta,\rho}(P_0)}{\rho} = Q^\top H_0(\theta)Q$$

The proofs are straightforward and thus omitted.

- **Limits empirical moments**: we now derive useful convergence results with empirical moments that we will use later on. Let us assume that $m_{n,\theta}$ is a $\sqrt{n}$-consistent estimator of m_θ as stated in Assumption 4.

We further assume that ρ_n is such that $\rho_n = cn^\alpha$ with $0 < \alpha < \frac{1}{2}$.

Proposition 2.1 *If we assume that all the moments are finite, then we have the following results:*

- $\|\nabla g_{\rho_n}(m_{\theta,n}) - \Gamma_0(\theta)\| \overset{P}{\to} 0$
- $\|\frac{H_{\rho_n}(m_{\theta,n})}{\rho_n} - H_0(\theta)\| \overset{P}{\to} 0$

Proof. • First consider the case where $\gamma \geq 0$ (i.e. $\theta \notin \Theta_I$).

First, we want to show that $e^{-\rho_n m_{\theta,n}} \overset{P}{\to} \{\mathbf{1_p} \in \mathcal{J}(\theta)\}$

$$\forall j, \quad e^{-\rho_n m_{\theta,n,j}} = e^{-\rho_n (m_{\theta,j} + O_p\left(\frac{1}{\sqrt{n}}\right))} = e^{-\rho_n m_{\theta,j}} \left(1 + O_p\left(\frac{\rho_n}{\sqrt{n}}\right)\right)$$
$$= \begin{cases} 1 + O_p\left(\frac{\rho_n}{\sqrt{n}}\right) \text{ if } m_{\theta,j} = 0 \\ O_p(\frac{1}{n^b}) \text{ for any } b > 0 \text{ if } m_{\theta,j} > 0 \end{cases}$$
$$\overset{P}{\to} \begin{cases} 1 \text{ if } m_{\theta,j} = 0 \\ 0 \text{ if } m_{\theta,j} > 0 \end{cases}$$

Likewise, let us show that for any j, $\rho_n m_{\theta,n,j} e^{-\rho_n m_{\theta,n,j}} \overset{P}{\to} 0$. We have that:

$$\rho_n m_{\theta,n,j} = \rho_n m_{\theta,j} + O_p\left(\frac{\rho_n}{\sqrt{n}}\right) = \begin{cases} O_p\left(\frac{\rho_n}{\sqrt{n}}\right) \text{ if } m_{\theta,j} = 0 \\ O_p\left(\rho_n\right) \text{ if } m_{\theta,j} > 0 \end{cases}$$

From what precedes:

$$e^{-\rho_n m_{j,n,\theta}} = \begin{cases} 1 + O_p\left(\frac{\rho_n}{\sqrt{n}}\right) \text{ if } m_{\theta,j} = 0 \\ O_p(\frac{1}{n^b}) \text{ for any } b > 0 \text{ if } m_{\theta,j} > 0 \end{cases}$$

Consequently, we have:

$$\rho_n m_{j,n,\theta} e^{-\rho_n m_{\theta,j,n}} = \begin{cases} O_p\left(\frac{\rho_n}{\sqrt{n}}\right) \text{ if } m_{\theta,j} = 0 \\ O_p(\frac{1}{n^b}) \text{ for any } b > 0 \text{ if } m_{\theta,j} > 0 \end{cases}$$

Let $Y_n = e^{-\rho_n m_{\theta,n}}$ and $Z_n = \rho_n m_{\theta,n,j} e^{-\rho_n m_{\theta,n,j}}$. By a simple application of the continuous mapping theorem to the random variables Y_n and Z_n we can show the following results:

(i) $\sum_{j'=1}^{p} e^{-\rho m_{\theta,n,j'}} \xrightarrow{P} J_0$

(ii) $\frac{e^{-\rho m_{\theta,n}}}{1+\sum_{j'=1}^{p} e^{-\rho m_{\theta,n,j'}}} \xrightarrow{P} \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{1+J_0}$

(iii) $\forall j,\ \rho_n m_{\theta,n,j} \frac{e^{-\rho m_{\theta,n,j}}}{1+\sum_{j'=1}^{p} e^{-\rho m_{\theta,n,j'}}} \xrightarrow{P} 0$

(iv) $\rho_n g_{\rho_n}(m_{\theta,n}) = \frac{\sum_{j=1}^{p} \rho_n m_{\theta,n,j} e^{-\rho m_{\theta,n,j}}}{1+\sum_{j'=1}^{p} e^{-\rho m_{\theta,n,j'}}} \xrightarrow{P} 0$

Therefore, by combining these different results, we have:

(v) $w_j(m_{\theta,n}) \xrightarrow{P} \frac{1\{j \in \mathcal{J}(\theta)\}}{1+J_0}$

(vi) $\nabla g_{\rho_n}(m_{\theta,n}) \xrightarrow{P} \Gamma_0(\theta)$

(vii) $\frac{H_{\rho_n}(m_{\theta,n})}{\rho_n} \xrightarrow{P} H_0(\theta)$

- Second, we consider the case where $\gamma < 0$ (i.e. $\theta \notin \Theta_I$).

  Let us show that:

$$g_{\rho_n}(m_{n,\theta}) \xrightarrow{P} \gamma$$

$$
\begin{aligned}
g_{\rho_n}(m_{\theta,n}) &= \frac{\sum_{j=1}^{p} m_{\theta,n,j} e^{-\rho_n m_{\theta,n,j}}}{1 + \sum_{j=1}^{p} e^{-\rho_n m_{\theta,n,j}}} \\
&= \frac{\sum_{j=1}^{p} m_{\theta,j} e^{-\rho_n m_{\theta,j}} (1 + O_p(\frac{1}{\sqrt{n}}))}{1 + \sum_{j=1}^{p} e^{-\rho_n m_{\theta,j}} (1 + O_p(\frac{1}{\sqrt{n}}))} \\
&= \frac{\sum_{j=1}^{p} m_{\theta,j} e^{-\rho_n (m_{\theta,j} - \gamma)} (1 + O_p(\frac{1}{\sqrt{n}}))}{e^{\rho_n \gamma} + \sum_{j=1}^{p} e^{-\rho_n (m_{\theta,j} - \gamma)} (1 + O_p(\frac{1}{\sqrt{n}}))} \\
&= \frac{o_p(1) + \sum_{j \in \mathcal{J}(\theta)} \gamma (1 + O_p(\frac{1}{\sqrt{n}}))}{o_p(1) + \sum_{j \in \bar{\mathcal{J}}(\theta)} (1 + O_p(\frac{1}{\sqrt{n}}))} \\
&\xrightarrow{P} \gamma
\end{aligned}
$$

The third line comes from multiplying the numerator and the denominator by $e^{\rho_n \gamma}$. Fourth line comes from the fact that for $j \notin \mathcal{J}(\theta), (m_{\theta,j} - \gamma) > 0$.

By similar arguments: we can show that:

(i) $\frac{e^{-\rho_n m_{\theta,n}}}{1+\sum_{j=1}^{p} e^{-\rho_n m_{\theta,n,j}}} \xrightarrow{P} \Gamma_0(\theta) = \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{J_0}$

(ii) $\nabla g_{\rho_n}(m_{\theta,n}) \xrightarrow{P} \Gamma_0(\theta) = \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{J_0}$

(iii) $\frac{H_{\rho_n}(m_{\theta,n})}{\rho_n} \xrightarrow{P} H_0(\theta)$

□

Proof of proposition 4.2

(i) First proposition is straightforward:

Following Equation (2.4.9), if $\theta \in \Theta_I$, for any $j = 1, \dots, p$,

$$m_{\theta,j} = \mathbb{E} m_j(Y, \theta) \geq 0.$$

Therefore, $g_\rho(m_\theta) = \frac{\sum_{j=1}^{p} m_{\theta,j} e^{-\rho m_{\theta,j}}}{1+\sum_{j=1}^{p} e^{-\rho m_{\theta,j}}}$ that is a weighted sum of the components of m_θ is necessarily greater or equal than 0.

(ii) TBD

■

Proof of Proposition 4.4

Proof. For all $n \in \mathbb{N}^*$, g_{ρ_n} is infinitely differentiable. Thus, by the mean value Taylor expansion,$\forall n \in \mathbb{N}^*, \exists \tilde{m}_{\theta,n} \in [m_{\theta,n}, m_\theta]$ such that the following expansion holds:

$$\begin{aligned}\sqrt{n} g_{\rho_n}(m_{\theta,n}) &= \sqrt{n} g_{\rho_n}(m_\theta) + \sqrt{n}(g_{\rho_n}(m_{\theta,n}) - g_{\rho_n}(m_\theta)) \\ &= \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{A_n} + \underbrace{\nabla g_{\rho_n}(m_\theta)\sqrt{n}(m_{\theta,n} - m_\theta)}_{B_n} + \underbrace{\frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n} - m_\theta)}_{C_n}\end{aligned}$$

we have:

- For the term B_n we have:

$$\begin{aligned}B_n &= \nabla g_{\rho_n}(m_\theta)^\top \sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \Gamma_0(\theta)^\top \sqrt{n}(m_{\theta,n} - m_\theta) + (\nabla g_{\rho_n}(m_\theta) - \Gamma_0(\theta))^\top \sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \Gamma_0(\theta)^\top \sqrt{n}(m_{\theta,n} - m_0) + o(1)O_p(1)\end{aligned}$$

  Third line comes from Assumption 4 and the simple convergence of $\nabla g_{\rho_n}(m_\theta)$

- Finally, for the term C_n, we have:

$$\begin{aligned}C_n &= \frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \frac{\rho_n}{\sqrt{n}}\left(\sqrt{n}(m_{\theta,n} - m_\theta)^\top H_0(\theta)\sqrt{n}(m_{\theta,n} - m_\theta) + \sqrt{n}(m_{\theta,n} - m_\theta)^\top (\frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n} - H_0(\theta))\sqrt{n}(m_{\theta,n} - m_\theta)\right) \\ &= \frac{\rho_n}{\sqrt{n}} O_p(1) + \frac{\rho_n}{\sqrt{n}} o_p(1) = O_p(\frac{\rho_n}{\sqrt{n}}) = o_p(1)\end{aligned}$$

  Third line is implied by the fact that $\tilde{m}_{\theta,n} = m_{\theta,n} + O_p(1/\sqrt{n})$. Thus, from Proposition 2.1, we have: $\frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n} - H_0(\theta) = o_p(1)$. Moreover, by assumption $\sqrt{n}(m_{\theta,n} - m_\theta) = O_p(1)$ and $H_0(\theta)$ is clearly bounded. This shows the result.

□

Proof of Proposition 4.3

Proof. From Proposition 4.4

$$\sqrt{n}g_{\rho_n}(m_{\theta,n}) = \sqrt{n}g_{\rho_n}(m_\theta) + \Gamma_0(\theta)\sqrt{n}(m_{\theta,n} - m_\theta) + o_p(1)$$

Let us consider the three different cases

(i) $\theta \in \partial\Theta_I$.

$\sqrt{n}g_{\rho_n}(m_\theta) = \sqrt{n}\frac{\sum_{j=1}^{p} m_\theta \exp(-\rho_n m_\theta)}{1+\sum_{j=1}^{p}\exp(-\rho_n m_\theta)} = \sqrt{n}\frac{\sum_{j\notin\mathcal{J}(\theta)} m_\theta \exp(-\rho_n m_\theta)}{1+\sum_{j=1}^{p}\exp(-\rho_n m_\theta)} \underset{n\to+\infty}{\to} 0$

because for $j \notin \mathcal{J}(\theta)$, $m_j(\theta) > 0$.

Thus, directly from the CLT and the continuous mapping theorem, we have:

$$\sqrt{n}g_{\rho_n}(m_{\theta,n}) \xrightarrow{d} \Gamma_0(\theta)^\top \mathcal{N}(0,\Sigma_0) = \mathcal{N}(0,\Gamma_0(\theta)^\top\Sigma_0\Gamma_0)$$

Because $\theta \in \partial\Theta_I$, $\Gamma_0(\theta) \neq 0$ and thus $\Gamma_0(\theta)^\top\Sigma_0\Gamma_0 > 0$.

From Proposition 2.1: $g_{\rho_n}(m_{\theta,n}) \xrightarrow{P} \Gamma_0(\theta)$. By assumption, $\Sigma_n \xrightarrow{P} \Sigma_0$

Therefore, by the CLT, $\nabla g_{\rho_n}(m_{\theta,n})^T\Sigma_n\nabla g_{\rho_n}(m_{\theta,n}) \xrightarrow{P} \Gamma_0(\theta)^\top\Sigma_0\Gamma_0$ and we can apply Slutsky's lemma to conclude.

(ii) $\theta \in int(\Theta_I)$. In this case, we can rewrite $\xi_n(\theta)$ as follows:

$$\begin{aligned}
\xi_n(\theta) &= \sqrt{n}\frac{g_{\rho_n}(m_{\theta,n})}{\sqrt{\nabla g_{\rho_n}(m_{\theta,n})^T\Sigma_n\nabla g_{\rho_n}(m_{\theta,n})}} \\
&= \frac{\sum_{j=1}^{p} m_{\theta,n,j}e^{-\rho m_{\theta,n,j}}}{1+\sum_{j=1}^{p} e^{-\rho_n m_{\theta,n,j}}} \Big/ \frac{\sqrt{\bar{\nabla} g_{\rho_n}(m_{\theta,n})^T\Sigma_n\bar{\nabla} g_{\rho_n}(m_{\theta,n})}}{1+\sum_{j=1}^{p} e^{-\rho_n m_{\theta,n,j}}} \\
&= \sqrt{n}\frac{\sum_{j=1}^{p} m_{\theta,n,j}e^{-\rho_n m_{\theta,n,j}}}{\sqrt{\bar{\nabla} g_{\rho_n}(m_{\theta,n})^T\Sigma_n\bar{\nabla} g_{\rho_n}(m_{\theta,n})}}
\end{aligned}$$

with $\bar{\nabla} g_{\rho_n}(m_{\theta,n})_j \equiv e^{-\rho m_{\theta,n,j}}\left(1-\rho_n m_{\theta,n,j} + \rho_n g_{\rho_n}(m_{\theta,n})\right)$ the numerator of $\nabla g_{\rho_n}(m_{\theta,n})$. We define $c = \underset{j}{min}\ m_{\theta,j}$ and $\tilde{\mathcal{J}}(\theta) = \left\{ j \in \underset{j}{argmin}\ m_{\theta,j} \right\}$ and $\tilde{J}_0 = card(\tilde{\mathcal{J}}(\theta))$. For the exposition, let us assume that $\tilde{J}_0 = 1$ with j^* the minimum (we can show the result in the general case). By standard arguments, we can show that:

$$\begin{aligned}\xi_n(\theta) &= \sqrt{n}\frac{ce^{-\rho_n c}(1+o_p(1))}{\sqrt{e^{-2\rho_n c}c^2\rho_n^2\sigma_{0j}^2(1+o_p(1))}} \\ &= \frac{\sqrt{n}}{\rho_n}\sigma_{0j^*}^{-1}\frac{(1+o_p(1))}{\sqrt{(1+o_p(1))}}\end{aligned}$$

Consequently, $\frac{\rho_n}{\sqrt{n}}\xi_n(\theta) \xrightarrow{P} \sigma_{0j^*}^{-1}$. Therefore,

$$Pr(\xi_n(\theta) > z_\alpha) = Pr(\frac{\rho_n}{\sqrt{n}}\xi_n(\theta) - \frac{\rho_n}{\sqrt{n}}z_\alpha > 0) \underset{n\to+\infty}{\to} 1$$

Last line uses the fact that by assumption we have $\frac{\rho_n}{\sqrt{n}} \underset{n\to+\infty}{\to} 0$.

(iii) $\theta \notin \Theta_I \implies \gamma < 0$. If the moment conditions satisfy the CLT of Assumption 4, then we have from Proposition 2.1:

- $g_{\rho_n}(m_{n,\theta}) \xrightarrow{P} \gamma$
- $\nabla g_{\rho_n}(m_{\theta,n}) \xrightarrow{P} \Gamma_0(\theta) = \frac{\{\mathbf{1}_{\mathbf{p}} \in \mathcal{J}(\theta)\}}{J_0}$

Therefore, by the continuous mapping theorem, $\frac{\xi_n(\theta)}{\sqrt{n}} = \frac{g_{\rho_n}(m_{\theta,n})}{\nabla g_{\rho_n}(m_{\theta,n})^T\Sigma_n\nabla g_{\rho_n}(m_{\theta,n})} \xrightarrow{P} \frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)} < 0$

$$\begin{aligned}\Pr\left(\xi_n(\theta) \geq z_\alpha\right) &= \Pr\left(\frac{\xi_n(\theta)}{\sqrt{n}} - \frac{z_\alpha}{\sqrt{n}} - \frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)} \geq -\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right) \\ &\leq \Pr\left(\left|\frac{\xi_n(\theta)}{\sqrt{n}} - \frac{z_\alpha}{\sqrt{n}} - \frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right| \geq -\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right) \underset{n\to\infty}{\to} 0\end{aligned}$$

Last line comes from the fact that $\frac{g_{\rho_n}(m_{\theta,n})}{\nabla g_{\rho_n}(m_{\theta,n})^T\Sigma_n\nabla g_{\rho_n}(m_{\theta,n})} \xrightarrow{P} \frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}$

□

Important remark: The proof above indicates that the asymptotic distribution exhibited in Proposition 4.3 remains valid if one replaces $g_{\rho_n}(m_{\theta,n})$ in $\xi_n(\theta)$ by any consistent estimator of $\Gamma_0(\theta)$. For instance, one could use: $\frac{\exp(-\rho_n m_{\theta,n})}{1+\sum_{j=1} p \exp(-\rho_n m_{\theta,n,j})}$ instead.

Proof Asymptotic distribution when the moments are standardized

Remember that the standardized version of the test statistic writes:

$$\tilde{\tilde{\xi}}_n(\theta) = \sqrt{n}\frac{g_{\rho_n}(\tilde{m}_{\theta,n})}{\nabla g_{\rho_n}(\tilde{m}_{\theta,n})^T \Omega_n \nabla g_{\rho_n}(\tilde{m}_{\theta,n})}$$

With Ω_n a consistent estimator of the correlation matrix $\Omega_0 = diag(\Sigma_0)^{-1/2}\Sigma_0 diag(\Sigma_0)^{-1/2}$.

Proposition 2.2 *Let ρ_n a divergent sequence of positive numbers such that $\rho_n = O(n^\alpha)$, $0 < \alpha < 1/2$. Let us further assume that $\forall j, \sigma_{j,\theta} > 0$. Then there are 3 different cases:*

- $\theta \in int(\Theta_I)$ *($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0):*

$$Pr(\tilde{\tilde{\xi}}_n(\theta) > z_\alpha) \underset{n\to+\infty}{\to} 1$$

- $\theta \in \partial\Theta_I$ *($\theta \in \Theta_I$ and $J_0 = card(\mathcal{J}(\theta)) > 0$, i.e. the number of binding moments is larger than 0). Then our test statistic is asymptotically distributed as a standard normal:*

$$\tilde{\tilde{\xi}}_n(\theta) \xrightarrow{d} \mathcal{N}(0,1)$$

- $\theta \notin \Theta_I$: $\quad Pr(\tilde{\tilde{\xi}}_n(\theta) > z_\alpha) \underset{n\to+\infty}{\to} 0$

Proof. The proof is very similar to the one in the non-normalized case. Under assumption **??** and by the continuous mapping theorem (exploiting the fact that $\sigma_{\theta,j}$ is positive so $Diag(\Sigma_0)^{-1/2}$ is well defined), the following holds:

$$diag(\Sigma_n)^{-1/2}\sqrt{n}(m_{\theta,n} - m_\theta) \xrightarrow{d} \mathcal{N}(0, diag(\Sigma_0)^{-1/2}\Sigma_0 diag(\Sigma_0)^{-1/2}) = \mathcal{N}(0,\Omega_0)$$

Now let us consider the 3 different cases:

- $\theta \in \partial\Theta_I$: By the same arguments as in the proof of Proposition 4.4

$$\sqrt{n} g_{\rho_n}(\tilde{m}_{\theta,n}) = \Gamma_0(\theta) diag(\Sigma_n)^{-1/2} \sqrt{n}(m_{\theta,n} - m_\theta) + o_p(1)$$

  Thus, we have that

$$\sqrt{n} g_{\rho_n}(\tilde{m}_{\theta,n}) \xrightarrow{d} \Gamma_0(\theta)^\top \mathcal{N}(0, \Omega_0)$$

  By standard arguments, we have:

$$\sqrt{n} \frac{g_{\rho_n}(\tilde{m}_{\theta,n})}{\nabla g_{\rho_n}(\tilde{m}_{\theta,n})^T \Omega_n \nabla g_{\rho_n}(\tilde{m}_{\theta,n})} \xrightarrow{d} \mathcal{N}(0,1)$$

- $\theta \in Int(\Theta_I)$: same argument as in Proposition 4.3

- $\theta \notin \Theta_I$: same argument as in Proposition 4.3

$\square$

Proof of Proposition 4.5

We have that the finite sample bias is equal to:

$$\mathbb{E}\left[\sqrt{n} g_{\rho_n}(m_{\theta,n})\right] = \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{\text{Identification bias}} + \underbrace{\frac{\rho_n}{\sqrt{n}} \mathbb{E}\left[\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n} \sqrt{n}(m_{\theta,n} - m_\theta)\right]}_{\text{Bias implied by the Taylor expansion's rest}}$$

No let us find an upper bound on both terms:

- **identification bias:** under $H_0 : \theta \in \Theta_I$, the identification bias corresponds to the difference between the approximation of the minimum and the minimum: $\sqrt{n} g_{\rho_n}(m_\theta) = \sqrt{n}(g_{\rho_n}(m_\theta) -$

$min\{0, m_{\theta,1}, ..., m_{\theta,p}\}$., this bias is equal to:

$$\begin{aligned}
|\text{Identification bias}| = \sqrt{n} g_{\rho_n}(m_\theta) &= \frac{\sum_{j \notin \mathcal{J}(\theta)} m_{\theta,j} e^{-\rho m_j}}{1 + J_0 + \sum_{j \notin \mathcal{J}(\theta)} e^{-\rho m_j}} \\
&\leq \frac{\sum_{j \notin \mathcal{J}(\theta)} m_{\theta,j} e^{-\rho m_j}}{1 + J_0} \\
&\leq (p - J_0) \frac{\frac{1}{\rho_n} e^{-1}}{1 + J_0}
\end{aligned}$$

Third line comes from the fact that $xe^{-\rho_n x}$ is maximized at $x = \frac{1}{\rho_n}$.

- **Bias implied by the Taylor expansion's rest.** Now we want to provide an upper bound on the rest of the Taylor expansion.

  - First, let us study the asymptotic behavior of the rest. We can show that:

$$\frac{\rho_n}{\sqrt{n}} \sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n} \sqrt{n}(m_{\theta,n} - m_\theta) = \frac{\rho_n}{\sqrt{n}} \sqrt{n}(m_{\theta,n} - m_\theta)^\top H_0(\theta) \sqrt{n}(m_{\theta,n} - m_\theta) + o_p\left(\frac{\rho_n}{\sqrt{n}}\right)$$

We have to compute $\mathbb{E}\left[\frac{\rho_n}{\sqrt{n}} \sqrt{n}(m_{\theta,n} - m_\theta)^\top H_0(\theta) \sqrt{n}(m_{\theta,n} - m_\theta)\right]$. Given that $\sqrt{n}(m_{\theta,n} - m_\theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$, we have by the continuous mapping theorem:

$$\sqrt{n}(m_{\theta,n} - m_\theta)^\top H_0(\theta) \sqrt{n}(m_{\theta,n} - m_\theta) \xrightarrow{d} Z^T \Sigma_0^{1/2} H_0(\theta) \Sigma_0^{1/2} Z = \sum_{j=1}^{p} \lambda_j \chi^2$$

with $Z \sim \mathcal{N}(0, I_p)$ and $(\lambda_j, ..., \lambda_p)$ the eigen values of $\Sigma_0^{1/2} H_0(\theta) \Sigma_0^{1/2}$. As a consequence,

$$\mathbb{E}\left[\frac{\rho_n}{\sqrt{n}} \sqrt{n}(m_{\theta,n} - m_\theta)^\top H_0(\theta) \sqrt{n}(m_{\theta,n} - m_\theta)\right] = \frac{\rho_n}{\sqrt{n}} \sum_{j=1}^{p} \lambda_j + o_p\left(\frac{\rho_n}{\sqrt{n}}\right).$$

Let us observe that H_0 is negative definite and thus the eigen values of $\Sigma_0^{1/2} H_0(\theta) \Sigma_0^{1/2}$ will be negative.

– Second, let us try to find an upper bound on the bias. We have:

$$\begin{aligned}
&\left|\mathbb{E}\left(\frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n}-m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n}-m_\theta)\right)\right| \\
&= \left|\mathbb{E}\left(tr\left(\frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n}-m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n}-m_\theta)\right)\right)\right| \\
&\leq \frac{\rho_n}{\sqrt{n}}\mathbb{E}\left|tr\left(\sqrt{n}(m_{\theta,n}-m_\theta)\sqrt{n}(m_{\theta,n}-m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\right)\right| \\
&\leq \frac{\rho_n}{\sqrt{n}}\mathbb{E}\left|tr\left(\sqrt{n}(m_{\theta,n}-m_\theta)\sqrt{n}(m_{\theta,n}-m_\theta)^\top\right) tr\left(\frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\right)\right| \\
&\leq \frac{\rho_n}{\sqrt{n}}\sqrt{\mathbb{E}\, tr\left(\sqrt{n}(m_{\theta,n}-m_\theta)\sqrt{n}(m_{\theta,n}-m_\theta)^\top\right)^2 \mathbb{E}\, tr\left(\frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\right)^2}
\end{aligned}$$

.

Second line uses $tr(AB) = tr(BA)$. Third line uses properties on the trace: $trace(AB)^2 \leq tr(A^2)tr(B^2) \leq tr(A)^2 tr(B)^2$. Last line obtained from Cauchy Schwarz

– Third, observe that the bias can also be estimated by simulations methods

- **Choice of ρ_n** In the end, we obtain that an upper bound on the absolute value bias (identification bias+ Taylor rest) is of the from:

$$\text{Bias} \leq (p - J_0)\frac{\frac{1}{\rho_n}e^{-1}}{1+J_0} + \frac{\rho_n}{\sqrt{n}}K_0$$

with $K_0 > 0$ a constant that increases with the number of binding moments and the variance of the moments. Thus, we want to choose ρ_n to minimize this upper bound and we find:

$$\rho_n = n^{1/4}\sqrt{\frac{(p-J_0)e^{-1}}{(1+J_0)K_0}}$$

LSE as smoothing function

Preliminary results:

Before expanding the test-statistic $g_\rho(\cdot)$, we need a few lemmas to characterize the higher order derivatives of

$$g_\rho: \ m_\theta \mapsto g_\rho(m_\theta).$$

Lemma 1 *The function $g_\rho: \ m_\theta \mapsto g_\rho(m_\theta)$ is infinitely differentiable in any $m_\theta \in \mathbb{R}^p$. Furthermore,*

- *we have a close form expression for the gradient:*

$$\nabla g_\rho(m_\theta) = w, \tag{2.B.18}$$

in which

$$w_j = \frac{\exp(-\rho m_{\theta,j})}{1 + \sum_{j=1}^{p} \exp(-\rho m_{\theta,j})}.$$

- *as well as for the Hessian:*

$$H_\rho(m_\theta) = \nabla g_\rho(m_\theta) \nabla^\top = \rho \left(-diag(w) + ww^T \right). \tag{2.B.19}$$

The proof is straightforward and thus omitted. Observe that the Hessian in m_θ is equal to ρ times a bounded matrix (for instance an upper bound is simply $2I_p$.

we now derive useful convergence results that we will use later on. Assume that moments $m_{\theta,n}$ satisfy the CLT of Assumption 4.

First, let us define $\gamma = \min\{m_{\theta,1}, ..., m_{\theta,2}\}$. We define

$$\mathcal{J}(\theta) = \{j \in \{1, \ldots, p\} \ | m_{\theta,j} = min\{0, \gamma\}\}$$

and $J_0 = card(\mathcal{J}(\theta))$. Finally, let $\mathbf{1_p} = (1, ..., p)'$ the vector of indices.

Lemma 2 *The following results hold:*

$$\lim_{\rho\to\infty} \nabla g_\rho(m_\theta) = \Gamma_0(\theta) = \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{1\{\gamma \geq 0\} + J_0}$$

and assume that ρ_n is a diverging sequence of numbers such that $\rho_n = O(n^\alpha)$ with $\alpha \in]0, 1/2[$.

$$\left\|\nabla g_\rho(m_{\theta,n}) - \Gamma_0(\theta)\right\|_2 \xrightarrow[\rho\to\infty]{P} 0.$$

Proof Let's prove the first claim:

- If $\gamma > 0$, then for any j, $m_{\theta,j} > 0$

$$\left[\nabla g_\rho(m_\theta)\right]_j = \frac{\exp(-\rho m_{\theta,j})}{1 + \sum_{j=1}^p \exp(-\rho m_{\theta,j})} \xrightarrow[\rho\to\infty]{} 0$$

- If $\gamma \leq 0$, then:

$$\begin{aligned}\left[\nabla g_\rho(m_\theta)\right]_j &= \frac{\exp(-\rho m_{\theta,j})}{1 + \sum_{j=1}^p \exp(-\rho m_{\theta,j})} \\ &= \frac{\exp(-\rho(m_{\theta,j} - \gamma))}{\exp(\rho\gamma) + \sum_{j=1}^p \exp(-\rho(m_{\theta,j} - \gamma))} \xrightarrow[\rho\to\infty]{} \frac{\{\mathbf{1_p} \in \mathcal{J}(\theta)\}}{1\{\gamma = 0\} + J_0}\end{aligned}$$

For the second claim:

$$\left\|\nabla g_\rho(m_{\theta,n}) - \Gamma_0(\theta)\right\|_2 \leq \left\|\nabla g_\rho(m_{\theta,n}) - \nabla g_\rho(m_\theta)\right\|_2 + \left\|\nabla g_\rho(m_\theta) - \Gamma_0(\theta)\right\|_2, \tag{2.B.20}$$

$$= \left\|H_\rho(\tilde{m}_{\theta,n})(m_{\theta,n} - m_\theta)\right\|_2 + \left\|\nabla g_\rho(m_\theta) - \Gamma_0(\theta)\right\|_2, \tag{2.B.21}$$

in which $\tilde{m}_{\theta,n}$ is a point in the segment $[m_{\theta,n}, m_n]$.

Following, Lemma 1, the first term is bounded by

$$\left\|H_\rho(\tilde{m}_{\theta,n})(m_{\theta,n} - m_\theta)\right\|_2 \leq \frac{\rho_n}{\sqrt{n}} \bar{H} \sqrt{n}(m_{\theta,n} - m_\theta)$$

As the moment are by assumption asymptotically normal, we have:

$$\left\|H_\rho(\tilde{m}_{\theta,n})(m_{\theta,n} - m_\theta)\right\|_2 \leq \frac{\rho_n}{\sqrt{n}} \times O_p(1),$$

and it tends to 0 in probability given the speed of divergence of ρ_n.

The second term in (2.B.21) tends trivially to 0 in probability from the definition of $\Gamma_0(\theta)$. ■

Proof of proposition 1.1

Proof. For all $n \in \mathbb{N}^*$, g_{ρ_n} is infinitely differentiable. Thus, by the mean value Taylor expansion,$\forall n \in \mathbb{N}^*, \exists \tilde{m}_{\theta,n} \in [m_{\theta,n}, m_\theta]$ such that the following expansion holds:

$$\begin{aligned}\sqrt{n} g_{\rho_n}(m_{\theta,n}) &= \sqrt{n} g_{\rho_n}(m_\theta) + \sqrt{n}(g_{\rho_n}(m_{\theta,n}) - g_{\rho_n}(m_\theta)) \\ &= \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{A_n} + \underbrace{\nabla g_{\rho_n}(m_\theta)\sqrt{n}(m_{\theta,n} - m_0)}_{B_n} + \underbrace{\frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n} - m_\theta)}_{C_n}\end{aligned}$$

we have:

- For the term B_n we have:

$$\begin{aligned}B_n &= \nabla g_{\rho_n}(m_\theta)^\top \sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \Gamma_0(\theta)^\top \sqrt{n}(m_{\theta,n} - m_\theta) + (\nabla g_{\rho_n}(m_\theta) - \Gamma_0(\theta))^\top \sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \Gamma_0(\theta)^\top \sqrt{n}(m_{\theta,n} - m_\theta) + o(1)O_p(1)\end{aligned}$$

  Third line comes from Assumption 4 and the simple convergence of $\nabla g_{\rho_n}(m_\theta)$ shown in Lemma

- Finally, for the term C_n, we have:

$$\begin{aligned}C_n &= \frac{\rho_n}{\sqrt{n}}\sqrt{n}(m_{\theta,n} - m_\theta)^\top \frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}\sqrt{n}(m_{\theta,n} - m_\theta) \\ &= \frac{\rho_n}{\sqrt{n}} O_p(1) = o_p(1)\end{aligned}$$

  Second line is implied by the fact that $m_{\theta,n} - m_{\theta,n} = O_p(1/\sqrt{n})$ and $\frac{H_{\rho_n}(\tilde{m}_{\theta,n})}{\rho_n}$ is bounded. This shows the result.

□

Proof of proposition 1.2

Step 1

To establish this result, let us first show that the expansion in Proposition 1.1 can be rewritten as follows:

$$g_{\rho_n}(m_{\theta,n}) = \sqrt{n}\min\{\gamma,0\} - \sqrt{n}\log\left(\frac{1\{\gamma \geq 0\} + p^*}{1+p}\right) + \Gamma_0(\theta)^\top\sqrt{n}(m_{\theta,n} - m_\theta) + o_p(1) \tag{2.B.22}$$

with $p* = J_0$ the number of binding moments. Also remember that $\gamma = \min_{j=1,\ldots,p} m_\theta$. To show this, it is sufficient to start from the expansion exhibited in 4 and to show:

$$\sqrt{n}g_{\rho_n}(m_\theta) = \sqrt{n}\min\{\gamma,0\} - \sqrt{n}\log\left(\frac{1+p^*}{1+p}\right) + o_p(1)$$

First, let us assume that $\gamma \geq 0$, then

$$\begin{aligned}
g_{\rho_n}(m_\theta) &= \frac{-1}{\rho_n}\log\left(\frac{1+\sum_{j=1}^{p}\exp(-\rho_n m_{\theta,j})}{p+1}\right) \\
&= \frac{-1}{\rho_n}\log\left(\frac{1+\sum_{j=1}^{p}\exp(-\rho_n m_{\theta,j})}{p+1}\right) \\
&= \frac{-1}{\rho_n}\log\left(\frac{1+p^*}{1+p} + \frac{\sum_{j\notin\mathcal{J}(\theta)}\exp(-\rho_n m_{\theta,j})}{p+1}\right) \\
&= \frac{-1}{\rho_n}\log\left(\frac{1+p^*}{1+p}\right) - \frac{1}{\rho_n}\log\left(1 + \frac{\sum_{j\notin\mathcal{J}(\theta)}\exp(-\rho_n m_{\theta,j})}{1+p^*}\right)
\end{aligned}$$

We can show that the second term tends to zero at any polynomial rate. Take $\eta = \min_{j\notin\mathcal{J}(\theta)} m_{\theta,j}$. By definition of $\mathcal{J}(\theta)$, $\eta > 0$. Therefore,

$$\sum_{j\notin\mathcal{J}(\theta)}\exp(-\rho_n m_{\theta,j}) \leq p\exp(-\rho_n\eta).$$

The last term tends to 0 when n tends to infinity and, using $x - x^2/2 \leq \log(1+x) \leq x$ in a neighborhood of 0, we obtain:

$$\frac{W_n - W_n^2/2}{\rho_n} \leq \frac{1}{\rho_n} \log\left(1 + \frac{\sum_{j \notin \mathcal{J}(\theta)} \exp(-\rho_n m_{\theta,j})}{1+p^*}\right) \leq \frac{W_n}{\rho_n},$$

with $W_n = \frac{p}{p+1} \exp(-\rho_n \eta)$.

Therefore, for any $b > 0$,

$$\frac{1}{\rho_n} \log\left(1 + \frac{\sum_{j \notin \mathcal{J}(\theta)} \exp(-\rho_n m_{\theta,j})}{1+p^*}\right) = o(1/n^b),$$

Thus,

$$\sqrt{n} g_{\rho_n}(m_\theta) = -\sqrt{n} \log\left(\frac{1+p^*}{1+p}\right) + o_p(1)$$

Next, let us assume that $\gamma < 0$. Then,

$$\begin{aligned}
g_{\rho_n}(m_\theta) &= \frac{-1}{\rho_n} \log\left(\frac{1 + \sum_{j=1}^{p} \exp(-\rho_n m_{\theta,j})}{p+1}\right) \\
&= \gamma + \frac{-1}{\rho_n} \log\left(\frac{\exp(\rho_n \gamma) + \sum_{j=1}^{p} \exp(-\rho_n (m_{\theta,j} - \gamma))}{p+1}\right) \\
&= \gamma + \frac{-1}{\rho_n} \log\left(\frac{p^*}{1+p} + \frac{\exp(\rho_n \gamma) \sum_{j \notin \mathcal{J}(\theta)} \exp(-\rho_n (m_{\theta,j} - \gamma))}{p+1}\right) \\
&= \gamma + \frac{-1}{\rho_n} \log\left(\frac{p^*}{1+p}\right) - \frac{1}{\rho_n} \log\left(1 + \frac{\sum_{j \notin \mathcal{J}(\theta)} \exp(-\rho_n (m_{\theta,j} - \gamma))}{1+p^*}\right)
\end{aligned}$$

By the same argument as previously, the third term is $o(1/n^b)$ for any $b > 0$. Thus, we have the first result.

Step 2

Next, we show that the estimation of p^* does not change the expansion exhibited in (2.B.22). Remember the definition of $\widehat{p^*}_n$

$$\widehat{p^*}_n = \sum_{j=1}^{p} \mathbf{1}\left\{\frac{m_{\theta,j}}{\sqrt{\sigma_j^2}} < \tau_n\right\}.$$

Using the expansion, we can rewrite it with $\widehat{p^*}_n$ as follows

$$\sqrt{n}\left(g_{\rho_n}(m_\theta)+\frac{1}{\rho_n}\log\left(\frac{1+\widehat{p^*}_n}{1+p}\right)\right)+\frac{\sqrt{n}}{\rho_n}\log\left(\frac{1+p^*}{1+\widehat{p^*}_n}\right)=\sqrt{n}\min\{\gamma,0\}+\Gamma_0(\theta)^\top\sqrt{n}(m_{\theta,n}-m_\theta)+o_p($$

We need to show that the second term of the left hand side of the equality above tends to 0 in probability.

Observe that

$$\Pr(\frac{\sqrt{n}}{\rho_n}\log\left(\frac{1+p^*}{1+\widehat{p^*}_n}\right)\neq 0)=\Pr(\widehat{p^*}_n\neq p^*).$$

Then, we calculate the probability for a binding moment to not be selected:

$$\begin{aligned}
P_{ns} &= \Pr((Z/\sqrt{n}+o_P(1/\sqrt{n}))>\tau_n)\\
&= 1-\Phi(\sqrt{n}\tau_n+o_P(1))\\
&\leq K\frac{1}{\sqrt{n}\tau_n}\exp(-n\tau_n^2/2)\text{following classical result on the upper tail of the } N(0,1)
\end{aligned}$$

Similarly the probability for a non-binding moment (its expectation is denoted $\mu>0$) to be selected is equal to

$$\begin{aligned}
P_s &= P((Z/\sqrt{n}+\mu+o_P(1/\sqrt{n}))<\tau_n)\\
&= \Phi(\sqrt{n}(\tau_n-\mu)+o_P(1))
\end{aligned}$$

τ_n tending toward 0 and μ being positive, the quantity inside $\Phi(\cdot)$ tends to $-\infty$. The probability tends to 0 at any polynomial rate.

As a result, for any $b>0$,

$$n^b P_0(\widehat{p^*}_n\neq p^*)\xrightarrow[n\to\infty]{P}0.$$

Therefore, the following expansion holds:

$$\sqrt{n}\left(g_{\rho_n}(m_\theta)+\frac{1}{\rho_n}\log\left(\frac{1+\widehat{p^*}_n}{1+p}\right)\right)=\sqrt{n}\min\{\gamma,0\}+\Gamma_0(\theta)^\top\sqrt{n}(m_{\theta,n}-m_\theta)+o_p(1). \tag{2.B.23}$$

Step 3

• $\theta\in\partial\Theta_I$ ($\min\{\gamma,0\}=0$), Then from expansion 2.B.23 and by a direct application of the central limit theorem and the continuous mapping theorem:

$$\sqrt{n}\left(g_{\rho_n}(m_\theta)+\frac{1}{\rho_n}\log\left(\frac{1+\widehat{p^*}_n}{1+p}\right)\right)\xrightarrow{d}\Gamma_0(\theta)\mathcal{N}(0,\Sigma_0)$$

From lemma 2.B and the continuous mapping theorem, we have that $\nabla g_{\rho_n}(m_{\theta,n})^\top\Sigma_n\nabla g_{\rho_n}(m_{\theta,n})\xrightarrow{\mathbb{P}}\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)$

Thus, by Slutsky,

$$\tilde{\xi}_n(\theta)\xrightarrow{d}\mathcal{N}(0,1)$$

• $\theta\notin\Theta_I$. Then from expansion 2.B.23 and by standard arguments:

$$\frac{\tilde{\xi}_n(\theta)}{n}\xrightarrow{\mathbb{P}}\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}<0$$

Thus,

$$\begin{aligned}\Pr\left(\tilde{\xi}_n(\theta)\geq z_\alpha\right)&=\Pr\left(\frac{\tilde{\xi}_n(\theta)}{\sqrt{n}}-\frac{z_\alpha}{\sqrt{n}}-\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\geq-\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right)\\&\leq\Pr\left(\left|\frac{\tilde{\xi}_n(\theta)}{\sqrt{n}}-\frac{z_\alpha}{\sqrt{n}}-\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right|\geq-\frac{\gamma}{\Gamma_0(\theta)^\top\Sigma_0\Gamma_0(\theta)}\right)\underset{n\to\infty}{\rightarrow}0\end{aligned}$$

• $\theta\in Int(\Theta_I)$ (tbd)

■

Inference with covariates

Proof of proposition 4.6

To show this, we just need to prove that the graph generated by $\Gamma(\theta, X)$ remains the same for all $X \in \mathcal{X}$. The graph $\Gamma(\theta, X)$ is such that there exists an edge between two elements y_1 and y_2 if their equilibrium regions $\mathcal{R}_\omega(X, y_1)$ and $\mathcal{R}_\omega(X, y_2)$ overlap. Now from proposition 3.2, we know that this is the case if

$\forall t \in \mathcal{T}$, such that $0 \leq N_t, \bar{N}_t \leq a$:

$$\begin{cases} -\pi_t(X, N_t, \mathbf{N}_{-\mathbf{t}}; \omega) < -\pi_t(X, \bar{N}_t + 1, \bar{\mathbf{N}}_{-t}; \omega) \\ -\pi_t(X, \bar{N}_t, \bar{\mathbf{N}}_{-t}; \omega) < -\pi_t(X, N_t + 1, \mathbf{N}_{-t}; \omega). \end{cases}$$

From the decomposition in Assumption 5. The previous conditions become:

$$\begin{cases} -\kappa_t(N_t, \mathbf{N}_{-\mathbf{t}}; \omega_2) < -\kappa_t(\bar{N}_t + 1, \bar{\mathbf{N}}_{-t}; \omega_2) \\ -\kappa_t(\bar{N}_t, \bar{\mathbf{N}}_{-t}; \omega_2) < -\kappa_t(N_t + 1, \mathbf{N}_{-t}; \omega_2). \end{cases}$$

These conditions do not depend on X. ■

The smoothed min approach

For the moment, we prove Propositions 4.8 and 4.9 in the special case related to entry games, where $m_{\theta,j}(X) = C_{\theta,j}(X) - q_j^T \mathbb{E}[\mathbf{1}(Y = y)|X]$. Moreover, let us define: $G_{\theta,\rho}(x, h_0) = g_\rho(C_\theta(X) - Q\mathbb{E}[\mathbf{1}(Y = y)|X = x])$. We are working on extending the proof to the more general case.

Let us consider the kernel estimator of $h_0(X) = \mathbb{E}[\mathbf{1}(Y = y)|X]$,

$$\hat{h}_0(x) = \frac{\sum_{i=1}^n K_\sigma(x - X_i)\mathbf{1}\{Y_i = y\}}{\sum_{i=1}^n K_\sigma(x - X_i)},$$

where σ is the bandwidth and $K_\sigma(x) = \frac{1}{\sigma^d} K\left(\frac{x}{\sigma}\right)$. Lets also define $\hat{f}_0(x) = \frac{1}{n}\sum_{i=1}^n K_\sigma(x - X_i)$ and $\hat{w}_0(x) = \frac{1}{n}\sum_{i=1}^n K_\sigma(x - X_i)\mathbf{1}\{Y_i = y\}$. Under assumption 6, we have

$$\sqrt{n}||\hat{h}_0 - h_0||^2 \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sqrt{n}||\hat{f}_0 - f_0||^2 \xrightarrow{\mathbb{P}} 0,$$

where $\|\cdot\|$ is a Sobolev norm, for non-negative integer k, defined as

$$\|f\| = \max_{j\leq k}\max_{x\in\mathcal{X}}\left\|\frac{\partial^j f(x)}{\partial x^j}\right\|.$$

Then, from Newey and McFadden (1994) Lemma 8.9, we have

$$\|f_0 - \mathbb{E}(\hat{f}_0)\| = O\left(\sigma^m\right),$$

if $m + k < \alpha$. Similar result holds for $\hat{h}_0$. For the variance term, following Newey and McFadden (1994), we have

$$\|\hat{f}_0 - \mathbb{E}(\hat{f}_0)\| = O_P\left[\left(\frac{\ln n}{n\sigma^{d+2k}}\right)^{\frac{1}{2}}\right].$$

Similar result holds for $\hat{w}_0$.

Lemma 3 *Under assumption 6, we have*

$$\left\|\hat{h}_0 - \mathbb{E}(\hat{h}_0)\right\| = O_P\left[\left(\frac{\ln n}{n\sigma^{d+2k}}\right)^{\frac{1}{2}}\right].$$

Lemmas

The next two lemmas verify technical conditions for kernel estimator which are needed to derive asymptotic distribution.

Lemma 4 (Mean Square Differentiability) *Under assumption 6, we have uniformly over smoothing parameter $\rho > 0$*

$$\sqrt{n}\left[\int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x) - h_0(x))\mathrm{d}x - \frac{1}{n}\sum_{i=1}^{n}\alpha(W_i)\right] \xrightarrow{\mathbb{P}} 0.$$

Proof:

We have

$$\begin{aligned}
&\int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^\top (\hat{h}_0(x) - h_0(x))\mathrm{d}x \\
&\quad = \int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^\top \left(\frac{\sum_{i=1}^n K_\sigma(x - X_i)\mathbf{1}\{Y_i = y\}}{\sum_{i=1}^n K_\sigma(x - X_i)} - h_0(x) \right) \mathrm{d}F_0(x) \\
&\quad = \sum_{i=1}^n \int \frac{K_\sigma(x - X_i)}{\sum_{i=1}^n K_\sigma(x - X_i)} \underbrace{\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^\top (\mathbf{1}\{Y_i = y\} - h_0(x))}_{\alpha(w)} \mathrm{d}F_0(x) \\
&\quad = \int \alpha(w) d\tilde{F}(w) \\
&\quad = \frac{1}{n} \sum_{i=1}^n \int \frac{K_\sigma(x - X_i)}{\frac{1}{n}\sum_{i=1}^n K_\sigma(x - X_i)} \underbrace{\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^\top [-h_0(x) \;\; I]}_{\beta(x)} \underbrace{\begin{bmatrix} \mathbf{1} \\ \mathbf{1}\{Y_i = y\} \end{bmatrix}}_{Z_i} \mathrm{d}F_0(x),
\end{aligned}$$

where distribution $\tilde{F}$ is mix of empirical conditional measure and density f_0. The random variable X follows density $f_0(\cdot)$ and conditional on $X = x$, $\mathbf{1}\{Y_i = y\}$ is distributed with empirical measure with points mass $\frac{K_\sigma(x - X_i)}{\sum_{i=1}^n K_\sigma(x - X_i)}$. So, we just need to show that

$$\sqrt{n} \left[\int \alpha(w) d\tilde{F}(w) - \int \alpha(w) d\hat{F}(w) \right] \xrightarrow{\mathbb{P}} 0,$$

where $\hat{F}$ is just empirical distribution. We have

$$
\begin{aligned}
&\sqrt{n}\left[\int \alpha(w)d\tilde{F}(w) - \int \alpha(w)d\hat{F}(w)\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int \frac{K_\sigma(x-X_i)}{\frac{1}{n}\sum_{i=1}^{n}K_\sigma(x-X_i)}\beta(x)\mathrm{d}F_0(x) - \beta(X_i)\right]Z_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int \frac{K_\sigma(x-X_i)}{f_0(x)}\beta(x)f_0(x)\mathrm{d}x - \beta(X_i)\right]Z_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int \frac{K_\sigma(x-X_i)(f_0(x)-\hat{f}_0(x))}{\hat{f}_0(x)}\beta(x)\mathrm{d}x\right]Z_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int \frac{K_\sigma(x-X_i)}{f_0(x)}\beta(x)f_0(x)\mathrm{d}x - \beta(X_i)\right]Z_i \\
&\quad + \sqrt{n}\left[\int (f_0(x)-\hat{f}_0(x))\beta(x)\left\{\frac{\frac{1}{n}\sum_{i=1}^{n}K_\sigma(x-X_i)Z_i}{\hat{f}_0(x)} - \begin{bmatrix}\mathbf{1}\\ h_0(x)\end{bmatrix}\right\}\mathrm{d}x\right] \\
&\quad + \sqrt{n}\underbrace{\left[\int (f_0(x)-\hat{f}_0(x))\beta(x)\begin{bmatrix}\mathbf{1}\\ h_0(x)\end{bmatrix}\mathrm{d}x\right]}_{=0} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\int \frac{K_\sigma(x-X_i)}{f_0(x)}\beta(x)f_0(x)\mathrm{d}x - \beta(X_i)\right]Z_i \\
&\quad + \sqrt{n}\int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(f_0(x)-\hat{f}_0(x))(h_0(x)-\hat{h}_0(x))\mathrm{d}x.
\end{aligned}
$$

For the first term, we will use Chebyshev's inequality and show its expectation and variance converge to zero. Let $\gamma_0(x) = \mathbb{E}(Z_i|X = x)$, then

$$\begin{aligned}
&\left\| \sqrt{n}\mathbb{E}\left[\left\{ \int K_\sigma(x - X_i)\beta(x)\mathrm{d}x - \beta(X_i) \right\} Z_i \right] \right\| \\
&\quad = \sqrt{n} \left\| \int \left\{ \int \beta(x + \sigma u)K(u)\mathrm{d}u \right\} \gamma_0(x) f_0(x)\mathrm{d}x - \int \beta(x)\gamma_0(x) f_0(x)\mathrm{d}x \right\| \\
&\quad = \sqrt{n} \left\| \iint \beta(x)K(u)\gamma_0(x - \sigma u) f_0(x - \sigma u)\mathrm{d}u\, \mathrm{d}x - \int \beta(x)\gamma_0(x) f_0(x)\mathrm{d}x \right\| \\
&\quad = \sqrt{n} \left\| \int \beta(x) \left\{ \int \left[f_0(x - \sigma u)\gamma_0(x - \sigma u) - f_0(x)\gamma_0(x) \right] K(u)\mathrm{d}u \right\} \mathrm{d}x \right\| \\
&\quad \leqslant \sqrt{n} \int \|\beta(x)\| \left\| \int \left[f_0(x - \sigma u)\gamma_0(x - \sigma u) - f_0(x)\gamma_0(x) \right] K(u)\mathrm{d}u \right\| \mathrm{d}x \leqslant C\sqrt{n}\sigma^m \int \|\beta(x)\|\mathrm{d}x,
\end{aligned}$$

where last bound, for some constant C, follows from Taylor expansion of $f_0(x - \sigma u)$, $\gamma_0(x - \sigma u)$ and assumption (6)(1). Therefore, $\left\| \sqrt{n}\mathbb{E}\left[\left\{ \int \beta(x)K_\sigma(x - X_i)\,\mathrm{d}x - \beta(X_i) \right\} Z_i \right] \right\| \leqslant C\sqrt{n}\sigma^m \to 0$.

Also, by almost everywhere continuity of $\beta(x), \beta(x + \sigma u) \to \beta(x)$ for almost all x. Also, on the bounded support of $K(u)$, for small enough $\sigma, \beta(x + \sigma u) \leqslant \sup_{\|v\| \leqslant \varepsilon} \beta(x + v)$, so by the dominated convergence theorem, $\int \beta(x + \sigma u)K(u)\mathrm{d}u \to \int \beta(x)K(u)\mathrm{d}u = \beta(x)$ for almost all x. The boundedness of $K(u)$ and dominated convergence theorem gives

$$\mathbb{E}\left[\left\| \int \beta(x)K_\sigma(x - X_i)\,\mathrm{d}x - \beta(X_i) \right\|^4 \right] \to 0,$$

so by the CauchySchwartz inequality, $\mathbb{E}\left[\|Z_i\|^2 \left\| \int \beta(x)K_\sigma(x - X_i)\,\mathrm{d}x - \beta(X_i) \right\|^2 \right] \to 0$. Mean square differentiability condition follows from the Chebyshev inequality, since the mean and variance of $n^{-1/2}\sum_{i=1}^n \left[\int \beta(x)K_\sigma(x - X_i)\,\mathrm{d}x - \beta(X_i) \right] Z_i$ go to zero.

The second term in the expression is bounded, uniformly over ρ, by

$$\begin{aligned}
&\left|\sqrt{n}\int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(f_0(x)-\hat{f}_0(x))(h_0(x)-\hat{h}_0(x))\mathrm{d}x\right| \\
&\quad\leq \sqrt{n}\int \left|\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(f_0(x)-\hat{f}_0(x))(h_0(x)-\hat{h}_0(x))\right|\mathrm{d}x \\
&\quad\leq \sqrt{n}\|f_0(x)-\hat{f}_0(x)\|\int \left|\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(h_0(x)-\hat{h}_0(x))\right|\mathrm{d}x \\
&\quad\leq \sqrt{n}\|f_0(x)-\hat{f}_0(x)\|\int \left|\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(h_0(x)-\hat{h}_0(x))\right|\mathrm{d}x \\
&\quad\leq \sqrt{n}\|\hat{h}_0-h_0\|\|\hat{f}_0-f_0\|\int\sum_i \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h_i}dx \\
&\quad= |\mathcal{X}|\sqrt{n}\|\hat{h}_0-h_0\|\|\hat{f}_0-f_0\|.
\end{aligned}$$

$\sqrt{n}\|\hat{h}_0-h_0\|\|\hat{f}_0-f_0\|$, which converges to zero in probability. ■

Lemma 5 (Stochastic Equicontinuity) *Under assumption 6, we have uniformly over smoothing parameter $\rho > 0$*

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}(\hat{h}_0(X_i)-h_0(X_i)) - \int \frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}F_0(x)\right] \xrightarrow{\mathbb{P}} 0.$$

Proof: It is easy to see that

$$\begin{aligned}
\hat{h}_0(x)-h_0(x) &= \frac{\hat{w}_0(x)}{\hat{f}_0(x)} - h_0(x) \\
&= \frac{\hat{w}_0(x)-h_0(x)\hat{f}_0(x)}{\hat{f}_0(x)} \\
&= \frac{\hat{w}_0(x)-h_0(x)\hat{f}_0(x)}{\hat{f}_0(x)}\left[\frac{\hat{f}_0(x)}{f_0(x)} + 1 - \frac{\hat{f}_0(x)}{f_0(x)}\right] \\
&= \underbrace{\frac{\hat{w}_0(x)-h_0(x)\hat{f}_0(x)}{f_0(x)}}_{I} + \underbrace{\frac{(\hat{h}_0(x)-h_0(x))(\hat{f}_0(x)-f_0(x))}{f_0(x)}}_{II}.
\end{aligned}$$

Based on this, we divide the problem into two sub-problems I and II. For the sub-problem I, we need to show[20]

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}}{f_0(X_i)}(\hat{w}_0(X_i)-h_0(X_i)\hat{f}_0(X_i))-\int\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{w}_0(x)-h_0(x)\hat{f}_0(x))\mathrm{d}x\right]$$

$$=\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}}{f_0(X_i)}[-h_0(X_i)\ I]\begin{bmatrix}\hat{f}_0(X_i)\\ \hat{w}_0(X_i)\end{bmatrix}-\int\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}[-h_0(x)\ I]\begin{bmatrix}\hat{f}_0(x)\\ \hat{w}_0(x)\end{bmatrix}\mathrm{d}x\right]$$

$$=\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\beta(X_i)}{f_0(X_i)}\frac{1}{n}\sum_{j=1}^{n}K_\sigma(X_i-X_j)Z_j-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_X\left[\frac{\beta(X)}{f_0(X)}K_\sigma(X-X_j)Z_j\right]\right]$$

$$=\sqrt{n}\Bigg[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X_j)Z_j-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_X\left[\frac{\beta(X)}{f_0(X)}K_\sigma(X-X_j)Z_j\right]$$

$$-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{X,Z}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X)Z\right]+\mathbb{E}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X_j)Z_j\right]\Bigg]$$

$$+\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{X,Z}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X)Z\right]-\mathbb{E}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X_j)Z_j\right]\right]$$

We will apply Lemma 6 from Newey and McFadden (1994) to verify that first bracket converges to 0 in probability, uniformly over ρ. It is easy to see that uniformly over ρ, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i-X_i)Z_i\right\|\right]&=\mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}}{f_0(X_i)}[-h_0(X_i)\ I]K_\sigma(0)Z_i\right\|\right]\\&=K_\sigma(0)\mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}}{f_0(X_i)}[-h_0(X_i)\ I]\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\{Y_i=y\}\end{bmatrix}\right\|\right]\\&\le K_\sigma(0)\mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}}{f_0(X_i)}\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}\right\|\right]\\&\le CK_\sigma(0),\end{aligned}$$

[20]We used the notation $\beta(x)=\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}[-h_0(x)\ I]$ and $Z_i=\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\{Y_i=y\}\end{bmatrix}$

where second last bound uses the fact that $-1 \leq \mathbf{1}\{Y_i = y_k\} - h_{0,k}(X_i) \leq 1$ and $\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h} \geq 0$. Similarly, uniformly over ρ, we have

$$
\begin{aligned}
\mathbb{E}\left[\left\|\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i - X_j)Z_j\right\|^2\right] &= \mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top}{f_0(X_i)}[-h_0(X_i)\ I]K_\sigma(X_i - X_j)Z_j\right\|^2\right] \\
&\leq M\mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top}{f_0(X_i)}[-h_0(X_i)\ I]\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\{Y_j = y\}\end{bmatrix}\right\|^2\right] \\
&\leq M\mathbb{E}\left[\left\|\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top}{f_0(X_i)}\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}\right\|\right] \\
&\leq CM,
\end{aligned}
$$

where M is bound on the kernel.

Second big bracket will converge in probability to zero if

$$
\mathbb{E}\left[\left\|\mathbb{E}_{X,Z}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i - X)Z\right]\right\|^2\right] \to 0,
$$

by Chebyshev's inequality. It is easy to see that

$$
\begin{aligned}
\mathbb{E}_{X,Z}\left[\frac{\beta(X_i)}{f_0(X_i)}K_\sigma(X_i - X)Z\right] &= \mathbb{E}_X\left[\frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top}{f_0(X_i)}K_\sigma(X_i - X)\left[h_0(X) - h_0(X_i)\right]\right] \\
&= \frac{\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top}{f_0(X_i)}\left[\mathbb{E}_X(K_\sigma(X_i - X)h_0(X)) - h_0(X_i)\right],
\end{aligned}
$$

and clearly $\mathbb{E}_X(K_\sigma(X_i - X)h_0(X)) \to h_0(X_i)$ as σ converges to zero. ■

Let $m_{n1}(z) = \int m_n(z,\tilde{z})\mathrm{d}F_0(\tilde{z}), m_{n2}(z) = \int m_n(\tilde{z},z)\mathrm{d}F_0(\tilde{z})$, and $\mu = \int m_n(z,\tilde{z})\mathrm{d}F_0(\tilde{z})\mathrm{d}F_0(z)$. The following lemma is taken from Newey (1994).

Lemma 6 (V-Statistic Convergence) *If $Z_1, Z_2, \ldots$ are i.i.d. then*

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_n\left(Z_i, Z_j\right) - n^{-1}\sum_{i=1}^{n}\left[m_{n1}\left(Z_i\right) + m_{n2}\left(Z_i\right)\right] + \mu$$
$$= O_P\left\{\frac{\mathbb{E}\left[\left\|m_n\left(Z_1, Z_1\right)\right\|\right]}{n} + \frac{\left(\mathbb{E}\left[\left\|m_n\left(Z_1, Z_2\right)\right\|^2\right]\right)^{1/2}}{n}\right\}.$$

Proof of Proposition 4.8

We are interested in deriving the asymptotic distribution of

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho}(X_i, \hat{h}_0) - \mathbb{E}\left[G_{\theta,\rho}(X_i, h_0)\right]\right).$$

Our proof proceeds in two steps. First, we exploit results in Ackerberg, Chen, Hahn, and Liao (2014), which allow us to recover the form of the adjustment term due to a noisy estimate of h_0. Second, we build upon Newey (1994) to derive the asymptotic distribution.

Step 1: (Derivation of the adjustment term $\alpha(W_i)$). $W_i = (X_i, Y_i)$. In comparison to the more general case treated in Ackerberg, Chen, Hahn, and Liao (2014), the derivation of the adjustment term simplifies because $G_{\theta,\rho}(X_i, h_0)$ depends on h_0 only through the values it takes $h_0(X_i)$. For the sake of thoroughness, we follow and reproduce some developments and borrow some of their notations. For more details, we refer the interested reader to Ackerberg, Chen, Hahn, and Liao (2014). Under H_0,

$$\mathbb{E}[G_{\theta,\rho}(X_i, h_0)] \geq 0,$$

and $h_0 = (h_{0,1}, ..., h_{0,L})$ consist of $L = \text{card}(\mathcal{Y}) - 1$ nuisance functions, which are identified by the following conditional moment restrictions:

$$\mathbb{E}\left[\rho_l(W_i, h_{0,l}(X_i))|X_i\right] = 0 \; a.s \; X_i \quad \text{for } l = 1, ...L$$

where $\rho_l(W_i, h_{0,l}(X_i)) = 1\{Y_i = y_l\} - h_{0,l}(X_i)$. The last function is directly inferred from $\sum_{\mathcal{Y}} 1\{Y_i = y_l\} = 1$ almost surely. Every nuisance function h_l is assumed to lie in the space $\mathcal{H}_l$, which is a linear subspace of the space of integrable functions with respect to X. Finally, we define

$$k_l(X_i, h_{0,l}) = \mathbb{E}\left[\rho_l(W_i, h_{0,l}(X_i))|X_i\right].$$

One can easily show that the regularity conditions (p 922) of Ackerberg, Chen, Hahn, and Liao (2014) are satisfied. In our case, k depends only on h through $h(X_i)$. Thus, the pathwise derivative with respect to h_l evaluated at h_0 in the direction $v_l \in \mathcal{V}_l \equiv \mathcal{H}_l - \{h_{0,l}\}$

$$\frac{\partial k_l(X_i, \theta, h_{0,l})}{\partial h_l}[v_l] = \frac{\partial \mathbb{E}[\rho_l(W_i, h_{0,l}(X_i) + \tau v_l(X_i))}{\partial \tau}\bigg|_{\tau=0} = \frac{\mathbb{E}\left[\rho_l(W_i, h_{0,l}(X_i))|X_i\right]}{\partial h_l} v_l(X) = -v_l(X).$$

with the second derivative being the ordinary derivative. Following Ackerberg, Chen, Hahn, and Liao (2014) the Riesz representation theorem implies that there is a unique $u_l^* \in \mathcal{V}_l$ (set of measureable functions of X) such that $\forall v_l \in \mathcal{V}_l$,

$$\frac{\partial \mathbb{E}[G_{\theta,\rho}(X_i, h_0)]}{\partial h_l}[v_l] = \mathbb{E}\left[u_l^*(X_i) v_l(X_i)\right] \tag{2.B.24}$$

Now let us show that the adjustment term is equal to $\alpha(W_i) = \sum_{l=1}^{L} u_l^*(X_i)\rho_l(W_i, h_0(X_i))$. To do this, we follow the same reasoning as the one developed for the proof of proposition 1 in Newey (1994). Without loss of generality, we assume $L = 1$. As shown in Newey (1994), the adjustment term of the sum corresponds to the sum of the adjustment terms.

Consider a path $\{F_\tau(w)\}$ of the distribution of random variable W. Let h_τ be the function indexed by τ such that $\mathbb{E}_\tau\left[\rho(W_i, h_\tau(X_i))|X\right] = 0$ where $\mathbb{E}_\tau[.|X]$ denotes the conditional expectation taken under $F_\tau(w)$ with the corresponding score $S(w)$. From the definition of $u^*(X_i)$:

$$\frac{\partial}{\partial \tau}\mathbb{E}[G_{\theta,\rho}(X_i, h_\tau)] = \mathbb{E}\left[u^*(X_i)\frac{\partial}{\partial \tau}k(X_i, h_\tau)\right] = \frac{\partial}{\partial \tau}\mathbb{E}\left[u^*(X_i)k(X_i, h_\tau)\right].$$

Now by assumption on h_τ, we have that for any square integrable function $w(X_i)$,

$$\mathbb{E}_\tau\left[w(X_i)\rho(W_i, h_\tau(X_i))\right] = 0.$$

By differentiating with respect to τ, we have:

$$\frac{\partial}{\partial\tau}\mathbb{E}_\tau[w(X_i)\rho(W_i,h_0(X_i))]+\frac{\partial}{\partial\tau}\mathbb{E}\left[w(X_i)k(X_i,h_\tau)\right]=0.$$

By combining the 2 previous equations, we have:

$$\frac{\partial}{\partial\tau}\mathbb{E}[G_{\theta,\rho}(X_i,h_\tau)]=-\frac{\partial}{\partial\tau}\mathbb{E}_\tau\left[u^*(X_i)\rho(W_i,h_0(X_i))\right]=\mathbb{E}_\tau\left[-u^*(X_i)\rho(W_i,h_0(X_i))S(W_i)\right].$$

Following equation (3.9) in Newey (1994), we have that the adjustment term writes $\alpha(W_i)=-u^*(X_i)\rho(W_i,h_0(X_i))$. Observe that we also have have that $\mathbb{E}[\alpha(W_i)]=0$. Now let us derive $u^*(X_i)$ by exploiting the structure of our model. The same applies for the pathwise derivative of m, which depends only on h through $h(X_i)$:

$$\frac{\partial\mathbb{E}[G_{\theta,\rho}(X_i,h_0)|X_i]}{\partial h}[v]=\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}[v]=\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}v(X_i).$$

From the Riesz representation equation 2.B.24, we have that for any v,

$$\mathbb{E}\left[\frac{\partial\mathbb{E}[G_{\theta,\rho}(X_i,h_0)|X_i]}{\partial h}[v]-u^*(X_i)v(X_i)\right]=\mathbb{E}\left[\left(\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}-u^*(X_i)\right)v(X_i)\right]$$

The last equality holds for any v and in particular for $v(X_i)=\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}-u^*(X_i)$. Thus, we have:

$$\mathbb{E}\left[\left(\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}-u^*(X_i)\right)^2\right]=0$$

Thus, $u^*(X_i)=\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}$ almost surely. This yields the result:

$$\alpha(W_i)=\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^\top(\mathbf{1}\{Y_i=y\}-h_0(X_i)).$$

Step 2: (Derivation of the asymptotic distribution). By the mean-value Taylor expansion of $G_{\theta,\rho}(X_i,\hat{h}_0)$ around $G_{\theta,\rho}(X_i,h_0)$, there exists $\tilde{h}\in\mathcal{H}$ such that:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho}(X_i,\hat{h}_0) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho}(X_i,h_0) + \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}(\hat{h}_0(X_i)-h_0(X_i)) \\
&+ \underbrace{\frac{\rho}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top} J_\rho(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i))}_{O_p(\rho n^{1/2-2\gamma}) \text{ by } \textbf{??}(i)} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho}(X_i,h_0) + \underbrace{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial G_{\theta,\rho}(X_i,h_0)}{\partial h}^{\top}(\hat{h}_0(X_i)-h_0(X_i)) - \int\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x\right]}_{o_p(1) \text{ by Lemma 5}} \\
&+ \int\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x + \frac{\rho}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top} J_\rho(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_{\theta,\rho}(X_i,h_0)+\alpha(W_i)\right] + \frac{\rho}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top} J_\rho(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i)) \\
&+ \underbrace{\sqrt{n}\left[\int\frac{\partial G_{\theta,\rho}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x - \frac{1}{n}\sum_{i=1}^{n}\alpha(W_i)\right]}_{o_P(1) \text{ by Lemma 4}} + o_p(1)
\end{aligned}
$$

■

Proof of Proposition 4.9

Step 1: By the mean-value Taylor expansion of $G_{\theta,\rho_n}(X_i,\hat{h}_0)$ around $G_{\theta,\rho_n}(X_i,h_0)$, there exists $\tilde{h}\in\mathcal{H}$ such that:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i,\hat{h}_0)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i,h_0)+\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial G_{\theta,\rho_n}(X_i,h_0)}{\partial h}^{\top}(\hat{h}_0(X_i)-h_0(X_i))\\
&+\underbrace{\frac{\rho_n}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top}J_{\rho_n}(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i))}_{O_p(\rho_n n^{1/2-2\gamma})\ by\ \mathbf{??}(i)}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i,h_0)+\underbrace{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial G_{\theta,\rho_n}(X_i,h_0)}{\partial h}^{\top}(\hat{h}_0(X_i)-h_0(X_i))-\int\frac{\partial G_{\theta,\rho_n}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x\right]}_{o_p(1)\text{ uniformly over }\rho_n\text{ by Lemma 5}}\\
&+\int\frac{\partial G_{\theta,\rho_n}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x+\frac{\rho_n}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top}J_{\rho_n}(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i))\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_{\theta,\rho_n}(X_i,h_0)+\alpha(W_i)\right]+\frac{\rho_n}{\sqrt{n}}\sum_{i=1}^{n}(\hat{h}_0(X_i)-h_0(X_i))^{\top}J_{\rho_n}(X_i,\tilde{h})(\hat{h}_0(X_i)-h_0(X_i))\\
&+\underbrace{\sqrt{n}\left[\int\frac{\partial G_{\theta,\rho_n}(x,h_0)}{\partial h}^{\top}(\hat{h}_0(x)-h_0(x))\mathrm{d}x-\frac{1}{n}\sum_{i=1}^{n}\alpha(W_i)\right]}_{o_P(1)\text{ uniformly over }\rho_n\text{ by Lemma 4}}+o_p(1)
\end{aligned}
$$

with $\alpha(W_i)=\frac{\partial G_{\theta,\rho_n}(X_i,h_0)}{\partial h}^{\top}(\mathbf{1}\{Y_i=y\}-h_0(X_i))$, $J_{\rho_n}(x_i,\tilde{h})$ such that $\frac{\partial^2 G_{\theta,\rho_n}(x_i,\tilde{h})}{\partial h\partial h^{\top}}=\rho J_\rho(x_i,\tilde{h})$ and for any $(x_i,\tilde{h})$, $||J_{\rho_n}(x_i,\tilde{h})||_\infty\leq 2m$. So, we have

$$
\begin{aligned}
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}G_{\theta,\rho_n}(X_i,\hat{h}_0)-\mathbb{E}\left[G_{\theta,\rho_n}(X_i,h_0)\right]\right)&=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}G_{\theta,\rho_n}(X_i,h_0)+\alpha_n(W_i)-\mathbb{E}\left[G_{\theta,\rho_n}(X_i,h_0)\right]\right)\\
&+O_p(\rho n^{1/2-2\gamma})+o_p(1).
\end{aligned}
$$

Step 2: From triangular array CLT, we have

$$
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}G_{\theta,\rho_n}(X_i,h_0)+\alpha_n(W_i)-\mathbb{E}\left[G_{\theta,\rho_n}(X_i,h_0)\right]\right)\xrightarrow{d}\mathcal{N}(0,V_0),
$$

with $V_0 = \lim_{n\to\infty} \text{Var}\left[\left[G_{\theta,\rho_n}(X_i, h_0) + \alpha_n(W_i)\right]\right]$ as long as Lyapunov condition is satisfies or

$$\frac{1}{n^{\frac{\delta}{2}}}\mathbb{E}\left[\frac{|G_{\theta,\rho_n}(X_i, h_0) + \alpha_n(W_i) - \mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right]|^{2+\delta}}{\text{Var}(G_{\theta,\rho_n}(X_i, h_0) + \alpha_n(W_i))^{1+\frac{\delta}{2}}}\right] \to 0,$$

for some $\delta > 0$. It is easy to check that this holds as

$$\begin{aligned} G_{\theta,\rho_n}(X_i, h_0) + \alpha_n(W_i) = -\rho^{-1}\log\Big(&\frac{1 + \sum_{j=1}^{p}\exp(\rho(q_j^\top h(X_i) - C_{\theta,j}))}{1+p}\Big) \\ &+ \frac{\partial G_{\theta,\rho_n}(X_i, h_0)}{\partial h}^\top (\mathbf{1}\{Y_i = y\} - h_0(X_i)) \\ &\le \frac{\log(p+1)}{\rho} + 2. \end{aligned}$$

Similarly, we can show that it is also bounded below. Since $|G_{\theta,\rho_n}(X_i, h_0) + \alpha_n(W_i)|$ is bounded, Lyapunov condition is automatically satisfied.

Step 3: Let V_n be a consistent estimator for V_0. Now, we show asymptotic validity and consistency. Under H_0, $\theta \in \Theta_I$,

$$\theta \in \Theta_I \implies G_{\theta,\rho_n}(X_i, h_0) \ge 0 \text{ a.s} \implies \mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right] \ge 0\ \forall n.$$

(i) Asymptotic validity:

$$\begin{aligned} &\Pr\left(\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0)}{\sqrt{V_n}} \ge z_\alpha\right) \\ &= \Pr\left(\sqrt{n}\frac{\mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right]}{\sqrt{V_n}} + \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0) - \mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right]}{\sqrt{V_n}} \ge z_\alpha\right) \\ &\ge \Pr\left(\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0) - \mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right]}{\sqrt{V_n}} \ge z_\alpha\right) \\ &= \Pr\left(\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho}(X_i, h_0) + \alpha(W_i) - \mathbb{E}\left[G_{\theta,\rho}(X_i, h_0)\right] + o_P(1)}{\sqrt{V_n}} \ge z_\alpha\right) \qquad \underset{n\to\infty}{\to} 1 - \alpha \end{aligned}$$

(ii) Consistency:

$$
\begin{aligned}
\theta \notin \Theta_I &\implies \mathbb{E}\left[\lim_{n\to\infty} G_{\theta,\rho_n}(X_i, h_0)\right] = \gamma < 0 \\
&\implies \frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0) \xrightarrow{\mathbb{P}} \gamma \\
&\implies \frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0)}{\sqrt{V_n}} - \frac{\gamma}{\sqrt{V_0}} \xrightarrow{\mathbb{P}} 0
\end{aligned}
$$

Second line: from the asymptotic distribution in step 2: $\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0) = \mathbb{E}\left[G_{\theta,\rho_n}(X_i, h_0)\right] + o_p(1)$. Thus,

$$
\begin{aligned}
&\Pr\left(\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0)}{\sqrt{V_n}} \geq z_\alpha\right) \\
&= \Pr\left(\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0)}{\sqrt{V_n}} - \frac{z_\alpha}{\sqrt{n}} - \frac{\gamma}{\sqrt{V_0}} \geq -\frac{\gamma}{\sqrt{V_0}}\right) \\
&\leq \Pr\left(\left|\frac{\frac{1}{n}\sum_{i=1}^{n} G_{\theta,\rho_n}(X_i, \hat{h}_0) + \log(p)/\rho_n}{\sqrt{V_n}} - \frac{z_\alpha}{\sqrt{n}} - \frac{\gamma}{\sqrt{V_0}}\right| \geq -\frac{\gamma}{\sqrt{V_0}}\right) \underset{n\to\infty}{\rightarrow} 0.
\end{aligned}
$$

■

2.C Uniformity

Chapter 3

Identification and Estimation of Incentive Contracts under Asymmetric Information: an Application to the French Water Sector

co-authored with Christian Bontemps and David Martimort

Abstract

We develop a Principal-Agent model to represent management contracting for public-service delivery. A firm (the Agent) has private knowledge of its marginal cost of production. The local public authority (the Principal) cares both about the consumers' net surplus from consuming the services and the (weighted) firm's profit. The contractual negotiation is modeled as the choice by the privately informed firm within a menu of options determining both the unit price charged to consumers and the fixed fee. Our theoretical model characterizes optimal contracting in this environment. We then explicitly study the nonparametric identification of the model and perform a semi-parametric estimation on a dataset coming from the 2004 wave of a survey from the French environment Institute (IFEN, Institut Français de l'Environnement).
Keywords: Principal-Agent, optimal contracts, structural model, nonparametric estimation, Instrumental Variable Quantile Regression.
JEL codes: C12, C15, D82.

3.1 Introduction

The increasing participation of the private sector in public service delivery is often motivated by the need to expand access to services, increase or update existing delivery networks, and operate public utilities more efficiently. As many services provided by public utilities are associated with health, environmental or household income considerations (gas, electricity, water, transportation), there is a broad consensus on the need for public regulation of these utilities. In particular, industries such as water, gas and electricity usually meet the conditions for a local natural monopoly (large fixed costs and constant or declining marginal cost), so that protecting consumers from large price increases is often advocated as the main reason behind public regulation of utilities in these industries.

Private-sector participation in public utilities may take very different forms: private ownership of networks and facilities, centralized or local regulation by a public or independent authority, or contracting-out utility operations to private companies. In the latter case, typical arrangements are lease or concession contracts which can be renegotiated over time between a local community and a private company in charge of operating the public utility. Although contracting out seems an interesting way of promoting public-private partnerships for public utilities because it combines flexibility with legal commitment, it can deteriorate consumer welfare if the arrangement concerning utility pricing is not carefully specified. A major reason behind the difficulty to design an optimal pricing rule for the utility is the fact that common information on the operator's ability to manage the utility efficiently is rare. For example, the operator's technical know-how and expertise may not correspond exactly to the actual state of the facilities (based on past maintenance). Such a situation of asymmetric information on the operator's efficiency in a contract-based relationship has been studied extensively in the literature on incentives and regulation (Salanié (2005), Laffont and Martimort (2009), Laffont and Tirole (1993)).

In the standard theory of contracts, agents (operators in our case) are indexed by a private-information parameter that ultimately determines their actions, within a contract-based relationship with a "principal" (the local community). Whether this parameter denotes an unobservable action (moral hazard) or an unknown characteristic of the agent (adverse selection), the principal is assumed to have prior information used to design an optimal contract (in most cases, maximizing social welfare). The range of industries in which production, marketing, and regulation activities are subject to contract-based relationships between economic agents is sufficiently large to guarantee an increasing number of empirical applications for such a theory.

The first approach in the literature on econometric estimation of delegated management models with asymmetric information was based on reduced-form models, or structural versions with restrictive parametric restrictions on the technology and the distribution of the private parameter. Examples of this first generation of models with asymmetric information include Wolak (1994), Thomas (1995), Ivaldi and Martimort (1994), Brocas, Chan, and Perrigne (2006),Gagnepain and Ivaldi (2002). See Chiappori and Salanie (2000) for references on reduced-form estimation of models with asymmetric information, and Lavergne and Thomas (1997) for a survey on structural and reduced-form models. Most empirical applications confirm the fact that neglecting asymmetric information in the estimation of structural models yields biased estimates of, e.g., marginal cost or consumer price elasticity. However, it is also true that specification crucially matters for this type of structural models, as misspecification is likely to affect estimates of agents' preferences or production technology as much as neglecting asymmetric information altogether.

The second approach is more recent in the literature and proposes a way round this problem, in a series of articles based on nonparametric approaches. The move from parametric to nonparametric methods for estimating structural models with asymmetric information followed the development of structural models of auctions, with which they share some common features, as well as the literature on nonparametric identification (D'Haultfœuille and Février (2020) , Luo, Perrigne, and Vuong (2018) among others).

Structural models of delegated management with asymmetric information share common features with models of auctions or nonlinear pricing. The structural equations are nonlinear in an unobserved heterogeneity component (the private-information parameter) whose distribution is explicitly part of the solution to the underlying economic model. Moreover, endogenous selection of companies may also occur because of individual-rationality or incentive-compatibility constraints, in a way similar to auction participants or consumers in nonlinear pricing problems. For instance, Luo, Perrigne, and Vuong

(2018) propose a nonparametric identification method for models of nonlinear pricing, whose estimating equations closely resemble the ones associated with structural models of management delegation.

We propose in this paper a new method to estimate optimal contracts under asymmetric information, which relaxes most assumptions considered in the literature. More precisely, we rely extensively on nonparametric techniques to derive the distribution of the private information type and functional estimates of the model. The paper discusses the nonparametric identification of the structural model under several assumptions regarding unobserved heterogeneity. The estimation method is applied to the case of delegated management of water utilities in France. The advantage of considering such an industry for our empirical application lies in the fact that contracts between local communities and private firms for the operation of water utilities in France exist for a long time, under various modes that make the application of the theory of contracts particularly relevant. In the case we consider, the local community manager has incomplete information on the efficiency of the (private) operator of the water utility, before signing the delegation contract. The optimal contract is a second-best solution that depends on marginal social surplus for water, marginal cost of water supply, as well as on the distribution of the private type. A particular aspect of the model is the existence of two sources of unobserved heterogeneity (to the econometrician). Beside the usual private parameter, the econometrician does not observed an heterogeneity term associated with the social surplus and which is community-specific.

The paper is organized as follows. Section 3.2 presents the French water sector. Section 3.3 presents a theoretical principal-agent model of contract-based regulation between the local community and the operator of the water utility in a context of information asymmetry. This model is tailored to the specificities of the sector under scrutiny. The system of equations describing the optimal contract are first derived. Those equations summarize the optimality of the behaviors of the municipality and the operator. They serve as the basis for the estimation procedure. In Section 3.4, we discuss the conditions for nonparametric identification of the structural model. Two different cases are considered, depending on the assumptions on unobserved heterogeneity. Section 3.6 introduces the empirical application to the French water utilities with a particular focus on residential water pricing rules. Some counterfactuals are presented in Section 3.7. Section 3.8 discusses alternative modeling of optimal contracting relationships and illustrates some specificities of our approach. Finally, Section 3.9 concludes.

3.2 The French Water Sector

3.2.1 Governance

As in other countries throughout the world, the provision of water services in France is a regulated activity, although it has always been fully decentralized at the local level since the 1789 French Revolution. The need for regulation comes from the fact that each water network is indeed a public monopoly, with high fixed costs and low variable costs, as well as a declining (long-run) average cost curve. In each urban area of significant size equipped with such a water network, a local authority (a single city or a group of cities) has full responsibility to contract with an operator for providing water to the corresponding population.

In terms of size, water utilities represent no less than 1% of French national GDP. Yet, modes of governance significantly differ across networks. Water utilities in charge of supply, distribution and sewage activities may indeed be public (the so-called *"régies municipales"*) or private (*"gestation déléguée"*).[1] Delegation of utility operation to a private company is sometimes viewed as improving efficiency although, empirical studies are not really conclusive as to the relative efficiency of a management mode or the other.[2]

When the service is delegated to the private sector, various kinds of contracts (*"concession"*, *"affermage"*, *"gérance"...*) rule the details of this public-private partnership related to water production and distribution, maintenance, and quality supervision.[3] For the so-called *régies intéressées*, the opera-

[1] As a matter of comparison, ownership is mostly public in Germany and Italy, while it remains mixed in Spain. In Great Britain, water utilities are always run by private companies under the aegis of an independent national regulator (OFWAT). In the U.S., water utilities operate under rate-of-return regulation. The fact that State commission for public utilities may differ in their mandates gives rise to fairly heterogeneous rates of return.

[2] See Bhattacharyya, Harris, Narayanan, and Raffiee (1995), Bhattacharyya, Parker, and Raffiee (1994) and Estache and Rossi (1999) among others.

[3] In France, the public authority itself is in charge of verifying the quality of the service.

tor does not own the network and is paid as a fraction of the benefits of the service that accrues to the municipality. For a contract d' *"affermage"* , the operator is again paid directly by consumers but the cost of maintaining and renewing assets remains borne by the municipality. If the contract is a lease contracts *"concession"*, the operator must cover the costs of building and maintaining the infrastructure which are given back to the municipality at the end of the contract. The operator is directly paid through consumers' bills (the so-called *"redevance"*). By 2007, although private arrangements only concerned 39 % of water services, its share presented 72 % of the overall population. By 2008, 28% of networks were under public management. At the same time, the private sector is highly concentrated with three large companies sharing the market: Veolia Eau France, 39%, Lyonnaise des Eaux, 19 %, Saur, 11% and others companies or joint agreements among big ones amounting to less than 3 %. Veolia is the leading firm in the private sector. The smallest company Saur is more present in rural areas. It is also interesting to notice that public management is more frequent in small municipalities.

Given the high concentration in the sector, practitioners, medias and even politicians have sometimes complained that the market may not be so competitive after all, leaving inexpert municipalities in a weak bargaining position in front of big private players and subject to corruptive behavior. As a result, the *Loi Sapin* was enacted in 1993 to improve transparence and competitive biding while the 1995 *Loi Mazeaud* was designed to improve control on the operator. The competitive bidding generally relies on a two-step procedure. First the public authority chooses an operator through tender, whose criteria need not be publicized. Then a winner is selected after negotiation within those operators having made the best offers.

3.2.2 Pricing

Importantly, heterogeneity is a key aspect of the sector both on the demand and the cost sides. Networks are disconnected from one municipality to the other and vary significantly in terms of length, age, and thus leakages. For instance, although leakages amount only to 3% in Paris, they may reach up to 40% in some rural areas to average overall 21,9%. More maintenance helps reducing leakages and save water resource. Even though existing infrastructures are often old, the replacement rate for years 2006-2008 remained rather small with only 0,6% of the networks being renewed over that period.[4]

Given the specificities of each network, the operator managing the service may learn over time the state of the network. Private information over the cost of providing the service is thus pervasive. It introduces a fundamental asymmetry between municipalities and operators. This asymmetry is also partly due to the difficulties in assessing how costs, including labor costs, overhead costs, and maintenance investments, are allocated between water supply, distribution or sewage treatment, especially when more than one of these three operations are shared by the same operator.[5]

Such heterogeneity explains also significant disparities in consumption prices, with an average bill of 183 euros per inhabitant by 2008 which slightly increases over time as a means to finance an improving quality or new investments.

As far as pricing is concerned, water is billed with two-part tariffs; a usual feature of pricing in network industries.[6] The fixed fee helps to cover fixed cost while the ariable part that depends on consumption aims at paying for variable costs. Fixed fees vary significantly across networks, but on average are equal to 32 euros which represents around 20% of the bill for an average consumption of 120 m3.

3.3 Theory

Our theoretical model of the contractual relationship between municipalities (sometimes referred to as “principals" in the sequel) and service providers (the “agent’) fits the actual contractual practices reviewed in Section 3.2. In particular, informational asymmetries, heterogeneity and the form of pricing are key ingredients of any formal description of the sector.

[4]See Commissariat Général au Developpement Durable (2010).

[5]For a discussion of cost in the French stare sector, see Garcia and Thomas (2001).

[6]Hall (2000).

3.3.1 Preliminaries

We first set up the stage before entering into more details into the characterization of an optimal contract in the environment under scrutiny.

Demand side. We consider a population of heterogeneous municipalities that differ in terms of the consumers' surplus that prevails locally. More precisely, we assume that there exists a shift parameter ε, common knowledge for contracting parties (although not observed by the econometrician), such that the surplus in a municipality characterized by ε writes as $S(q,\varepsilon)$, where $S(q,\varepsilon)$ is increasing and concave in the consumption q and increasing in ε. This parameter ε allows to take into account heterogeneity on the demand side.

In a given municipality, aggregate demand for water at price p is then denoted as $D(p,\varepsilon) = (S'_q)^{-1}(p,\varepsilon)$ with $D'_p(p,\varepsilon) = \frac{1}{S''_{qq}(D(p,\varepsilon),\varepsilon)} < 0)$ for any realization of ε. Of course, the following identity holds

$$p = S'_q(D(p,\varepsilon),\varepsilon).$$

Supply side. The cost function of the service operator is parameterized as $\theta C_0(q)$, where q is the amount produced and θ is an efficiency parameter that enters multiplicatively. We assume that the function C_0 is strictly increasing and convex. Observe that this cost function satisfies the usual Spence-Mirrlees assumption; an operator with a more efficient technology (θ lower) also produces at a lower marginal cost. As usual in the screening literature, this assumption ensures that different operators can be sorted according to their marginal cost of producing the service and choose accordingly to produce under different contractual terms.

Information. In this paper, we are instead interested in the case where the cost parameter θ is the firm's private information. This parameter is distributed according to a common knowledge atomless distribution F, with a positive density function f on a bounded support $\Theta = [\underline{\theta}, \bar{\theta}]$. In accordance with the screening literature,[7] we impose the familiar *monotone hazard rate property*[8] that ensures fully separation allocations at an optimal contract.

[7] Guesnerie and Laffont (1984) and Laffont and Martimort (2009) (Chapter 3).

[8] See Bagnoli and Bergstrom (2006).

Assumption 1 (MHR)
$$\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \geq 0, \quad \forall\theta \in \Theta.$$

Contracts. The contract between the municipality and the firm stipulates not only a price p per unit of water produced but also an upfront subsidy A under the form of (per capita) subscription fees paid by consumers to access the service. That subsidy distributes the overall surplus between the consumers and the operator. Implicit in this specification of the contract is the idea that controlling the unit price amounts to controlling demand and thus the production that meets this demand.[9]

Following the incentive regulation literature, we will envision the result of contractual negotiations between the firm and the municipality as the choice of an item by the privately informed party within a menu of options. Two equivalent approaches might be used to model this choice. The first one relies on the so-called *Revelation Principle*[10]that states that there is no loss of generality in looking for contracts that are direct and truthful mechanisms of the kind $\{A(\hat{\theta},\varepsilon), p(\hat{\theta},\varepsilon)\}_{\hat{\theta}\in\Theta}$. Note that we index the contract by the demand shock ε which is commonly known by contracting parties. With such direct communication, the operator picks a subsidy/unit price according to his efficiency parameter by communicating information on its cost parameter. The mechanism is incentive compatible when each operator ends up preferring the option targeted to his own type.

An alternative approach based on the so-called *Taxation Principle*[11] gives up the abstract direct communication process underlying the Revelation Principle and focuses instead on the true economic choice made by the privately informed party. Facing a nonlinear scheme $A(p,\varepsilon)$[12] linking the value of the subscription fee to the actual per unit price chosen by the firm, the firm chooses optimally at

[9]See Baron (1988) for a general formulation of regulatory mechanisms relying on such approach. It is only incidentally different from the standard approach that focuses on the direct control of quantities that is developed in Baron and Myerson (1982) and Laffont and Tirole (1993).

[10]See Myerson (1982) and Laffont and Martimort (2009) for a textbook approach.

[11]See for instance Rochet (1985).

[12]Without fear of confusion and for the sake of simplifying presentation, we slightly abuse notations here by denoting similarly the fixed fee viewed as a function of the agent's type in a direct mechanism and viewed as a function of the price in an indirect scheme. Notice also that the nonlinear scheme $A(\cdot,\varepsilon)$ is again indexed on the commonly observable variable ε that characterizes the relationship under scrutiny.

which price it stands to produce.[13] Again, this choice can be viewed as a metaphor for more complex negotiation procedures where firms and public authorities negotiate over both the fixed fee and the unit price charged to consumers. We will favor this second approach since it relates to actual observables available to us (price, quantity, fees) that will be used in our econometric analysis.

Objective functions of the contracting parties. With our previous notations at hand, the expression of the firm's profit becomes:

$$\mathcal{U}(\theta, \varepsilon, p, A) = A + pD(p, \varepsilon) - \theta C_0(D(p, \varepsilon)).$$

Following Baron and Myerson (1982), a municipality maximizes a welfare function which includes not only consumers' net surplus from consuming the service but also the firm's profit weighted by some parameter $\gamma \in [0, 1[$. This parameter can be viewed as an index of the firm's bargaining power at the stage of tenders or during contract negotiations. It can also be inherited from how local political forces interact as argued in Baron (1989).[14] The corresponding welfare function $W(\cdot)$ can thus be written as:

$$\mathcal{W}(\theta, \varepsilon, p, A) = S(D(p, \varepsilon), \varepsilon) - A - pD(p, \varepsilon) + \gamma \mathcal{U}(\theta, \varepsilon, p, A).$$

Using the expression of the fixed fee as a function of the firm's profit, the latter definition becomes:

$$\mathcal{W}(\theta, \varepsilon, p, A) = S(D(p, \varepsilon), \varepsilon) - \theta C_0(D(p, \varepsilon)) - (1 - \gamma)\mathcal{U}(\theta, \varepsilon, p, A).$$

This latter expression stresses the rent/efficiency trade-off faced by the local government in designing the regulatory contract. On the one hand, the principal would like to charge a price $p(\theta, \varepsilon)$ close to $p^*(\theta, \varepsilon)$ as defined in (3.3.1) so that the overall surplus is maximized. On the other hand, the principal would also like to reduce the firm's informational rent which is viewed as socially costly. Under asymmetric information, rents and outputs are linked altogether through incentive compatibility conditions and this leads to an important trade-off between the conflicting objectives of promoting efficiency and extracting rents.

[13] Of course, any incentive compatible direct mechanism $\{A(\hat{\theta}, \varepsilon), p(\hat{\theta}, \varepsilon)\}_{\hat{\theta} \in \Theta}$ can be transformed into a nonlinear scheme by setting $A(p, \varepsilon) = A(\hat{\theta}, \varepsilon)$ if $p = p(\hat{\theta}, \varepsilon)$ and $A(p, \varepsilon) = -\infty$ otherwise.

[14] See also Gagnepain, Ivaldi, and Martimort (2013) for a model of the French transportation sector that relies on a similar specification of the preferences of local authorities.

Remark 1 *The expressions of the objective functions above can easily be extended to account for some fixed cost F in the operatopr's cost function. Suppose indeed that the operator's profit can be written as $\mathcal{U}(\theta,\varepsilon,p,A) = A' + pD(p,\varepsilon) - \theta C_0(D(p,\varepsilon)) - F$. for some fixed-fee A'. Setting $A \equiv A' - F$ then amounts to having the principal pays for the fixed-cost in the first place, which is basically an accounting convention.*

Benchmark. Had θ been common knowledge, efficiency would require to produce a quantity $q^*(\theta,\varepsilon) = D(p^*(\theta,\varepsilon),\varepsilon)$ such that the marginal social value of production is equal to marginal cost:

$$S'_q(q^*(\theta,\varepsilon),\varepsilon) = p^*(\theta,\varepsilon) = \theta C'_{0q}(q^*(\theta,\varepsilon)). \tag{3.3.1}$$

Then, the firm's operates if $A^*(\theta,\varepsilon)$ extracts revenues from the service:

$$A^*(\theta,\varepsilon) = -p^*(\theta,\varepsilon)D(p^*(\theta,\varepsilon),\varepsilon) + \theta C_0(D(p^*(\theta,\varepsilon),\varepsilon)). \tag{3.3.2}$$

Remark 2 *In the empirical part of our analysis, the two functions $S(\cdot,\varepsilon)$ and $C_0(\cdot)$ will depend on a set of explanatory variables. For example, the treatment made for making the water drinkable has an impact on the unit cost. This treatment is however observed in the data. For the exposition, we omit the dependance in the explanatory variables without loss of generality for the results derived in the analysis of the theoretical model.*

3.3.2 Optimal Contract

Incentive compatibility constraints. Let define the firm's information rent $U(\theta,\varepsilon)$ and an optimal price[15] respectively as:

$$U(\theta,\varepsilon) = \max_p A(p,\varepsilon) + pD(p,\varepsilon) - \theta C_0(D(p,\varepsilon)) \tag{3.3.3}$$

and

$$p(\theta,\varepsilon) = \arg\max_p A(p,\varepsilon) + pD(p,\varepsilon) - \theta C_0(D(p,\varepsilon)). \tag{3.3.4}$$

[15] Or at least, a selection within the best-response correspondence.

From (3.3.3), $U(\theta,\varepsilon)$ is the maximum of a family of decreasing linear functions in θ. As such it is decreasing, convex in θ and absolutely continuous so that one can write:

$$U(\theta,\varepsilon) = U(\bar{\theta},\varepsilon) + \int_{\theta}^{\bar{\theta}} C_0\left(D(p(x,\varepsilon),\varepsilon)\right) dx. \tag{3.3.5}$$

At any point of differentiability in θ (i.e., almost everywhere), we get:

$$U'_{\theta}(\theta,\varepsilon) = -C_0\left(D(p(\theta,\varepsilon),\varepsilon)\right). \tag{3.3.6}$$

Because $U(\theta,\varepsilon)$ is convex in θ, $C_0\left(D(p(\grave{},\varepsilon),\varepsilon)\right)$ is non-decreasing in θ, which in turn implies:

$$p(\theta,\varepsilon) \text{ is non-increasing in } \theta. \tag{3.3.7}$$

This condition expresses the fact that firms endowed with less efficient technologies produce lower volumes at higher prices . From this monotonicity, it also follows that $p(\theta,\varepsilon)$ is almost everywhere differentiable in θ.

Let us turn now to the expression of the nonlinear schedule $A(p,\varepsilon)$ that plays an important role in our empirical analysis. By a standard duality argument of (*generalized*) convex analysis, we may first rewrite:[16]

$$A(p,\varepsilon) + pD(p,\varepsilon) = \min_{\theta} U(\theta,\varepsilon) + \theta C_0(D(p,\varepsilon)).$$

From which, it also follows that $A(p,\varepsilon)$ is absolutely continuous in p and thus such that

$$\begin{aligned} A(p,\varepsilon) + pD(p,\varepsilon) =& A(p(\bar{\theta},\varepsilon),\varepsilon) + p(\bar{\theta},\varepsilon)D(p(\bar{\theta},\varepsilon),\varepsilon) \\ &+ \int_{p(\bar{\theta},\varepsilon)}^{p} \vartheta(p,\varepsilon)C'_{0q}\left(D(p,\varepsilon)\right) D'_p(p,\varepsilon)dp \end{aligned} \tag{3.3.8}$$

where $\vartheta(p,\varepsilon) = \min_{\theta} U(\theta,\varepsilon) + \theta C_0(D(p,\varepsilon))$ is an assignment function[17] (a selection within the best-response monotonically increasing p (thus almost everywhere differentiable) and such that $\vartheta(p(\theta,\varepsilon),\varepsilon) \equiv \theta$.

At any point of differentiability in $p = p(\theta,\varepsilon)$, we thus have also:

$$A'_p(p(\theta,\varepsilon),\varepsilon) = -\Big(p(\theta,\varepsilon) - \theta C'_{0q}\left(D(p(\theta,\varepsilon),\varepsilon)\right)\Big) D'_p(p(\theta,\varepsilon),\varepsilon) - D(p(\theta,\varepsilon),\varepsilon). \tag{3.3.9}$$

[16] See for instance Basov (2005) Chapter 7).

[17] See Nöldeke and Samuelson (2005).

Participation constraints. The firm chooses to always operate the service irrespectively of its costs when it at least breaks even. For a fixed ε, this participation constraint can be written as:

$$U(\theta,\varepsilon) \geq 0 \qquad \forall \theta.$$

As usual in the incentive regulation literature,[18] this constraint is binding for the worst type $\bar{\theta}$ at the optimal contract. Otherwise reducing uniformly the fixed fee by some small amount would improve the principal's expected payoff while maintaining incentive compatibility. From this observation, an immediate manipulation of (3.3.5) yields the following expression of the firm's information rent as:

$$U(\theta,\varepsilon) = \int_{\theta}^{\bar{\theta}} C_0\left(D(p(x,\varepsilon),\varepsilon\right) dx. \tag{3.3.10}$$

Observe that the rent left to the operator is greater as prices are lower. The intuition is a standard one. By pretending being slightly less efficient, an operator with efficiency parameter θ can produce the same quantity than this slightly less efficient type $\theta + d\theta$ but at a lower marginal cost. To induce this operator to report truthfully his type he must be given an extra fee that equals the corresponding cost saving $d\theta C_0\left(D(p(\theta+d\theta,\varepsilon),\varepsilon)\right) \approx d\theta C_0\left(D(p(\theta+d\theta,\varepsilon),\varepsilon)\right)$. The right-hand side of (3.3.10) expresses how those marginal information rents just pill up over all supra-marginal types.

Optimal contracts. Under asymmetric information, an optimal contract maximizes the expected welfare of the municipality subject to incentive and participation constraints. From our observations above, that incentive feasible set can be summarized by constraints (3.3.7) and (3.3.10). As usual, the monotonicity condition (3.3.7) will be omitted in a first step and checked ex post on the solution to the so relaxed problem. Formally, this relaxed problem can be written as:

$$\max_{\{p(\cdot,\varepsilon),U(\cdot,\varepsilon)\}} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(D(p(\theta,\varepsilon),\varepsilon)) - \theta C_0\left(D(p(\theta,\varepsilon),\varepsilon)\right) - (1-\gamma)U(\theta,\varepsilon)\right] dF(\theta) \text{ subject to (3.3.10).}$$

Using (3.3.10) and integrating by parts yields the following expression of the expected rent left to the operator:

$$\int_{\underline{\theta}}^{\overline{\theta}} U(\theta,\varepsilon)dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} \frac{F(\theta)}{f(\theta)} C_0\left(D(p(\theta,\varepsilon),\varepsilon)\right) dF(\theta).$$

[18]See Armstrong and Sappington (2007), Baron and Myerson (1982) and Laffont and Tirole (1993) among others.

This expression can be incorporated into the maximand above before proceeding to pointwise optimization. This last step leads to the following expression of the price per-unit of consumption $p(\theta, \varepsilon)$.

$$p(\theta, \varepsilon) = \left(\theta + (1-\gamma)\frac{F(\theta)}{f(\theta)}\right) C'_{0q}\left(D(p(\theta, \varepsilon), \varepsilon)\right). \tag{3.3.11}$$

The corresponding volume that is supplied $q(\theta, \varepsilon)$ is then defined as:

$$S'_q(q(\theta, \varepsilon), \varepsilon) = p(\theta, \varepsilon). \tag{3.3.12}$$

Equation (3.3.11) indicates that the price is now above marginal costs and, as a result of (3.3.12). Equilibrium quantities are also lower than at the first best. Increasing the unit price above marginal cost reduces the demand addressed to the operator. It thus reduces the latter's information rent. Formally, everything happens as if the cost parameter was now replaced by a *virtual cost parameter* $H(\theta, \gamma)$ which is greater:

$$H(\theta, \gamma) = \theta + (1-\gamma)\frac{F(\theta)}{f(\theta)}.$$

This expression first illustrates the rent/efficiency trade-off that arises under asymmetric information and, second, how this trade-off is modified as parameters of the model change. Indeed, the virtual cost parameter is greater as the public authority is more concerned by rent extraction (i.e., γ lower) and as the types distribution is more front-loaded, in the sense of having a greater hazard rate $F(\theta)/f(\theta)$.

Turning now to sufficient conditions for optimality, observe that Assumption 1 ensures that $p(\theta, \varepsilon)$ so defined by (3.3.11) is non-decreasing in θ as requested by condition (3.3.7). Hence, the solution of the relaxed problem really characterizes the optimal contract. When we get to our empirical analysis, we will actually check on our estimated distribution that it indeed satisfies Assumption 1.

3.4 Nonparametric Identification

In this section, we study the nonparametric identification of our model. It is indeed important to figure what are the structural functions that can be fully recovered from the available data. In our dataset, we observe for each local community, the unit price p, the quantity consumed q and the fixed fee A. We also observe some explanatory variables W and Z which are related to respectively the cost function $C_0(q, W)$ and the surplus function $S(q, Z, \varepsilon)$. The vectors W and Z do not have any variable in common in our case but the results are unchanged when the reverse holds as long as there are some exclusion restrictions, i.e., there exists an explanatory variable in Z which is not part of W and an explanatory variable in W which is not part of Z. Our purpose is to identify the production technology $C_0(q, W)$, the distribution $F(\theta)$ of the types θ, the consumers surplus function $S_0(q, Z, \varepsilon)$, and the weight γ of the firm's profit in the principal's objective.

A scale normalization Multiplying θ by a positive scalar λ and dividing the cost function by the same value would give the same equilibrium outcome. θ and the marginal cost are therefore identified up to a scale and we first need to impose a normalization for the distribution of θ, by assuming that a prespecified quantile is equal to a given value. Standard normalizations are $\underline{\theta}$ equal to 1 or the median of θ equal to 1. In the latter case, the function $C'_{0q}(q, W)$ is then interpreted as the marginal cost function for the median type firm, given the observed characteristics W.

Assumption 2 *[Normalization]*

$$Median(\theta) = 1.$$

3.4.1 The Simple Case Without Heterogeneity

As a starting point, we omit explanatory variables (no W and no Z)and unobserved heterogeneity (*i.e.* $\varepsilon \equiv 0$), simplifying notations accordingly. For exposition purposes, it is important to know what are the identifying power of the first-order conditions (3.3.9), (3.3.11) and (3.3.12) and what do the additional assumptions help us identify. The system of first-order conditions reduces in this case to[19]

$$p = H(\theta, \gamma) C'_{0q}(q), \tag{3.4.13}$$

$$S'(q) = p, \tag{3.4.14}$$

$$A'(p) = -q - \left(p - \theta C'_{0q}(q)\right) D'_p(p). \tag{3.4.15}$$

[19] We omit, in the notations, the dependance of p and q in θ for the clarity of exposition.

First, let us observe that equation (3.4.14) directly identifies $S'(q)$ on the support of the equilibrium quantities and hence provides the expression of $D'_p(p) = \frac{\partial}{\partial p}\left(S'^{-1}\right)(p)$ that we insert into (3.4.15). Finally, (3.4.15) provides information about the operator's price-cost margin

$$\frac{p - \theta C'_{0q}(q)}{p} = \frac{A'(p) + q}{pD'_p(p)}.$$

This is so because the terms on the right-hand side are either observed ($q = D(p)$ and p) or derived directly from the observations ($A'(p)$ and $D'(p)$).

Making the dependence of the price-cost margin on θ explicit, we define a price-cost margin $r(\theta, \gamma)$ as:

$$r(\theta, \gamma) = \frac{p(\theta) - \theta C'_{0q}(q(\theta))}{p(\theta)} = \frac{1}{1 + \frac{\theta f(\theta)}{(1-\gamma)F(\theta)}} \tag{3.4.16}$$

where the last equality immediately follows from (3.4.13). Thus, the price-cost margin only depends on the efficiency parameter θ and the bargaining power γ. In the sequel, we shall assume that different operators can be perfectly sorted according to that price-cost margin. Formally, we require that there is a one-to-one mapping between price-cost margins and efficiency parameters, i.e., $r(\theta, \gamma)$ is a monotonically decreasing transformation of θ which always lies between 0 and 1 and is worth 0 at $\underline{\theta}$, i.e., for the most efficient type who produces efficiently:

Assumption 3 (MPCM)

$$\frac{d}{d\theta}\left(\frac{F(\theta)}{\theta f(\theta)}\right) \geq 0, \quad \forall \theta \in \Theta.$$

This assumption is actually **stonger** than Assumption 1 and is satisfied for all standard parametric distributions.

To give a bit more intuition about the role played by Assumption 3, let us come back to (3.4.15) which can be rewritten as:

$$A'(p(\theta)) = -D(p(\theta))\left(1 + r(\theta)\varepsilon_D(p(\theta))\right) \tag{3.4.17}$$

where we denote the demand elasticity by $\varepsilon_D(p) = -\frac{pD'(p)}{D(p)}$. Assuming also that demand is more elastic at greater price, i.e., $\varepsilon_D(p)$ is increasing with p, it is straightforward to check by differentiating (3.4.17) that Assumption 3 ensures that $-A'(p(\theta))/D(p(\theta))$ is increasing in θ, or alternatively that $A(p)$ is quasi-concave in p.

Equipped with this one-to-one relationship between cost-price margins and efficiency parameters, we now let $G(\cdot)$ (resp. $g(\cdot)$) be the cumulative distribution function (resp. probability density function) of that margin. Of course, we have $G(r(\theta,\gamma)) = F(\theta)$ and by differencing the last equality with respect to θ, we also get $g(r) = f(\theta)\theta'_r(r,\gamma)$ where $\theta(r,\gamma)$ denotes the inverse function of $r(\theta,\gamma)$. Reintroducing the last two equalities into (3.4.16), we obtain after some manipulations:

$$(1-\gamma)\frac{\theta_r(r,\gamma)}{\theta(r,\gamma)} = \frac{r}{1-r}\frac{g(r)}{G(r)}.$$

For exposition, let us assume that we impose the normalization $\underline{\theta} = 1$.[20] As $r = 0$ when $\theta = \underline{\theta} = 1$, we can solve the latter differential equation and get:

$$\theta(r,\gamma)^{1-\gamma} = \exp\left[\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right]. \tag{3.4.18}$$

The density function for θ is then derived from $f(\theta) = \frac{g(r)}{\dot{\theta}(r)}$:

$$f(\theta) = (1-\gamma)\frac{1-r}{r\theta}G(r). \tag{3.4.19}$$

Observe that the last expression gives the density of the types as a function of the c.d.f. of the price cost margin. Reintroducing (3.4.18) into (3.4.13) and observing that $\theta = (1-r(\theta,\gamma))H(\theta,\gamma)$ lead to

$$C'_{0q}(q) = \frac{p}{H(\theta,\gamma)} = \frac{p(1-r)}{\exp\left[\frac{1}{1-\gamma}\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right]}.$$

Therefore, when γ is known or predetermined from some additional source of information, the model is thus identified since the marginal cost and the distribution of the firms' types are identified up to the standard normalization.

[20] If we use the normalization related to the median, Equation (3.4.18) below becomes

$$\theta(r,\gamma) = \underline{\theta}\exp\left[\frac{1}{1-\gamma}\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right].$$

$\underline{\theta}$ is determined to ensure that $Median(\theta) = 1$.

Instead, when γ is not known, only $\theta^{1-\gamma}$ is identified. Consequently, for any β in $[0,1[$, the two sets $(\theta, C'_{0q}(q), \gamma)$ and $(\theta^{\frac{1-\gamma}{1-\beta}}, C'_{0q}(q)\theta^{\frac{\gamma-\beta}{1-\beta}}, \beta)$ are observationally equivalent. We thus need additional assumptions or information to identify the model. It is worth noting that the conclusions of this section do not depend on the existence of observed explanatory variables. Adding an explanatory variable in either the cost function or the surplus one does not change the previous results. Observe also that we can exploit the bounds on β to bound the distributions of interest. In particular

$$C'_{0q}(q) \leq \frac{p}{H(\theta,\gamma)} = \frac{p(1-r)}{\exp\left[\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right]},$$

and

$$f(\theta) \leq \frac{1-r}{r\theta}G(r), \text{ on } [1;+\infty[.$$

3.4.2 Full Identification of the Model with Explanatory Variables and Heterogeneity

We now consider the full model with explanatory variables, Z and W, that appears respectively in the surplus function and the cost function.

Identification of the Marginal Surplus Function under Completeness Assumption

• In a first step, we assume that the marginal cost function is separable and additive in both the explanatory variables, Z,[21] and the unobserved heterogeneity term, ε, i.e. $S'_q(q,Z,\varepsilon) = S'_0(q) + \beta_Z Z + \varepsilon$, Equation (3.3.12) can be written

$$p = S'_0(q) + \beta Z + \varepsilon, \tag{3.4.20}$$

with $[\varepsilon|Z,W] = 0$.

[21] The separability in Z does not play any role in the identification result. This is nevertheless the model we estimate in the empirical application.

The equilibrium quantity q being endogenous, Equation (3.4.20) is a typical nonparametric instrumental regression like the one studied in Florens, Johannes, and Van Bellegem (2012).[22] Following Theorem 3 in Florens, Johannes, and Van Bellegem (2012), we recover the identification of the primitives of the model under a condition of completeness of W with respect to q. Observe that here, in the partial linear regression like the one in (3.4.20), we need additional regularity conditions than the completeness assumption due to the linear part $\beta_Z Z$ (see Florens, Johannes, and Van Bellegem (2012))[23]. The completeness assumption is nevertheless the most restrictive assumption though relatively standard in this literature. The set of instruments W is complete for q if, for any measurable function ΔS in L^1,

$$E[\Delta S(q)|W] = 0 \; a.s. \; \Rightarrow \Delta S(q) = 0 \; a.s.$$

Sufficient conditions for completeness can be found in Newey and Powell (2003), Chernozhukov and Hansen (2005), and Andrews (2011). It can be replaced by weaker concepts like the bounded completeness in Chernozhukov and Hansen (2005), Blundell, Chen, and Kristensen (2007), D'Haultfoeuille (2011), and Andrews (2011).[24]

• We now consider the non-separable case,

$$p = S'_q(q, Z, \varepsilon),$$

with $[\varepsilon|Z, W] = 0$. Let V be indeed a given c.d.f., i.e. an injective function from $\mathbb{R}$ to $[0,1]$. Let $\tilde{\varepsilon} = V^{-1} \circ \Phi(\varepsilon)$, where $\Phi(\cdot)$ denotes the c.d.f. of ε, and $\tilde{S}(q, Z, \tilde{\varepsilon}) = S(q, Z, \Phi^{-1} \circ V(\tilde{\varepsilon}))$. The equilibrium quantities A, p and q are invariant to this transformation.

The distribution of the parameter ε is therefore non identifiable as such and we therefore need to normalize the distribution of ε. The standard one is to assume that it is uniformly distributed, i.e. $\varepsilon \sim U[0;1]$. It is worth noting, that this normalization does not restrict the economic interpretation of our

[22] See also Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011) and Ai and Chen (2003) for additional references.

[23] It is required that the conditional expectations of Z, q and any L^2 function of q given W are in $L^2(\mathbb{R}^{dimW})$ and that the matrix $E\left[E(Z|W)E(Z|W)^\top\right]$ is full rank.

[24] See also Chen et al. for a general discussion. This assumption is however not testable as recently shown by Canay, Santos, and Shaikh (2013).

model. The value of ε does not have an interpretation in its own and is only an index of the position of the firm in a demand ranking. We therefore identify the marginal surplus (given Z) of any quantile of the distribution of ε, like the marginal surplus of the median local community, given Z.

The identification arises from the monotonicity constraints implied by the economic model (see, for example Chesher (2007), or Chernozhukov and Hansen (2005)). For a given rank $\alpha \in [0,1]$:

$$\begin{aligned}\mathbb{P}\left(p \leq \left(S'_q(q,\alpha,Z)\right)|Z,W\right) &= \mathbb{E}_{q|Z,W}\mathbb{P}\left(S'_q(q,\varepsilon,Z) \leq \left(S'_q(q,\alpha,Z)\right)|q,Z,W\right)\\ &= \mathbb{E}_{q|Z,W}\mathbb{P}(\varepsilon \leq \alpha|q,Z,W) \text{ due to the monotonicity in } \varepsilon \\ &= \mathbb{P}(\varepsilon \leq \alpha|Z,W) = \alpha\end{aligned} \tag{3.4.21}$$

Equation (3.4.21) derived above can be reexpressed as a conditional moment restriction:

$$\mathbb{E}\left(\mathbf{1}\{p \leq S'_q(q,\alpha,Z)\} - \alpha|Z,W\right) = 0. \tag{3.4.22}$$

This is a standard quantile IV equation and we need essentially an assumption of completeness of W with respect to q to identify any quantile of the marginal surplus function given Z. The following result summarizes the main conclusion if this section.

Proposition 1 *If W is complete with respect to q the marginal surplus function given Z is identified.*

3.4.3 Identification of the marginal cost function and the distribution of types

We now prove that the marginal cost function is identified without the knowledge of γ when there is unobserved heterogeneity of the demand function. Assume initially that we do not have explanatory variables for the marginal cost.

We consider the points (q_i, p_i) of the quantities consumed and marginal prices paid for all the contracts (see Figure 3.1). For the most efficient firms, $\theta = \underline{\theta}$, the marginal price equation (3.3.11) reveals that the marginal cost function for the lower type $\underline{\theta}C'_{0q}(q)$ can be estimated using the lower envelope of the points in the space (q,p) between the quantity $q^l_{\underline{\theta}}$ which corresponds to the most efficient firm contracting with the lowest demand city (the contract which has the lowest price) and the quantity $q^u_{\underline{\theta}}$

which corresponds to the most efficient firm contracting with the highest demand city (i.e. the contract with the highest quantity). This is the dashed curve in the figure.

Similarly, for the less efficient firms, $\theta = \overline{\theta}$, the marginal prices are equal to $kC'_{0q}(q)$ where k is equal to the limit of $\theta + (1-\gamma)/f(\theta)$ on the upper bound, when it exists. This function is therefore the upper envelope between the minimum quantity among all contracts $q^l_{\overline{\theta}}$ (i.e. the quantity of the contract of the less efficient firm contracting with the local community with the lowest surplus) and the quantity corresponding to the maximum price $q^u_{\overline{\theta}}$ (the quantity consumed in the local community with the highest surplus which contracts with the less efficient firm).

If $q^u_{\overline{\theta}} \geq q^l_{\underline{\theta}}$ and if k is finite, we can estimate the ratio of the two extreme costs with the same quantity q and therefore estimate $C'_{0q}(q)$ up to a scale between the minimum and the maximum quantities of the population of contracts. We need to have sufficient heterogeneity in ε. If there are explanatory variables W, we can use the same argument conditionally on W.

Alternatively, if k is infinite, or if there is not enough heterogeneity, we can use a moment equation like the one used for estimating the surplus function in (3.4.21). Using the monotone property of $H(\theta) = \theta + (1-\gamma)\frac{F(\theta)}{f(\theta)}$ in θ, we can indeed write for any quantile $\alpha \in]0,1[$, θ_α being the corresponding quantile of θ,

$$\mathbb{P}\left(p \leq H(\theta_\alpha, \gamma)C'_{0q}(q, W)|Z, W\right) = \alpha. \tag{3.4.23}$$

Again, $C'_{0q}(q, W)$ is identified through (3.4.23) up to a scale under an assumption of completeness of Z. We can in a first step normalize $H(\theta_{0.5}, \gamma)$ to 1 before changing the scale to ensure that the median of θ is equal to 1.

Once $C'_{0q}(q, W)$ is known up to a scale, we obtain θ from (3.3.9) and estimate the scale to ensure $Median(\theta) = 1$. The marginal cost function is therefore identified. The knowledge of θ provides the identification of $F(\theta)$ and $f(\theta)$.

Proposition 2 *If Z is complete with respect to q the marginal cost function given W, $C'_{0q}(q, W)$ and the distribution of types are identified.*

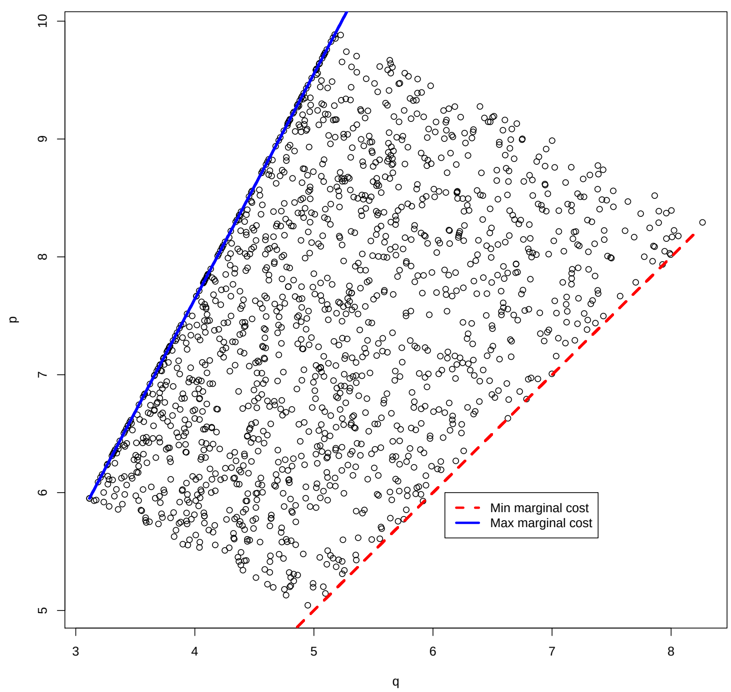


Figure 3.1: The set of contracts in the $(q \times p)$ space.

3.4.4 Identification of γ

Once, all the functions are identified, $H(\theta,\gamma)$ is known from (3.3.11). We can now identify the bargaining weight γ using

$$H(\theta,\gamma)=\theta+(1-\gamma)\frac{F(\theta)}{f(\theta)}.$$

Identification when γ is not unique It is worth noting that the identification strategy does not require the unicity of γ across the cities as long as the distribution of γ is exogeneous. Assume that γ varies across cities, its c.d.f being denoted by $\Gamma(\cdot)$, and that the observed heterogeneity Z and W are both complete with respect to q as required in Proposition 1 and 2. The equilibrium equations are not changed and we identify similarly $S_q'(q,Z)$.

Equation (3.4.23) is now modified as we have to condition on γ. For any quantile $\alpha\in]0,1[$,

$$\mathbb{P}\left(p\leq H(\theta_\alpha,\gamma)C_{0q}'(q,W)|Z,W,\gamma\right)=\alpha. \tag{3.4.24}$$

For any $\gamma\in[0,1[$, the function $H(\cdot,\gamma)$ is a one-to-one mapping from $[\underline{\theta},\overline{\theta}[$ to $[\underline{\theta},\lim_{\theta\to\overline{\theta}}\theta+(1-\gamma)\frac{F(\theta)}{f(\theta)}[$.

Let $\theta(k,\gamma)=\max_{\theta\in\Theta}\{H(\theta,\gamma)\leq k\}$. For any γ, $\theta(\underline{\theta},\gamma)=\underline{\theta}$ and $\theta(\cdot,\cdot)$ is increasing in k and decreasing in γ. Therefore

$$\mathbb{P}\left(p\leq kC_{0q}'(q,W)|Z,W,\gamma\right)=F(\theta(k,\gamma)),$$

and, integrating with respect to gamma,

$$\mathbb{P}\left(p\leq kC_{0q}'(q,W)|Z,W\right)=\int F(\theta(k,\gamma))d\Gamma(\gamma)=r(k). \tag{3.4.25}$$

$r(k)$ is increasing in k and maps $[\underline{\theta},\lim_{\theta\to\overline{\theta}}\theta+\frac{F(\theta)}{f(\theta)}[$ into $[0,1[$. We have therefore derived the same structure of quantiles than in the case where γ is unique. Again, we can identify the marginal cost up to a scale before deriving θ from (3.3.9). Once θ is known, we can rescale it to meet our normalization $Median(\theta)=1$. $H(\theta,\gamma)$ is therefore identified from $H(\theta,\gamma)=p/C_{0q}'(q,W)$.

Finally the distribution of γ is identified from

$$H(\theta,\gamma)=\theta+(1-\gamma)\frac{F(\theta)}{f(\theta)}.$$

3.5 Simulations and Estimation Procedure

In this section, we expose the different steps of our estimation strategy. Some Monte Carlo simulations are also displayed to assess the sample properties of the proposed estimator.

The results are based on $n_s = 100$ replications of samples of size $n = 1000$. In this scenario the marginal cost function is a linear function $C'_{0q}(q, W) = q + W$, where $W \sim \mathbb{U}_{[0,1]}$. θ the type of the firm follows a Beta distribution up to a location-scale transformation, i.e. $(\theta - 1)/5 \sim B(2, 5)$ and therefore $\Theta = [1, 6]$. $S'_0(q, Z) = 15 - 5q + Z$, where $Z \sim \mathbb{U}_{[0,1]}$ and $\gamma = 0.3$. The unobserved heterogeneity ε is drawn from a centered uniform distribution.

We display for the estimated functions the integrated square bias, integrated variance and integrated mean square error based on 100 simulations. Let $\hat{f}_s(\cdot)$ be the estimated function of $f(\cdot)$ for simulation s. Let $x_1, ..., x_{99}$ be the 99 percentiles of the variable x in the population. We define $\overline{f}(x) = \frac{1}{n_s}\sum_{s=1}^{n_s} \hat{f}_s(x)$.

The integrated square bias is estimated by $ISB = \sum_{i=1}^{99}\left(\overline{f}(x_i) - f(x_i)\right)^2$, the integrated square variance by $ISV = \sum_{i=1}^{99}\frac{1}{n_s}\sum_{s=1}^{n_s}\left(\hat{f}_s(x_i) - \overline{f}(x_i)\right)^2$. Finally the integrated Mean Square Error IMSE is estimated by $IMSE = \sum_{i=1}^{99}\frac{1}{n_s}\sum_{s=1}^{n_s}\left(\hat{f}_s(x_i) - f(x_i)\right)^2$.

3.5.1 Estimation of the marginal surplus function

We assume that the marginal surplus is separable and additive in both the explanatory variables, Z and the unobserved heterogeneity ε like in Equation (3.4.20):

$$p = S'_q(q) + \beta_Z Z + \varepsilon. \tag{3.5.26}$$

We apply Ai and Chen (2003) to estimate the marginal surplus using cubic spline approximation. This is attractive as $S'_q(q)$ can be forced to be decreasing by imposing that the sequence of coefficients related to the spline approximation is also decreasing.

1. We choose the cubic splines $B_4(t)$ as basis for our sieve estimation and renormalize q on $[0, 1]$ (q^* is

the normalized quantity) for the choice of the knots (G denotes the c.d.f. of q):

$$S'_q \circ G^{-1}(q^*) = \phi(q^*) = \sum_{l=-3}^{2^k-1} B_4(2^k q^* - l)$$
$$= P_{q^*}\beta^{K_n},$$

where $K_n = 2^{d_n} + 4$ is the number of parameters that are estimated. d_n is a tuning parameter which drives the approximation of the true (normalized) marginal surplus function.

2. We then project the moment equation

$$m(Z, W, \beta^{K_n}, \beta_Z) = \mathbb{E}\left[p - P_{q^*}\beta^{K_n} - \beta_Z Z | Z, W\right]$$

on a basis of the square-integrable function of Z, W, $p_j(Z, W)$, $j = 1, \ldots, J_n$, where J_n tends slowly to infinity as $n \to \infty$. We denote by $P^{J_n}(Z, W) = (p_1(Z, W), \ldots, p_{J_n}(Z, W))$ and

$$P = (P^{J_n}(Z_1, W_1), \ldots, P^{J_n}(Z_n, W_n)).$$

An empirical estimator of $m(Z, W, \beta^{K_n}, \beta_Z)$ is therefore:

$$\hat{m}(Z, W, \beta^{K_n}, \beta_Z) = \sum_{j=1}^{n} \left(p_j - P_{q_j^*}\beta^{K_n} - \beta_Z Z_j\right) P^{J_n}(Z_j, W_j) \left(P'P\right)^{-} P^{J_n}(Z, W).$$

We can now compute the (empirical) distance function $\hat{Q}(\beta_Z, \beta^{K_n})$ as:

$$\hat{Q}(\beta_Z, \beta^{K_n}) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}(Z_i, W_i, \beta^{K_n}, \beta_Z)\hat{m}(Z_i, W_i, \beta^{K_n}, \beta_Z).$$

We could have used another metric but it appears that the result are not very sensitive to the choice of the metric.

3. Finally, we estimate the parameters β_Z and β_{K_n} by minimizing the penalized criterion:

$$\hat{Q}(\beta_Z, \beta^{K_n}) + \lambda_n(d_0 + d_2).$$

The second term in the expression above is a penalization term added to control for the ill-posedness where $d_0 = \int_0^1 \phi(q^*)^2 dq^*$ and $d_2 = \int_0^1 \phi''(q^*)^2 dq^*$. d_2 is controlling for the oscillations of the estimated marginal surplus (if the monotonicity is not imposed in the estimation step). λ_n is a tuning parameter which controls for the strength of the penalisation term.

Table 3.1 displays the integrated square bias, the integrated variance and the integrated MSE for various choices of d_n, J_n and λ_n.

The results are quite good except for small values of λ_n for which the integrated variance increases a lot. The MSE decreases with the improvement of the sieve approximation (K_n) and the size of the basis used to project the conditional moment on (J_n).

3.5.2 Estimation of the marginal cost function

We estimate the marginal cost function using the SMD procedure of Chen and Pouzo (2009) which is, among additional contributions, a generalization of Ai and Chen (2003) to the case of nonsmooth moments. It is therefore particularly attractive here as we estimate the marginal cost function by quantile IV methods. The marginal cost function $C'_{0q}(q)$ is estimated by another spline approximation. This is a similar procedure than the one explained above. Here the moment condition is

$$\mathbb{E}\left[\mathbf{1}\{\left(p \leq \lambda_\alpha (P_{q^*}\delta^{J_n} + \beta_W W)\right)\} - \alpha | Z, W\right] = 0,$$

for a choice of 10 quantiles (from 0.05 to 0.95) plus the median that is imposed to be equal to one (in a first step). λ_α is an increasing sequence of parameter which corresponds to the value $H(\theta_\alpha, \gamma)$ for the α-quantile of the firm type, θ_α.

From the estimation of the marginal cost, we can now estimate θ in (3.3.9) as all the other quantities can be derived from the first two functions that have been estimated. We finally renormalize θ and the marginal cost function to ensure that the median is equal to 1.

Table 3.2 and 3.3 reports for respectively the estimated marginal cost and distribution of types the integrated square variance, mean bias and MSE for various choices of d_n, J_n and λ_n.

The integrated squared bias is higher than for the estimation of the marginal surplus as the estimation is now based on quantile IV methods. However, the variance is smaller and not very sensitive to the tuning parameters.

Table 3.1: Monte Carlo Study - Integrated MSE of Sieve IV estimator of $S_q'(q)$

	λ_n											
	$K_n = 4$						$K_n = 5$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	0.005	0.002	0.001	0.016	0.016	0.034	0.006	0.002	0.002	0.016	0.015	0.059
	0.230	0.221	0.294	0.974	3.392	4.881	0.256	0.205	0.283	0.955	4.319	9.399
	0.235	0.223	0.295	0.990	3.408	4.915	0.262	0.207	0.284	0.971	4.334	9.457
3	0.008	0.004	0.010	0.028	0.014	0.007	0.010	0.001	0.006	0.020	0.009	0.159
	0.280	0.221	0.327	1.049	1.730	2.356	0.260	0.230	0.277	0.847	2.618	2.939
	0.288	0.225	0.337	1.077	1.745	2.363	0.271	0.231	0.283	0.867	2.628	3.098
6	0.014	0.003	0.019	0.020	0.033	0.030	0.015	0.016	0.024	0.012	0.011	0.051
	0.204	0.171	0.275	0.477	0.602	0.656	0.195	0.179	0.292	0.556	0.884	0.900
	0.219	0.174	0.294	0.497	0.635	0.686	0.210	0.195	0.316	0.568	0.895	0.950
11	0.009	0.018	0.011	0.014	0.036	0.017	0.026	0.015	0.033	0.016	0.022	0.031
	0.132	0.141	0.194	0.269	0.329	0.325	0.121	0.111	0.193	0.294	0.380	0.432
	0.141	0.159	0.205	0.283	0.365	0.342	0.147	0.126	0.226	0.309	0.401	0.463

	λ_n											
	$K_n = 7$						$K_n = 11$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	0.004	0.003	0.008	0.027	0.010	0.324	0.002	0.004	0.018	0.031	0.100	0.056
	0.194	0.219	0.306	0.809	4.223	8.404	0.033	0.049	0.201	0.988	2.477	6.113
	0.198	0.222	0.313	0.835	4.233	8.729	0.036	0.053	0.219	1.019	2.578	6.169
3	0.009	0.014	0.001	0.003	0.131	0.172	0.002	0.005	0.026	0.078	0.042	0.210
	0.205	0.181	0.306	0.940	2.023	5.660	0.023	0.044	0.198	0.757	2.168	6.379
	0.214	0.195	0.307	0.943	2.153	5.831	0.025	0.049	0.224	0.835	2.210	6.589
6	0.012	0.000	0.012	0.029	0.043	0.032	0.004	0.007	0.038	0.030	0.036	0.063
	0.146	0.157	0.298	0.664	1.117	1.477	0.029	0.038	0.177	0.586	1.385	2.182
	0.158	0.158	0.310	0.693	1.160	1.509	0.033	0.045	0.215	0.616	1.421	2.245
11	0.025	0.015	0.028	0.027	0.039	0.026	0.008	0.015	0.024	0.041	0.046	0.020
	0.120	0.135	0.192	0.356	0.532	0.623	0.031	0.050	0.146	0.325	0.725	0.887
	0.146	0.150	0.219	0.384	0.571	0.649	0.039	0.065	0.170	0.366	0.772	0.907

Note: for each value of (λ_n, K_n, J_n) we report in each cell, the integrated squared bias, integrated variance and integrated MSE (divided by 100).

Table 3.2: Monte Carlo Study - Integrated MSE of Sieve IV estimator of $C'_{0q}(q)$

	λ_n											
	$K_n = 4$						$K_n = 5$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	0.19	0.14	0.18	0.12	0.18	0.13	0.14	0.12	0.14	0.17	0.13	0.16
	0.42	0.24	0.36	0.48	0.54	0.29	0.30	0.26	0.28	0.32	0.26	0.32
	0.61	0.39	0.54	0.59	0.72	0.41	0.44	0.39	0.42	0.50	0.39	0.49
3	0.11	0.16	0.12	0.12	0.15	0.17	0.14	0.17	0.13	0.13	0.14	0.16
	0.34	0.41	0.27	0.33	0.34	0.44	0.35	0.34	0.26	0.28	0.18	0.18
	0.45	0.57	0.39	0.45	0.49	0.61	0.48	0.51	0.39	0.41	0.32	0.34
6	0.11	0.12	0.18	0.11	0.11	0.10	0.10	0.10	0.09	0.12	0.07	0.09
	0.21	0.24	0.28	0.21	0.25	0.21	0.19	0.24	0.21	0.23	0.13	0.19
	0.32	0.36	0.46	0.32	0.36	0.31	0.29	0.34	0.30	0.35	0.20	0.28
11	0.16	0.11	0.13	0.11	0.11	0.07	0.08	0.07	0.08	0.07	0.10	0.08
	0.23	0.18	0.23	0.38	0.24	0.24	0.14	0.12	0.21	0.16	0.22	0.20
	0.39	0.30	0.36	0.49	0.35	0.30	0.21	0.20	0.28	0.23	0.32	0.28

	λ_n											
	$K_n = 7$						$K_n = 11$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	1.02	0.78	1.09	0.84	1.26	0.71	2.06	1.97	1.88	2.15	2.09	2.40
	0.95	0.69	1.20	0.79	1.04	0.82	0.99	0.66	0.96	0.80	0.74	1.08
	1.97	1.47	2.30	1.64	2.30	1.53	3.05	2.63	2.83	2.94	2.83	3.48
3	0.73	0.70	0.74	0.97	0.69	0.83	1.54	1.85	2.03	1.65	1.87	1.81
	0.56	0.76	0.91	0.75	0.57	0.92	0.81	0.97	0.84	0.66	1.24	0.86
	1.29	1.46	1.66	1.72	1.26	1.75	2.34	2.82	2.86	2.30	3.11	2.67
6	0.64	0.50	0.53	0.53	0.51	0.52	1.05	0.82	1.08	1.08	1.37	1.19
	0.55	0.44	0.56	0.32	0.52	0.47	0.64	1.66	1.26	0.82	0.59	0.60
	1.18	0.94	1.09	0.85	1.02	0.98	1.69	2.48	2.34	1.90	1.96	1.79
11	0.40	0.42	0.66	0.40	0.38	0.58	0.67	0.89	0.72	0.75	0.90	0.88
	0.40	0.39	0.91	0.41	0.33	0.42	0.75	0.67	1.40	0.80	1.29	0.53
	0.80	0.80	1.57	0.80	0.71	1.00	1.42	1.57	2.11	1.55	2.18	1.41

Note: for each value of (λ_n, K_n, J_n) we report in each cell, the integrated squared bias, the integrated variance and the integrated MSE (divided by 100).

Table 3.3: Monte Carlo Study - Integrated MSE of the nonparametric estimator of $F(\theta)$

	λ_n											
	$K_n = 4$						$K_n = 5$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	0.03	0.03	0.04	0.04	0.05	0.08	0.04	0.05	0.04	0.05	0.08	0.11
	0.03	0.03	0.04	0.05	0.16	0.23	0.03	0.03	0.04	0.06	0.19	0.41
	0.07	0.06	0.08	0.09	0.21	0.31	0.07	0.08	0.08	0.11	0.27	0.51
3	0.03	0.03	0.04	0.04	0.06	0.05	0.04	0.04	0.04	0.06	0.07	0.08
	0.03	0.03	0.04	0.05	0.08	0.12	0.04	0.04	0.03	0.06	0.15	0.15
	0.06	0.06	0.07	0.09	0.14	0.17	0.08	0.07	0.08	0.12	0.22	0.23
6	0.04	0.03	0.03	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.06	0.06
	0.03	0.03	0.03	0.04	0.04	0.05	0.04	0.03	0.03	0.04	0.06	0.08
	0.07	0.06	0.06	0.08	0.08	0.09	0.08	0.07	0.07	0.09	0.12	0.14
11	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.04
	0.02	0.03	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.04	0.05	0.04
	0.05	0.06	0.06	0.07	0.08	0.07	0.07	0.07	0.07	0.09	0.09	0.09

	λ_n											
	$K_n = 7$						$K_n = 11$					
$J_n - K_n$	0.2	0.1	0.01	0.001	1e-4	1e-5	0.2	0.1	0.01	0.001	1e-4	1e-5
1	0.05	0.05	0.05	0.06	0.10	0.20	0.08	0.08	0.07	0.07	0.12	0.25
	0.04	0.03	0.04	0.05	0.22	0.37	0.03	0.02	0.03	0.06	0.21	0.38
	0.09	0.08	0.09	0.11	0.32	0.57	0.11	0.10	0.11	0.13	0.33	0.63
3	0.06	0.05	0.07	0.07	0.09	0.15	0.07	0.07	0.07	0.07	0.16	0.44
	0.03	0.03	0.04	0.07	0.14	0.33	0.03	0.03	0.03	0.05	0.19	0.42
	0.09	0.08	0.11	0.14	0.23	0.49	0.10	0.10	0.10	0.13	0.36	0.87
6	0.06	0.06	0.05	0.06	0.08	0.11	0.08	0.07	0.07	0.08	0.16	0.37
	0.03	0.03	0.04	0.05	0.10	0.15	0.02	0.03	0.03	0.05	0.16	0.24
	0.09	0.08	0.09	0.11	0.18	0.26	0.10	0.10	0.10	0.13	0.32	0.61
11	0.06	0.05	0.05	0.06	0.09	0.08	0.06	0.06	0.08	0.09	0.14	0.22
	0.03	0.02	0.04	0.04	0.06	0.07	0.03	0.03	0.03	0.05	0.14	0.18
	0.09	0.08	0.09	0.10	0.15	0.15	0.09	0.09	0.11	0.13	0.28	0.39

Note: for each value of (λ_n, K_n, J_n) we report in each cell, the integrated squared bias, integrated variance and integrated MSE (divided by 100).

3.5.3 Estimation of the weight parameter γ and the parametric component

Finally we present in Table 3.4 the same results for the parameters involved in the simulation. We only report them for $\lambda_n = 0.2$ or 0.1 and $J_n - K_n = 3$ as the MSE is of the same order of magnitude across the different cases.[25] β_Z is the coefficient related to the parametric component in the marginal surplus, β_W to the component in the marginal cost and γ is the bargaining weight (assumed here unique) estimated from the following equation:

$$H(\theta, \gamma) = \theta + (1 - \gamma)\frac{F(\theta)}{f(\theta)}.$$

Table 3.4: Monte Carlo Study - Estimated bias, variance and MSE of the parameters of the model, $J_n - K_n = 3$.

		λ_n							
		$K_n = 4$		$K_n = 5$		$K_n = 7$		$K_n = 11$	
γ	Bias	0.010	0.059	-0.038	-0.056	-0.074	-0.033	-0.034	-0.034
	Variance	0.035	0.039	0.028	0.028	0.030	0.030	0.028	0.023
	MSE	0.035	0.042	0.001	0.001	0.001	0.001	0.001	0.001
β_Z	Bias	0.008	0.001	0.016	0.003	0.011	0.002	0.003	0.005
	Variance	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000
	MSE	0.001	0.001	0.005	0.006	0.009	0.003	0.002	0.001
β_W	Bias	-0.012	-0.011	-0.008	-0.005	-0.002	0.007	-0.007	-0.011
	Variance	0.009	0.011	0.005	0.004	0.003	0.005	0.001	0.004
	MSE	0.009	0.011	0.005	0.004	0.003	0.005	0.001	0.004

Note: for each value of (λ_n, K_n) we report in each cell, the mean bias, the variance and the MSE.

The estimation of γ is much noisier than for the other parameters but this is still good. In the estimation procedure, we use $K_n = 7$ and $\lambda_n = 0.1$.

[25] The full results are available upon request.

3.6 Empirical Application

We apply the estimation method detailed above to the case of water utilities in France. We first present the data set, before estimating our structural model of regulation. The counterfactuals are presented in the next section.

3.6.1 The Data

The production and distribution of water to households in France are decided at a local level. A survey has been conducted by the French environment institute (IFEN, Institut Français de l'Environnement) amongst local municipalities in metropolitan France.[26] All municipalities with a population of more than 10000 are in the survey and the sampling rate is decreasing with the population. The sample is therefore representative from the population of French local communities (36203 local municipalities in 2004).

We select the observations for which the chosen mode is either operating through a public company (*régie*) or through a private operator with a lease contract (this is the majority of the cases). Then we select the typical contracts, i.e. the two-part tariff contracts (around 95% of the observations). We also merge our dataset with an administrative dataset which reports the median income in a given municipality. For anonymity reasons, it is not reported for city size less than 150 inhabitants. Our sample does not contain such small villages. In the end, we have 3959 observations which each represent one municipality. Table 3.5 presents the management mode of municipalities ranked according to their population. When delegation is the chosen mode, it is in most of the cases conducted by one of the three major firms of the sector: Veolia-Environnement, Lyonnaise des Eaux-Suez Environnement, and Saur-Cise Veolia is the leading firm in the private sector, Saur is more present in rural areas. Observe also that public management is more frequent in small municipalities.

In our dataset, we observe the fix fee of the contract A and the variable part p paid by meter-cube consumed. The overall quantity q is also observed. The characteristics of the network are also given (length and density, quality of the water used as an input, treatment applied before distribution...). Some

[26] Corsica and overseas departments are excluded from our study.

Table 3.5: Management mode by municipality sizes

	Population N of the different municipalities					
	$N < 400$	$400 \leq N$ $N < 1\,000$	$1\,000 \leq N$ $N < 2\,000$	$2\,000 \leq N$ $N < 3\,500$	$3\,500 \leq N$ $N < 10\,000$	$10\,000 \leq N$
Number n of municipalities						
n	813	652	490	356	1210	636
Public	0.53	0.47	0.42	0.33	0.3	0.27
Delegation	0.47	0.53	0.58	0.67	0.7	0.73
Providers in the delegation mode						
Veolia	0.33	0.32	0.37	0.4	0.44	0.49
Suez	0.15	0.2	0.21	0.2	0.24	0.23
Saur	0.36	0.33	0.29	0.27	0.15	0.07
Other	0.16	0.15	0.13	0.13	0.17	0.2

additional information related to the local municipality (population, median income, size of the city, local weather measures) are also recorded. We present in Table 3.6 some descriptive statistics of these variables. For continuous variables, we present the median and the 1st and 3rd quartiles (Q_1 and Q_3) which are respectively robust measures of the location and the dispersion of the empirical distribution. For the two qualitative variables related to the quality of the water and the treatment (they are both recoded with three modalities), we report the empirical frequencies. The fix fee represents around one quarter of the transfer paid to the provider. The variable part is around 1.25 euro per m^3. The total bill charged to the consumer is much higher than what is charged by the firm. First, there are taxes paid to the local water agencies and to the state (value added tax). Then, mechanically, the amount given to the entity in charge of the waste collection is charged proportionally to the quantity consumed. This discrepancy between the prices is taken into account in our analysis.

The price paid by the consumer is generally lower in a local municipalities which are publicly managed. On the other hand, private firms have a higher probability to operate in more dense cities and to distribute water which requires heavier treatment to be drinkable.

Finally, we also collect variables which characterize the environment of a given city, i.e. number of houses, temperature, amount of rain, having in mind that water can be used for watering gardens.

3.6.2 Estimation of the model

We now follow the estimation strategy that is exposed in Section 3.5.

Fitting with Real-World Practices

In the real world, the observed per-unit prices paid by consumers are in fact affected by various taxes. Indeed, we can decompose the price p paid by consumers as:

$$p = (p_1 + p_2)(1 + \tau)$$

where p_1 is the price received by the producer, p_2 is a price paid to another party to finance waste water and τ is an ad-valorem tax imposed by the State. In the sequel, we will take a partial equilibrium

Table 3.6: Descriptive statistics of the main variables

	All mode			Public			Delegation		
	median	Q_1	Q_3	median	Q_1	Q_3	median	Q_1	Q_3
q per household	119.05	98.46	143.55	116.28	96.15	142.79	120.99	99.82	144.37
Fixed fee firm	33.62	19.86	53.84	27	15.54	46.74	37.86	22.63	59.26
Fixed fee	45.2	23.42	77.11	37.83	18.64	66.19	49.23	27	86.29
Fixed fee firm, tax incl.	35.01	20.96	56.35	28.3	16.07	48.87	40.12	23.96	62.59
Fixed fee, tax incl.	47.61	24.82	81	39.38	19.45	69	51.88	28.89	91.02
variable price firm	1.25	1.01	1.53	1.11	0.88	1.35	1.35	1.09	1.62
variable price	2.14	1.48	2.68	1.82	1.17	2.36	2.31	1.76	2.8
variable price, tax incl.	2.24	1.54	2.82	1.89	1.21	2.48	2.42	1.85	2.96
Transfert firm	189.73	155.22	231.96	167.2	133.33	198.23	205.28	170.8	248.95
Bill, tax incl.	323.11	246.01	401.19	272.68	190.74	341.86	353.78	284.34	432.92
Population	2412	540	6383	1214	368	4757	3616	776	7509
Median income	23943	20222	28858	23129	19730	27687	24457	20546	29790
Household size	2.61	2.41	2.78	2.57	2.39	2.75	2.63	2.42	2.8
% of Houses	71.29	53.97	82.38	71.3	55.28	82.38	71.27	52.7	82.38
% of Secondary res	4.39	1.67	12.61	5.88	2.04	15.81	3.64	1.51	10.26
% of Pop under 20	25.03	22.13	27.63	24.55	21.39	27	25.31	22.49	27.99
Temperature	25.1	23.6	26.3	25.1	23.6	26.3	25.1	23.6	26.4
Rain (mm)	194	151	223	200.5	151	251	190	149	212
Sunshine (in h)	581	547	665	591	547	669	579	535	665
Population density	138.87	40.88	480.6	80.37	25.93	261.26	184.27	56.89	631.4
Network characteristics									
network length (in km)	35	14	70	26	9	60	40	18	75
Nb connections	1024	264	2507	600	199	1950	1380	347	2805
Total tank volume	0	0	350	0	0	500	0	0	200
Nb of sensors	1	1	3	1	1	3	1	1	3
	Frequency			Frequency			Frequency		
Basic treatment	0.52			0.53			0.51		
High treatment	0.22			0.13			0.27		
Deep water	0.72			0.77			0.68		
Superficial water	0.28			0.23			0.32		

approach, taking p_2 and τ as given (which may vary across municipalities) and we will sometimes denote $p = P(p_1)$.

Taking into account this specification of the prices, we can rewrite the agent's and the principal's objectives respectively as:

$$\mathcal{U}(\theta, \varepsilon, p_1, A) = A + p_1 D(P(p_1), \varepsilon) - \theta C_0(D(P(p_1), \varepsilon))$$

and

$$\mathcal{W}(\theta, \varepsilon, p, p_1, A) = S(D(P(p_1), \varepsilon), \varepsilon) - A - P(p_1) D(P(p_1), \varepsilon) + \gamma \mathcal{U}(\theta, \varepsilon, p_1, A)$$

or

$$\mathcal{W}(\theta, \varepsilon, p, p_1, A) = S(D(P(p_1), \varepsilon), \varepsilon) - \theta C(D(P(p_1), \varepsilon)) + (p_1 - P(p_1)) D(P(p_1), \varepsilon) - (1-\gamma) \mathcal{U}(\theta, \varepsilon, p_1, A)$$

The fact that $P(p_1) \neq p_1$ creates a discrepancy between consumers expenditures and producer's revenues. Yet, the optimization of the principal's problem proceeds as above and the system (3.4.13)-(3.4.14)-(3.4.15) should now be replaced by the new set of optimality conditions:

$$S'(q) = p, \tag{3.6.27}$$

$$p_1 - \frac{\tau D(p)}{D'_p(p)} = H(\theta, \gamma) C'_{0q}(q), \tag{3.6.28}$$

$$A'(p)(1+\tau) = -q - \left(p_1 - \theta C'_{0q}(q)\right) D'_p(p)(1+\tau). \tag{3.6.29}$$

3.6.3 Estimation of the demand function

The individual demand function is estimated with different specification though we keep working with the log-log model in the following.

$$\log(D(p, Z, \varepsilon)) = \beta_p \log p + \beta'_Z Z + \varepsilon.$$

As usual in the case of a system of simultaneous equations, equilibrium quantities and prices are co-determined and, hence, p is endogenous. To circumvent this issue , we instrument p by a vector W

of exogenous cost shifters ($\mathbb{E}[\varepsilon|Z, W] = 0$) and we estimate the demand parameters by a simple IV regression. The instruments are the following: treatment of the water and characteristics of the water pumped before treatment. These variables increase the cost of distribution by requiring more advanced techniques to make the water drinkable but are independent from the demand shocks. The results are presented in Table 3.7, with standard errors in parenthesis.

The price elasticity is estimated at -0.171 (0.073), which is in the usual intervals estimated for residential water in France (see Reynaud (2003)). Observe that it is the total price charged to the consumer, p, which is used here, not the one charged to the firm, p_1.

Except for the percentage of the houses, the control variables have the expected sign. Richer cities consume more water and the climate conditions matter. Finally, everything else equal, cities with more secondary residences consume less. The Sargan test is not rejected with a p-value of 6%.

3.6.4 Estimation of the cost function

In our data, the size of the different cities vary a lot. Therefore, we make the following assumption about the cost function:

$$C(q_{tot}, W, N) = C_0(q_{tot}/N)N^{\alpha} \exp(\beta_W^{\top} W),$$

in which N is the number of households of the city and W, the other control variables (quality of the water, density of the network, treatment). In the following, $q = q_{tot}/N$ denotes the mean quantity per household consumed. α measures the return to scale to deliver the same quantity of water per household in a bigger city. We should expect α to be lower than one. Fundamentally, we make the assumption that the technology is the same across all providers but that they differ from one city to another through the observed heterogeneity, i.e., quality of the water pumped before treatment, density of the network, level of chemical treatment and through θ which measures the efficiency of the firm in charge of the distribution of the water.

As explained above, Equation 3.6.28 rearranged to take into account the price paid to the other

Table 3.7: Demand estimation

	Dependent variable:	
	q	log(q)
	(1)	(2)
p	-8.837^{*}	
	(4.560)	
log(p)		-0.171^{**}
		(0.073)
income	1.314^{***}	0.012^{***}
	(0.120)	(0.001)
household size	31.682^{***}	0.204^{***}
	(3.729)	(0.031)
% of houses	-0.320^{***}	-0.003^{***}
	(0.050)	(0.0004)
% secondary res	-0.655^{***}	-0.007^{***}
	(0.095)	(0.001)
temperature 2004	1.631^{***}	0.014^{***}
	(0.401)	(0.003)
rain 2004	-0.059^{***}	-0.001^{***}
	(0.010)	(0.0001)
sunshine 2001	0.009^{***}	0.0001^{***}
	(0.003)	(0.00002)
pop 0-20	-0.864^{***}	-0.005^{**}
	(0.241)	(0.002)
constant	40.895^{*}	4.141^{***}
	(23.661)	(0.159)
Observations	3,690	3,690
R^2	0.229	0.249
Adjusted R^2	0.228	0.248
Note:	*p<0.1; **p<0.05; ***p<0.01	

stakeholders and the tax implies that for any quantile $\lambda \in]0,1[$, we have:

$$\mathbb{P}\left(p_1 - \tau \frac{D(p,Z,\varepsilon)}{\partial D/\partial p D(p,Z,\varepsilon)} \leq H(\theta_\lambda,\gamma) C_0'(q) N^{\alpha-1} \exp(\beta_W^\top W) | Z, W, N\right) = \lambda, \tag{3.6.30}$$

with θ_λ being the corresponding quantile of θ and p_1 and p are the prices charged respectively by the firm and the other stakeholders, before tax. $C_{0q}'(q,W)$ is identified up to a scale under an assumption of completeness of Z. In practice, we first estimate $\log C_{0q}'(q,N,W)$ before changing the scale to ensure that the median of θ is equal to 1 for publicly managed firms. Equation 3.6.30 can also be expressed as a conditional moment condition:

$$\mathbb{E}\left(\mathbf{1}\left\{p_1 - \tau \frac{D(p,Z,\varepsilon)}{\partial D/\partial p D(p,Z,\varepsilon)} \leq H(\theta_\lambda,\gamma) C_0'(q) N^{\alpha-1} \exp(\beta_W^\top W)\right\} - \lambda | Z, W, N\right) = 0.$$

This conditional moment equality can be easily be transformed into a moment equality to estimate the marginal cost function and the quantiles $H(\theta_\lambda,\gamma)$ through a GMM procedure. As it has been highlighted in the literature, in practice, the global minimization of the GMM criterion is arduous due to the discontinuous indicator function $\mathbf{1}\{\cdot\}$. Following **?**, we smooth the indicator function in order to facilitate the minimization. In the estimation, 10 quantiles are used[27]. We run simultaneously the estimation for both the private and the public providers assuming that they face the same marginal cost but that each type of firm has its own distribution of θ. We run the estimation 10 times with different starting values and select the estimates which yield the lowest objective function. The results are presented in Table 3.8.

Observe that, as expected, $\hat{\alpha} = 1 - 0.1035 < 1$. Again, the estimates have the expected signs because more treatments increase the cost of the m^3 of water. Pumping deep water is less costly, because on average this water is cleaner.

Estimation of the distribution of types

In this part, we estimate the types θ taking into account the fact that there is no asymmetric information for public firms and that empirical evidence suggest that $\gamma = 0$ when the city designs its menu of

[27] The quantiles are 0.05,0.15,0.25,...

Table 3.8: GMM estimation of $C'_{0q}(q, W)$

	estimate	standard error
constant	-2.9621	4.7e-02
log(q)	0.7806	3.1e-03
log(N)	-0.1035	5.1e-04
basic treatment	0.7625	3.6e-03
high treatment	1.1660	3.6e-03
deep water	-0.6762	2.9e-03
population density	0.0002	3.1e-07

contracts. As a matter of fact, it is politically risky to give a higher rent to the private operator. Having at hand, the estimate of the marginal cost function, we do the following:

1. From the estimates of $C'_{0q}(q)$, we estimate the values θ for the public firms using Equation (3.6.28) with θ in replacement of $H(\theta, \gamma)$.

2. We readjust our estimates to fit with the normalization that the median of the θ is equal to 1. $C'_{0q}(q)$ represents the variable cost function for a median public firm.

3. We estimate
$$H(\theta) = \theta + \frac{F(\theta)}{f(\theta)},$$
from Equation (3.6.28).

4. We invert the expression above to recover θ from the observation of $H(\theta)$. Let $G(\cdot)$ (resp. $g(\cdot)$) the cumulative distribution function (resp. p.d.f) of the $h = H(\theta)$. Standard arguments allow us to back cast θ by expliciting the inverse mapping from $h = H(\theta)$ to θ well-known in the literature on auctions (see, for example Guerre, Perrigne, and Vuong (2000)). $\theta(h) = h - \frac{1}{G(h)} \int_{\underline{h}}^{h} G(x)dx$.

Table 3.9 compare our estimates of the distribution of the type within each population of publicly managed and private firms. The distribution of private types (see Figure 3.2) taking into account the observed heterogeneity and the asymetric information is shifted toward the left and, also, more concentrated.

Table 3.9: Estimation of the distribution of types

	1	2
1%	0.21	0.01
5%	0.41	0.07
15%	0.63	0.23
25%	0.83	0.35
50%	1.35	0.70
75%	2.00	1.14
85%	2.45	1.45
95%	3.61	2.22
99%	5.20	3.51

3.7 Counterfactuals

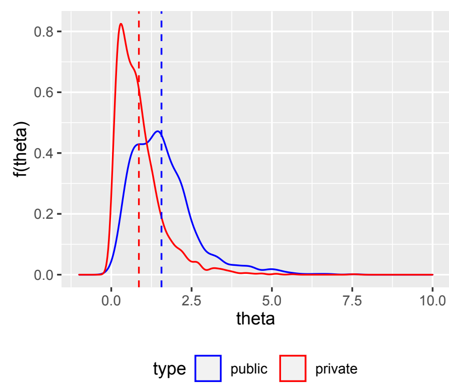
In this part, we run two counterfactuals of interest.

3.7.1 Complete versus Asymmetric Information

3.7.2 Private versus Public Ownership

Upper bound on the value of investment. Denote now by F_{pr} and F_{pu} the cumulative distributions of cost parameters under the private and the public scenario respectively. By investing an amount I ex ante (i.e., before the realization of the cost parameters), a private firm can shift the types distribution from F_{pu}

Figure 3.2: Density of the types public/private



to F_{pr}. As argued by Riordan (1990), Laffont and Tirole (1993) (Chapters 3 and 17) and Schmidt (1996), asymmetric information and the prospects of getting some information rent under private ownership is here the sole engine of investment. Instead, a public firm because it earns no rent under complete information won't invest. The terms of the trade-off are clear. By relying on private firms, the types distribution is shifted to the left while the counterpart of such efficiency gains is that information rent should be given up and prices end up above marginal costs.

Following Riordan (1990), we assume that this investment is non-verifiable, so that although whether it is incurred or not is perfectly anticipated by the local authority at the time of designing the contract, this contract cannot directly influence that decision. Such investment is meant for all the know-how, organizational and technological innovations, and expertise that may be done and that can hardly be explicitly included into a contract.

Our purpose in this section is to get an upper bound on the investment incurred by private firms. Given the contracting scenario so depicted, investment takes place whenever the following incentive constraint holds:

$$E_{\varepsilon}\left(\int_{\underline{\theta}}^{\overline{\theta}}(f_{pr}(\theta)-f_{pu}(\theta))U_{pr}(\theta,\varepsilon)d\theta\right)\geq I.$$

Integrating by parts the left-hand side, we may rewrite this incentive constraint as:

$$E_{\varepsilon}\left(\int_{\underline{\theta}}^{\overline{\theta}}(F_{pr}(\theta)-F_{pu}(\theta))C(D(p_{pr}(\theta),\varepsilon),\varepsilon)d\theta dG(\varepsilon)\right)\geq I.$$

The left-hand side above is the firm's expected gains when shifting the type distribution towards from F_{pu} to F_{pr} when the regulator anticipates that the investment is incurred and, as a result, offers a contract inducing the output profile $D(p_{pr}(\theta),\varepsilon)$ that prevails under a scenario of private ownership. Of course, this term is positive whenever $F_{pr}(\theta)\geq F_{pu}(\theta)\geq 0$ for all θ, a first-order stochastic dominance condition that holds from our previous empirical findings. The right-hand side is just the investment outlay.

Our previous econometric analysis allows us to compute the upper bound on any such investment because all terms on the left-hand side have been previously derived.

3.8 Alternative Formulations

This section discusses alternative formulation for the contracting environment and how our identification procedure applies or not. While Section 3.8.1 highlights environment where our basic identification strategy would still be useful, Section 3.8.3 shows the importance of the kind of contracts observed (relying on fixed fees and per-unit of consumption prices) to get such positive results.

3.8.1 Non-Separability in the Cost Function

Let us now suppose that the cost function is no longer separable and can be more generally written as $C(q,\theta)$, i.e., costs are not necessarily linear in θ. We assume that $C(q,\theta)$ is increasing and concave in q with on top $C'_{\theta}>0$ (operators with lower types produce at lower costs) and the Spence-Mirrlees condition $C''_{q\theta}>0$ (those operators also produce at lower marginal costs) being satisfied. It is routine to check tha the system (3.3.9)-(3.3.11)-(3.3.12) now becomes:

$$A'_{p}(p(\theta,\varepsilon),\varepsilon)=-\Big(p(\theta,\varepsilon)-C'_{q}\left(D(p(\theta,\varepsilon),\varepsilon),\theta\right)\Big)D'_{p}(p(\theta,\varepsilon),\varepsilon)-D(p(\theta,\varepsilon),\varepsilon). \qquad (3.8.31)$$

$$p(\theta,\varepsilon) = C'_q\left(D(p(\theta,\varepsilon),\varepsilon),\theta\right) + (1-\gamma)\frac{F(\theta)}{f(\theta)}C''_{q\theta}\left(\left(D(p(\theta,\varepsilon),\varepsilon)\right),\theta\right), \tag{3.8.32}$$

$$S'_q(D(p(\theta,\varepsilon),\varepsilon),\varepsilon) = p(\theta,\varepsilon), \tag{3.8.33}$$

3.8.2 Identification of a More General Specification

We now prove the identification of this more general model with non-separability in both the surplus and cost functions (Equations (3.8.31)-(3.8.32)-(3.8.33)). First, observe that the identification of the marginal surplus is similar in this general model to what we have done above.

For the marginal cost function, we need an additional normalization like for the surplus function. We shall assume that θ is assumed to be uniformly distributed on $[0,1]$ and we therefore identify the marginal cost function (given W) for the α-th quantile firm in an efficiency ranking (ranked from the most efficient , $\alpha = 0$, to the less efficient, $\alpha = 1$).

Equation (3.8.31) identifies $c'_q\left(q,\alpha,W\right)$ for any quantile α of θ under an assumption of completeness of $\tilde{Z}$ in q. We can indeed derive a moment condition similarly than what have been done in the seprable case earlier. We are indeed able to know, after having derived the marginal surplus function for any quantile of ε the quantity $d = p(\theta,\varepsilon) - \frac{A'_p(p(\theta,\varepsilon),\varepsilon) - D(p(\theta,\varepsilon),\varepsilon)}{D'_p(p(\theta,\varepsilon),\varepsilon)}$, which is equal to the marginal cost (given W) for some (unknown) quantile of θ. Moreover the monotonicity of the marginal cost in θ (given W) ensures that (see above for a similar derivation):

$$\mathbb{P}\left(\left(c'_q(\alpha,q,W) \leq d\right)|Z,W\right) = \alpha. \tag{3.8.34}$$

The completeness assumption ensures therefore the identification of any marginal cost function.

Coming back now to Equation (3.8.32), the only unknown quantity is now γ and we also recover the identification of the bargaining power in the non-separable case.

3.8.3 Contracting on Outputs

The fact that the model can be fully identified (at least when $\gamma = 0$) might seem quite surprising at first glance. Indeed, the literature on the structural estimations of incentive contracts has repeatedly failed to obtain such joint identification of cost functions and types distributions because the data available to the econometrician do not provide enough information. The key difference between our model and those of the existing literature, especially D'Haultfœuille and Février (2020) and Luo, Perrigne, and Vuong (2018), comes actually from the set of contracting variables available under different scenarios that fit with specific institutional environments.

In the regulatory context under scrutiny, for instance, consumers adopt a competitive behavior, expressing individual demand for water at the stipulated unit price while the operator stands ready to supply aggregate demand at that a market clearing price. Instead, had the principal being a single big customer dealing directly with the operator, contracting on the quantity bought would be a feasible option. This is such scenario that is analyzed in D'Haultfœuille and Février (2020).

To facilitate comparison with Section 3.3.1, let us thus assume that contracts are based only on the operator's quantity. In this hypothetical scenario where output could be directly contracted upon, a nonlinear contract might now specify a payment $T(q)$ to the operator if he offers a volume q.[28] For the sake of simplifying exposition, we also assume in this section that there is no fluctuations in demand (i.e., $\varepsilon \equiv 0$) and $\gamma = 0$. Under those conditions, we already know from Section 3.4.1 that our base model is actually identified.

Theoretical results. Mimicking some of the earlier steps of our above analysis, we can rewrite the operator's information rent and optimal supply respectively as:

$$U(\theta) = \max_q T(q) - \theta C_0(q) \tag{3.8.35}$$

and

$$q(\theta) = \arg\max_q T(q) - \theta C_0(q). \tag{3.8.36}$$

From there, it again follows from incentive compatibility that $U(\cdot)$ is absolutely continuous and admits

[28] Observe that, even in this output scenario, the principal still does not observe costs. Otherwise, he could retrieve from the joint observation of costs and outputs information on the agent's type. See Laffont and Tirole (1993) for a theoretical model where such cost observation is available and yet, because of an extra moral hazard variable, asymmetric information still matters. Perrigne and Vuong (2011) for the corresponding empirical study.

an integral representation as:

$$U(\theta) = \int_{\theta}^{\bar{\theta}} C_0\left(q(x)\right) dx \tag{3.8.37}$$

where again, we take into account that the operator's participation constraint is binding at $\bar{\theta}$ for the optimal contract. Observe that this rent is again greater as the operator is requested to produced more.

Under asymmetric information, an optimal contract maximizes the expected welfare of the municipality subject to incentive and participation constraints. From our above observations, that incentive feasible set can be summarized by constraints (3.8.37) and a monotonicity condition (namely $q(\theta)$ weakly decreasing in θ) that will be omitted in a first step and checked ex post on the solution to the solution of the so relaxed problem. Formally, this relaxed problem can be written as:

$$\max_{\{q(\cdot),U(\cdot)\}} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(q(\theta)) - \theta C_0\left(q(\theta)\right) - U(\theta)\right] dF(\theta) \text{ subject to (3.8.37).}$$

Using (3.8.37), integrating by parts as we did above and optimizing pointwise leads to the following expression of the optimal output $q(\theta)$:

$$S'_q(q(\theta)) = \left(\theta + \frac{F(\theta)}{f(\theta}\right) C'_{0q}\left(q(\theta)\right). \tag{3.8.38}$$

The structural model is then defined also by appending to that optimality condition for the principal's problem a second optimality condition related to the agent's problem, namely:

$$T'_q(q(\theta)) = \theta C'_{0q}\left(q(\theta)\right). \tag{3.8.39}$$

From an economic viewpoint, whether outputs as here or prices as in our base line model are the relevant contracting variables leads to the same optimal outputs. The principal's optimality condition (3.8.38) is similar to that obtained in our base model (3.3.11) (for $\gamma = 0$) and downward output distortions also follow from replacing cost parameter by their virtual counterparts. Since optimal outputs in the two models are the same, the agent's information rents, consumer's net surpluses and welfares are also identical. This economic equivalence should come at no surprise. Whether the demand side of the market adopts a "command and control" approach to specify how much should be supplied to satisfy aggregate needs or whether individual consumers are left to express demand is equivalent since that demand side is not the source of any informational asymmetry.

Differences in the scope for identification. Although economically equivalent, the two contracting scenarios differ not only in terms of the set of observables that are available to the econometrician but also in terms of the possibilities left for identification. In other words, whether prices or quantities are controlled gives two alternative implementations of the same allocation which are equivalent from the point of view of the players in the regulatory game which are equally informed on cost and surplus functions and distribution functions.

Yet, those two implementations differ from the perspective of the econometrician. The econometrician must infer from market conduct some information that is common knowledge among players. At a rough level, this outside observer faces a much harder signal extracting problem. In the scenario where quantities are directly controlled, no information on marginal surplus can be learned from the firm's choices. Instead, when only prices are controlled, information on marginal surplus gets communicated to outside observer. This simple fact allows identification of the model.

More precisely, for the output scenario, the only observables are the output q and the overall payment $T(q)$. As in our base model, S'_q and C'_{0q} are not observed in the data so that the principal's optimality condition (3.8.38) cannot be used to retrieve information either on types or on their distribution. On top, per-unit consumption prices are by construction not observed, a key difference with our base model.

Defining an implicit per-unit consumption price for this output scenario as $p(\theta) = S'_q(q(\theta))$, we may follow a procedure similar to that developed in our base model and introduce a fictitious price-cost margin. Accordingly, we thus rewrite the principal's optimality condition (3.8.38) as a familiar condition:

$$\frac{p(\theta) - \theta C'_{0q}(q(\theta))}{p(\theta)} = r(\theta, 0).$$

Even if such price is not observed, one could hope as in our base line model to retrieve the relevant information from the agent's optimality condition (3.8.39). However, this step is no longer possible here because the agent's optimality condition (3.8.39) now refers only to the marginal nonlinear price $T'_q(q(\theta))$ and not on the price as in our baseline scenario. Indeed, this marginal nonlinear price $T'_q(q(\theta))$ always differs from the per-unit consumption price $p(\theta) = S'_q(q(\theta))$ under asymmetric information, there is always a wedge between $T_q'(q(\theta))$ and $p(\theta) = S_q'(q(\theta))$, as:

$$p(\theta) - T_q'(q(\theta)) = \frac{F(\theta)}{f(\theta)} C_{0q}'(q(\theta)) > 0.$$

Thus, the agent's optimality condition does not bring new information on marginal surplus at the equilibrium point, making it impossible to identify this function.

In response to such impossibility, different alleys have been pursued in the literature. Luo, Perrigne, and Vuong (2018) consider a model of "false moral hazard"[29] where costs, which results from innate types but also no verifiable effort variables, can be observed. Such observation of course brings extra information. Although positive results are now obtained, the analysis remains complex because the additional effort variable that characterizes such model must also be disentangled from innate types. Developing an approach tailored to the specificities of their data set, D'Haultfœuille and Février (2020) work with some discrete heterogeneity on the demand side while still assuming that such information is observed by the econometrician.[30] Three points are in this case sufficient to recover identification. Finally, Luo, Perrigne, and Vuong (2018) consider a nonlinear pricing model with fixed quantity which *de facto* determines marginal surplus at the equilibrium point. Let us conclude this section by pointing out an alternative route. Indeed, parametric assumptions on both the surplus function and the cost function, or the cost function and the firms' types could help identifying the model but, of course, at a risk of misspecification as these assumptions cannot be tested.

3.9 Conclusion

In this paper, we consider a principal-agent model à la Baron-Myerson. We show that the contract negociation can be modeled by a choice within a menu of options of a two-part tariff scheme. Interestingly, we show that the model is non parametrically identified from the observation of the tariff (fixed fee and per unit price) and the quantity consumed.

[29] In the parlance of the incentive literature, see Laffont and Martimort (2009) (Chapter 7).

[30] This is in striking contrast with our assumption in the base line model that such heterogeneity is actually not observed.

We apply our methodology to the distribution of water in France, in which we quantify the amount of asymmetric information and its impact on the per-unit price charged.

Some analysis remain to be done. First, we would like to design a more flexible approach than the parametric estimates, borrowing from the recent literature on nonparametric quantile IV. Then, we proved that we the dsitribution of the γs could be identified. However it relies on the estimation of derivatives. We need to find a stable way to estimate it.

Bibliography

AI, C. AND X. CHEN (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1843.

ANDREWS, D. W. (2011): "Examples of l2-complete and boundedly-complete distributions," .

ARMSTRONG, M. AND D. E. SAPPINGTON (2007): "Recent developments in the theory of regulation," *Handbook of industrial organization*, 3, 1557–1700.

BAGNOLI, M. AND T. BERGSTROM (2006): "Log-concave probability and its applications," in *Rationality and Equilibrium: A Symposium in Honor of Marcel K. Richter*, Springer, 217–241.

BARON, D. P. (1988): "Regulation and legislative choice," *The RAND Journal of Economics*, 467–477.

——— (1989): "Design of regulatory mechanisms and institutions," *Handbook of industrial organization*, 2, 1347–1447.

BARON, D. P. AND R. B. MYERSON (1982): "Regulating a monopolist with unknown costs," *Econometrica: Journal of the Econometric Society*, 911–930.

BASOV, S. (2005): "The Multidimensional Screening Model," *Multidimensional Screening*, 135–175.

BHATTACHARYYA, A., T. R. HARRIS, R. NARAYANAN, AND K. RAFFIEE (1995): "Specification and estimation of the effect of ownership on the economic efficiency of the water utilities," *Regional science and urban Economics*, 25, 759–784.

BHATTACHARYYA, A., E. PARKER, AND K. RAFFIEE (1994): "An examination of the effect of ownership on the relative efficiency of public and private water utilities," *Land Economics*, 197–209.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75, 1613–1669.

BROCAS, I., K. CHAN, AND I. PERRIGNE (2006): "Regulation under asymmetric information in water utilities," *American Economic Review*, 96, 62–66.

CANAY, I. A., A. SANTOS, AND A. M. SHAIKH (2013): "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 81, 2535–2559.

CHEN, X. AND D. POUZO (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152, 46–60.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

CHESHER, A. (2007): "Identification of nonadditive structural functions," *ECONOMETRIC SOCIETY MONOGRAPHS*, 43, 1.

CHIAPPORI, P.-A. AND B. SALANIE (2000): "Testing for asymmetric information in insurance markets," *Journal of political Economy*, 108, 56–78.

DAROLLES, S., Y. FAN, J.-P. FLORENS, AND E. RENAULT (2011): "Nonparametric instrumental regression," *Econometrica*, 79, 1541–1565.

D'HAULTFŒUILLE, X. AND P. FÉVRIER (2020): "The provision of wage incentives: A structural estimation using contracts variation," *Quantitative Economics*, 11, 349–397.

D'HAULTFOEUILLE, X. (2011): "On the completeness condition in nonparametric instrumental problems," *Econometric Theory*, 27, 460–471.

ESTACHE, A. AND M. A. ROSSI (1999): *Comparing the performance of public and private water companies in the Asia and Pacific region: what a stochastic costs frontier shows*, vol. 2152, World Bank Publications.

FLORENS, J.-P., J. JOHANNES, AND S. VAN BELLEGEM (2012): "Instrumental regression in partially linear models," *The Econometrics Journal*, 15, 304–324.

GAGNEPAIN, P. AND M. IVALDI (2002): "Incentive regulatory policies: the case of public transit systems in France," *RAND Journal of Economics*, 605–629.

GAGNEPAIN, P., M. IVALDI, AND D. MARTIMORT (2013): "The cost of contract renegotiation: Evidence from the local public sector," *American Economic Review*, 103, 2352–2383.

GARCIA, S. AND A. THOMAS (2001): "The structure of municipal water supply costs: application to a panel of French local communities," *Journal of Productivity analysis*, 16, 5–29.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first-price auctions," *Econometrica*, 68, 525–574.

GUESNERIE, R. AND J.-J. LAFFONT (1984): "A Complete Solution to a Class of Principal-Agent Probems with an Application to the Control of a Self-Managed Firm," *Journal of public Economics*, 25, 329–369.

HALL, D. C. (2000): "Public choice and water rate design," *The political economy of water pricing reforms*.

IVALDI, M. AND D. MARTIMORT (1994): "Competition under nonlinear pricing," *Annales d'Economie et de Statistique*, 71–114.

LAFFONT, J.-J. AND D. MARTIMORT (2009): "The theory of incentives," in *The Theory of Incentives*, Princeton university press.

LAFFONT, J.-J. AND J. TIROLE (1993): *A theory of incentives in procurement and regulation*, MIT press.

LAVERGNE, P. AND A. THOMAS (1997): "Semiparametric estimation and testing in models of adverse selection, with an aplication to environmental regulation," .

LUO, Y., I. PERRIGNE, AND Q. VUONG (2018): "Structural analysis of nonlinear pricing," *Journal of Political Economy*, 126, 2523–2568.

MYERSON, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of mathematical economics*, 10, 67–81.

NEWEY, W. K. AND J. L. POWELL (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

NÖLDEKE, G. AND L. SAMUELSON (2005): "Optimal bunching without optimal control," *Journal of Economic Theory*.

PERRIGNE, I. AND Q. VUONG (2011): "Nonparametric identification of a contract model with adverse selection and moral hazard," *Econometrica*, 79, 1499–1539.

REYNAUD, A. (2003): "An econometric estimation of industrial water demand in France," *Environmental and Resource Economics*, 25, 213–232.

RIORDAN, M. H. (1990): "Asset specificity and backward integration," *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 133–146.

ROCHET, J.-C. (1985): "The taxation principle and multi-time Hamilton-Jacobi equations," *Journal of Mathematical Economics*, 14, 113–128.

SALANIÉ, B. (2005): *The economics of contracts: a primer*, MIT press.

THOMAS, A. (1995): "Regulating pollution under asymmetric information: The case of industrial wastewater treatment," *Journal of environmental economics and management*, 28, 357–373.

WOLAK, F. A. (1994): "An econometric analysis of the asymmetric information, regulator-utility interaction," *Annales d'Economie et de Statistique*, 13–69.