Are unobservables separable?*

Andrii Babii† — UNC Chapel Hill
Jean-Pierre Florens‡ — Toulouse School of Economics

January 28, 2020

Abstract

It is common to assume in empirical research that observables and unobservables are additively separable, especially, when the former are endogenous. This is done because it is widely recognized that identification and estimation challenges arise when interactions between the two are allowed for. Starting from a nonseparable IV model, where the instrumental variable is independent of unobservables, we develop a novel nonparametric test of separability of unobservables. The large-sample distribution of the test statistics is nonstandard and relies on a novel Donsker-type central limit theorem for the empirical distribution of nonparametric IV residuals. Using a dataset drawn from the 2015 US Consumer Expenditure Survey, we find that the test rejects the separability in Engel curves for most of the commodities.

Keywords: unobservables, endogeneity, separability test, nonparametric IV regression, nonparametric IV residuals, Engel curves.
JEL Classifications: C12, C14, C26

1 Introduction

It is common to assume in empirical research that observables and unobservables are additively separable, especially when the former are endogenous. This is done because it is widely recognized that identification and estimation challenges arise when

*We thank Ivan Canay, Tim Christensen, Elia Lapenta, Pascal Lavergne, Thierry Magnac, Nour Meddahi, and Ingrid Van Keilegom for helpful discussions. All remaining errors are ours.

†Department of Economics, University of North Carolina–Chapel Hill - Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. Email: babii.andrii@gmail.com.

‡Department of Economics, Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). Email: jean-pierre.florens@tse-fr.eu.

interactions between the two are allowed for.[1] However, the economic theory and considerations often may lead to nonseparable models. Prominent examples are demand functions, where price or income effects might be heterogeneous in unobserved preferences; production functions, where observed input choices may be heterogeneous in input choices unobserved by the econometrician; labor supply functions with heterogeneous wage effects; wage equations, where the returns to schooling might vary with unobserved ability; or more generally, treatment effect models, where causal effects may be heterogeneous in unobservables.

Since a fully nonparametric estimation of the nonseparable model is more challenging and, at the same time, the separability may rule out the heterogeneity of marginal effects due to unobservables, detecting separability is desirable in empirical applications. If the separability is rejected, then the more sophisticated nonseparable models should not be neglected, while if it turns out that the structural relation is separable, then the conventional empirical practice is well-justified.

Despite the significant efforts focused on understanding the identification and the estimation of nonseparable IV models and the popularity of separable IV models in the empirical practice, little work has been done on developing formal testing procedures that could discriminate empirically between the two. Lu and White (2014) and Su, Tu, and Ullah (2015) are notable exceptions that develop separability tests under the *conditional independence restriction* and additional identifying restrictions imposed by the nonseparable model. The conditional independence restriction is different from the *mean-independence restriction* imposed by the separable nonparametric IV model and does not allow justifying the separable nonparametric IV model that we are interested in here. Other recent specification tests for the nonseparable model include a monotonicity test of Hoderlein, Su, White, and Yang (2016), an endogeneity test of Fève, Florens, and Van Keilegom (2018), and a specification test for the quantile IV regression of Breunig (2020).

In this paper, we design a novel fully nonparametric separability test. Our test does not rely on the specific identifying restrictions imposed by the nonseparable model, such as the monotonicity or the rank similarity. The test is based on the insight that the structural function in the separable model can be estimated us-

[1] The nonparametric identification of nonseparable models with endogeneity and independent instrumental variables under different sets of additional restrictions attracts significant attention; see Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Torgovitsky (2015), and D'Haultfœuille and Février (2015) among others. The nonparametric estimation of the nonseparable model may lead to a difficult nonlinear ill-posed inverse problem; see Carrasco, Florens, and Renault (2007), Horowitz and Lee (2007), Gagliardini and Scaillet (2012), and Dunker, Florens, Hohage, Johannes, and Mammen (2014).

ing the nonparametric IV approach.[2] If the separable model is correct, then the nonparametric IV residuals should approximate unobservables independent of the instrumental variables in the nonseparable IV model. This intuition suggests that it should be possible to detect the separability with the classical Kolmogorov-Smirnov or Cramér-von Mises independence tests between the nonparametric IV residuals and the instrumental variable. To the best of our knowledge, no such test is currently available in the literature, and it is not known whether the empirical distribution of the nonparametric IV residuals satisfies the Donsker property.

Formalizing this intuition is far from trivial, since the regression residuals are different from the true regression errors and the nonparametric IV regression is an example of an *ill-posed inverse* problem and requires regularization. Moreover, the empirical distribution function of the nonparametric IV residuals is a *non-smooth* function of the estimated nonparametric IV regression. The weak convergence of the empirical distribution of regression residuals in the parametric linear case is a classical problem in statistics; see, e.g., Durbin (1973), Loynes (1980), and Mammen (1996). The extension to the nonparametric regression is more complex, and it is remarkable that the empirical distribution of nonparametric regression residuals still converges weakly as was shown in Akritas and Van Keilegom (2001).[3] The limiting Gaussian process often has a covariance function contaminated by the uncertainty coming from the fact that residuals are different from the true regression errors.[4]

The additively separable nonparametric IV regression differs from the problems discussed above in two important directions. First, its finite-sample and the asymptotic performance depend both on the smoothness of the regression function and the smoothing properties of the conditional expectation operator. Second, it features an additional dependence between the endogenous regressor and the regression error that cannot be neglected in large-sample approximations. We show that the empirical distribution of the nonparametric IV residuals satisfies the Donsker property. To obtain this result, we rely on the insight that the Tikhonov regularization in Sobolev spaces, studied recently in Florens, Johannes, and Van Bellegem (2011), Gagliar-

[2]See Florens (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), and Darolles, Fan, Florens, and Renault (2011).

[3]Note that the *weak convergence* of the empirical process based on estimated nuisance parameters is more difficult to achieve than the asymptotic distribution of a *low-dimensional* parameter in the presence of nonparametric or high-dimensional nuisance components; see, e.g., Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) for a recent treatment of the latter and further references.

[4]Interestingly, Einmahl and Van Keilegom (2008) show that the independence empirical process based on the nonparametric regression residuals may lead to a distribution-free test, cf., Blum, Kiefer, and Rosenblatt (1961).

dini and Scaillet (2012), Carrasco, Florens, and Renault (2014), and Gagliardini and Scaillet (2017),[5] provides a natural link between the modern empirical process theory and the theory of ill-posed inverse problems. Interestingly, our test statistics have nonstandard limiting distributions that are not free from the nonparametric nuisance components. We find in Monte Carlo experiments that the standard bootstrap fails, and we use the subsampling to compute the critical values.

The paper is organized as follows. In Section 2, we present two motivating examples, where economic considerations lead to nonseparable models with endogeneity and introduce our separability test. In Section 3, we characterize the large sample approximation to the distribution of the residual-based independence test statistics and introduce a resampling procedure to compute its critical values. The results of Section 3 follow to a large extent from a more fundamental result on the uniform asymptotic expansion and the Donsker CLT of the empirical distribution of the nonparametric IV residuals presented in Section 4. The latter provides all the main insights to the problem in a simplified setting. We report on a Monte Carlo study in Section 5 which provides insights about the validity of our asymptotic approximations in finite samples. In Section 6, we test the separability of Engel curves for a large set of commodities and find that the separability is rejected most of the time. Conclusions appear in Section 7. All technical details and proofs are collected in the Appendix. In Appendix A.1, we discuss the regularization of the nonparametric IV regression in Sobolev scales. Appendix A.2 provides proofs of all results stated in the main part of the paper.

2 Separability of unobservables

2.1 Two motivating examples

The instrumental variable models with additively separable unobservables constitute a workhorse of the modern empirical practice. At the same time, the economic theory often suggests that unobservables may be functionally related to observables and that the economic effects may be heterogeneous in unobservables; see, e.g., Heckman (2001) or Imbens (2007). Structural economic modeling typically leads to nonseparable unobservables as illustrated below.

[5] See also Blundell, Chen, and Kristensen (2007), who consider sieve regularization in Sobolev spaces without the Tikhonov penalization. It is an open question whether the sieve nonparametric IV estimator admits uniform asymptotic expansions similar to ours.

Example 2.1 (Demand function)**.** *Consider a random utility maximization problem*

$$Q = \underset{q \in \mathbf{R}^J :\, P^\top q = I}{\operatorname{argmax}} \; U(q, \varepsilon),$$

where $U(.,.)$ is a utility function, Q is a vector of demanded quantities, ε is an individual preference variable, unobserved by the econometrician, P is a vector of prices, and I is an income. The solution to this optimization problem leads to the nonseparable demand functions $Q_j = \Phi(P, I, \varepsilon)$ for each good $j = 1, \ldots, J$, see Brown and Walker (1989) for more details and Hoderlein and Vanhems (2010) for the welfare analysis based on the nonseparable model. The nonseparable demand functions may lead in turn to the nonseparable Engel curves.

Example 2.2 (Production function/frontier)**.** *Simar, Vanhems, and Van Keilegom (2016) consider a production process with unobserved heterogeneity that leads to the production function/frontier ϕ such that $Y = \phi(Z, \varepsilon) - U$, where Y is an output, Z are observed inputs, ε is an environmental factor, and $U \geq 0$ is a measure of inefficiency. In this example, the nonseparable model is generated by the fact that the environmental factor is taken into account along with other input choices by firms and, at the same time, the former is not observed by the econometrician.*

2.2 Separability test

Let (Y, Z, W) be observed random variables admitting a nonseparable representation

$$Y = \Phi(Z, \varepsilon), \qquad \varepsilon \perp\!\!\!\perp W,$$

where Y is a scalar outcome, $Z \in \mathbf{R}^p$ are regressors, ε is a scalar unobservable,[6] $W \in \mathbf{R}^q$ is a vector of instrumental variables, and $\Phi : \mathbf{R}^p \times \mathbf{R} \to \mathbf{R}$ is a structural function. We assume that W are valid instrumental variables satisfying the exclusion restriction, $\varepsilon \perp\!\!\!\perp W$, and the relevance condition, $W \not\perp\!\!\!\perp Z$.

A separability test in this setting can be developed using different approaches. For instance, one could nonparametrically identify and estimate the nonseparable model and check the separability, e.g., with a derivative-based statistics. This approach corresponds to the principle behind the Wald test for parametric models. Alternatively, since the nonparametric identification and estimation of the separable model is easier, one could estimate the separable model and check the independence

[6]The multidimensional unobservables $\xi \in \mathbf{R}^r$ can be easily accommodated using the index structure $\varepsilon = \xi^\top \gamma$ for some $\gamma \in \mathbf{R}^r$, where ξ may contain product terms to allow for functional interactions between the unobserved components.

condition of the nonseparable model. This approach corresponds to the principle behind the Rao's score test in the parametric setting and is the one adopted in this paper.

In the nonparametric IV regression, (Y, Z, W) admit a separable representation

$$Y = \varphi(Z) + V, \qquad \mathbb{E}[V|W] = 0$$

for some structural function $\varphi : \mathbf{R}^p \to \mathbf{R}$ and unobservable V. The mean-independence IV restriction implies that the nonparametric IV regression solves the functional equation

$$\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W].$$

Formally, we are interested in testing the null hypothesis

$$H_0 : \quad \Phi(Z, \varepsilon) = \psi(Z) + g(\varepsilon), \qquad \text{for some measurable} \quad \psi, g$$

against the alternative hypothesis, denoted H_1, that this is not the case.

To test H_0, we assume that the distribution of (Z, W) is such that the conditional expectation operator $\phi \mapsto \mathbb{E}[\phi(Z)|W]$ is injective on the space of square-integrable functions. Let φ be the solution to $\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W]$ and put $U \triangleq Y - \varphi(Z)$. Note that U is a well-defined random variable and that $\mathbb{E}[U|W] = 0$, even if the model is nonseparables. Under H_0,

$$Y = \psi(Z) + g(\varepsilon)$$

for some functions ψ, g and $\varepsilon \perp\!\!\!\perp W$. By the independence property, $\mathbb{E}[Y|W] = \mathbb{E}[\psi(Z) + \mathbb{E}g(\varepsilon)|W]$, whence we can identify $\varphi(Z) = \psi(Z) + \mathbb{E}g(\varepsilon)$ and $U = g(\varepsilon) - \mathbb{E}g(\varepsilon)$ by the uniqueness of φ. Since $\varepsilon \perp\!\!\!\perp W$, under H_0, we obtain a testable implication that $U \perp\!\!\!\perp W$. Note that under the alternative hypothesis when the model is nonseparable, we have $U = h(Z, \varepsilon)$, for some non-degenerate function h of (Z, ε), which is typically not independent of W, since W are not independent of Z (relevance condition).

Therefore, we can reduce the problem of testing the separability of unobservables to the problem of testing the independence between the instrumental variables W and the nonparametric IV residual U. Our test relies on the completeness condition commonly used in the nonparametric IV literature, cf., Babii and Florens (2017) and references therein, and does not require identifying the nonseparable model; see, e.g., Chernozhukov and Hansen (2005).

3 Large sample distribution

3.1 Distribution of statistics

Under the mean-independence IV condition, the structural function φ solves the ill-posed functional equation

$$r(w) \triangleq \mathbb{E}[Y|W=w]f_W(w) = \int \varphi(z)f_{ZW}(z,w)\mathrm{d}z \triangleq (T\varphi)(w),$$

where $T : L_2(\mathbf{R}^p) \to L_2(\mathbf{R}^q)$ is an integral operator. We focus on the Tikhonov-regularized estimator $\hat{\varphi}$ solving the penalized least-squares problem

$$\hat{\varphi} = \arg\min_{\phi \in H^s} \left\| \hat{T}\phi - \hat{r} \right\|^2 + \alpha_n \|\phi\|_s^2, \tag{1}$$

where $(\hat{T}, \hat{r})$ are kernel-smoothed estimators of (T, r), computed using a sample $(Y_i, Z_i, W_i)_{i=1}^n$, $(H_s, \|.\|_s)$ is a Sobolev space, and $\|.\|$ is the L_2 norm, see Appendix A.1 for more details.

Following the discussion in Section 2, we consider the independence test between the nonparametric IV residuals and the instrumental variables. Since the nonparametric IV estimator $\hat{\varphi}$ is consistent for φ, the difference between the estimated residuals $\hat{U}_i = Y_i - \hat{\varphi}(Z_i)$ and the nonparametric IV regression error $U_i = Y_i - \varphi(Z_i)$ should become negligible asymptotically. The test-statistics can be built around the following residual-based independence empirical process

$$\mathbb{G}_n(u,w) = \sqrt{n}\left(\hat{F}_{\hat{U}W}(u,w) - \hat{F}_{\hat{U}}(u)\hat{F}_W(w)\right),$$

where $\hat{F}_{\hat{U}W}$ is the empirical distribution function of $(\hat{U}_i, W_i)_{i=1}^n$ and $\hat{F}_{\hat{U}}$ and $\hat{F}_W$ are corresponding marginal empirical distribution functions. We will also use f_{UZW} to denote the joint density of (U, Z, W).

The Donsker central limit theorem cannot be applied directly to the process $\mathbb{G}_n$, since the process is based on the pseudo-observations $\hat{U}_i$. Moreover, the empirical distribution function depends on residuals in a non-smooth way. Using residuals instead of the true regression errors leads to the parameter uncertainty problem, and as the following result shows, it affects the large-sample distribution of the independence empirical process.

Theorem 3.1. *Suppose that Assumptions A.1.1, A.1.2, and A.2.1 are satisfied. Then uniformly over $(u,w) \in \mathbf{R} \times \mathbf{R}^q$*

$$\mathbb{G}_n(u,w) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{ \mathbf{1}_{\{U_i \le u, W_i \le w\}} - \mathbf{1}_{\{U_i \le u\}}F_W(w) - \mathbf{1}_{\{W_i \le w\}}F_U(u) + F_{UW}(u,w) + \delta_i(u,w) \right\} + o_P(1),$$

with

$$\delta_i(u,w) = U_i\left(T(T^*T)^{-1}g(u,w,.)\right)(W_i),$$
$$g(u,w,z) = \int^w f_{UZW}(u,z,v)\mathrm{d}v - f_{UZ}(u,z)F_W(w).$$

Note the parameter uncertainty problem disappears when $U \perp\!\!\!\perp (Z,W)$, in which case $g = 0$. Heuristically, this explains why the endogeneity in the IV model leads to the independence empirical process that behaves differently from the one in a simpler nonparametric conditional mean regression, cf., Einmahl and Van Keilegom (2008).

As a consequence of Theorem 3.1, we obtain the following Donsker-type central limit theorem.

Corollary 3.1. *Suppose that assumptions of Theorem 3.1 are satisfied. Then under the null hypothesis*

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \qquad \text{in} \qquad L_\infty(\mathbf{R}\times\mathbf{R}^q),$$

where $(u,w) \mapsto \mathbb{G}(u,w)$ is a tight centered Gaussian process with uniformly continuous sample paths and the covariance function

$$(u,w,u',w') \mapsto \mathbb{E}\Big[\left(\mathbf{1}_{\{U\leq u, W\leq w\}} - \mathbf{1}_{\{U\leq u\}}F_W(w) - \mathbf{1}_{\{W\leq w\}}F_U(u) + F_{UW}(u,w) + \delta(u,w)\right)\times$$
$$\times\left(\mathbf{1}_{\{U\leq u', W\leq w'\}} - \mathbf{1}_{\{U\leq u'\}}F_W(w') - \mathbf{1}_{\{W\leq w'\}}F_U(u') + F_{UW}(u',w') + \delta(u',w')\right)\Big].$$

Corollary 3.1 allows us to characterize a large-sample approximations to the distribution of the Cramér-von Mises and the Kolmogorov-Smirnov test statistics, defined as

$$T_{n,2} = \iint |\mathbb{G}_n(u,w)|^2\mathrm{d}\hat{F}_{\hat{U}W}(u,w) \qquad \text{and} \qquad T_{n,\infty} = \sup_{u,w}|\mathbb{G}_n(u,w)|.$$

To describe the behavior of both statistics under the alternative hypothesis, put also

$$d_2 = \iint |F_{UW}(u,w) - F_U(u)F_W(w)|^2\mathrm{d}F_{UW}(u,w)$$
$$d_\infty = \sup_{u,w}|F_{UW}(u,w) - F_U(u)F_W(w)|.$$

Corollary 3.2. *Suppose that assumptions of Theorem 3.1 are satisfied. Then under the null hypothesis*

$$T_{n,2} \xrightarrow{d} \iint |\mathbb{G}(u,w)|^2\mathrm{d}F_{UW}(u,w) \qquad \textit{and} \qquad T_{n,\infty} \xrightarrow{d} \sup_{u,w}|\mathbb{G}(u,w)|,$$

while under the alternative hypothesis $T_{n,2} \xrightarrow{\text{a.s.}} \infty$, provided that $d_2 > 0$ and $T_{n,\infty} \xrightarrow{\text{a.s.}} \infty$, provided that $d_\infty > 0$.

The asymptotic distribution in both cases is not pivotal, in contrast to the nonparametric regression without endogeneity cf., Einmahl and Van Keilegom (2008). While obtaining the distribution-free statistics is possible in simpler residual-based testing problems, see Escanciano, Pardo-Fernández, and Van Keilegom (2018); these methods do not seem to extend naturally to our setting. Therefore, the bootstrap seems to be an attractive alternative that can be used to obtain critical values of our separability test. Note that for the Kolmogorov-Smirnov statistics, one could also use the bootstrap approximations of Chernozhukov, Chetverikov, and Kato (2014) and Chernozhukov, Chetverikov, and Kato (2016) instead of the weak convergence arguments. Interestingly, as we discuss in the following section, the naive nonparametric and the multiplier bootstrap do not work.

3.2 Critical values

The asymptotic distributions in Corollary 3.1 are nonstandard and depend on several nuisance nonparametric components. This calls for resampling methods to compute the critical values. As can be seen from the proof of Theorem 3.1, our uniform asymptotic expansion relies on the differentiability of the CDF. This leads to a dependence of the asymptotic distribution on the probability density function f_{UZW} in Corollary 3.1; see also the proof of Theorem 4.1 and Corollary 4.1. Such uniform asymptotic expansion cannot be obtained in the same way for the bootstrapped statistics since in the bootstrap world the empirical distribution function is not differentiable.

The lack of smoothness of the empirical distribution function suggests that the standard bootstrap procedures may fail in approximating the asymptotic distribution of the test statistics. The problem of a similar nature occurs with the bootstrap of the cube-root consistent estimators; see, e.g., Babii and Kumar (2019) and references therein. Another complication with the bootstrap is that we typically need to resample from the distribution obeying the constraints of the null hypothesis and that the validity of the bootstrap has to be established case-by-case.[7]

Consequently, we suggest relying on the subsampling or the m out of n bootstrap to compute the critical values of the test. The resampling procedure is as follows:

1. Draw a sample of size m from $(Y_i, Z_i, W_i)_{i=1}^n$ without replacements (subsampling) or with replacements (m out of n bootstrap), where $m = m_n$ is a sequence

[7]Note that the (smoothed) residual bootstrap, cf., Neumeyer and Van Keilegom (2019), does not preserve the dependence between the endogenous regressor and the unobservables and does not mimic the data generating process of the IV regression under the null hypothesis. In Section 5, we find in Monte Carlo experiments that the standard nonparametric bootstrap does not work.

such that $m_n \to \infty$ and $m_n/n \to 0$ and as $n \to \infty$.

2. Compute the Kolmogorov-Smirnov or the Cramér-von Mises statistics using the simulated sample.

3. Repeat the first two steps many times and compute the critical values using empirical quantiles of the statistics over all simulated samples.

An attractive feature of the subsampling is that it is valid for general hypothesis testing problems; see Politis, Romano, and Wolf (2001), Theorem 3.1., and there is no need to show its validity in each specific application. An adaptive data-driven rule to select m_n is considered, e.g., Bickel and Sakov (2008).

4 Nonparametric IV regression residuals

In this section, we investigate the behavior of the empirical distribution function of the nonparametric IV residuals. The uniform asymptotic expansion and the Donsker central limit theorem obtained in this section are fundamental to our results in the previous section and illustrate all the main issues. Let ∂_u be a partial derivative with respect to the variable u (corresponding to U). We impose the following conditions on the distribution of the data and tuning parameters.

Assumption 4.1. *(i) $\|\partial_u f_{UZ}\|_\infty < \infty$ and $\sup_u \|f_{UZ}(u,.)\|_\kappa < \infty$ with $\kappa > 2a$; (ii) $h_n \to 0$ and $\alpha_n \to 0$ are such that $\alpha_n^{\frac{4a}{b+a}} n h_n^{2q} \to \infty$, $\alpha_n^{\frac{2(a+c)}{b+a}} n h_n^q \to \infty$, $\alpha_n^{-\frac{4a}{b+a}\vee 1} n h_n^{4t} \to 0$, $\alpha_n^2 n \to 0$, $\alpha_n n h_n^{p+2q} \to \infty$ $n h_n^{2b}/\alpha_n \to 0$, and $h_n^{2t-p-q}/\alpha_n \to 0$ with $c > p/2$, $t > q/2$, $2s = b - a$, $b \geq a$ and a, b, t, p, q are as in Assumptions A.1.1 and A.1.2.*

Assumption 4.1 (i) requires the existence of a uniformly bounded partial derivative of the joint density f_{UZ} in the unobservable variable and a sufficient degree of the Sobolev smoothness. The former is standard in the residual-based specification testing literature, while the latter is standard in the semiparametric literature, where a sufficient smoothness of nonparametric components is typically needed to achieve the parametric convergence rate. Given that the strength of the instrument is characterized by the smoothness of the joint density of the regressor and the instrument f_{ZW}, this condition in a sense rules out the extreme endogeneity cases.

Assumption 4.1 (ii) is satisfied whenever the smoothness indices b, t are sufficiently large compared to a, p, q. To illustrate that conditions on tuning parameters are not contradictory, suppose for simplicity that $p = q = c = 1$ and that $h_n \sim n^{-c_1}$ and

$\alpha_n \sim n^{-c_2}$ for some $c_1, c_2 > 0$. Then for b and t large enough, the binding conditions are $c_1 < 1/3$ and $c_2 \in (1/2, 1 - 3c_1)$.

We are interested in the behavior of residual-based empirical distribution function

$$\sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)) \triangleq \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,u]}(\hat{U}_i) - \Pr(U \leq u)\right),$$

where $\hat{U}_i = Y_i - \hat{\varphi}(Z_i), i = 1, \dots, n$ are the nonparametric IV residuals.

The following result shows that the residual empirical process admits a convenient uniform asymptotic expansion.

Theorem 4.1. *Suppose that Assumptions 4.1, A.1.1, and A.1.2, are satisfied. Then uniformly over $u \in \mathbf{R}$*

$$\sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathbf{1}_{(-\infty,u]}(U_i) - F_U(u) + U_i\left[T(T^*T)^{-1}f_{UZ}(u,.)\right](W_i)\right\} + o_P(1).$$

The influence function of the empirical distribution of residuals in the nonparametric IV model differs substantially from the one in the nonparametric conditional mean regression, cf., Akritas and Van Keilegom (2001). One common feature is that the latter depends on the marginal density of regression errors f_U, while the former features the joint density of unobservables and endogenous regressors f_{UZ}. This is probably not surprising given that the two are statistically dependent under the endogeneity. In our case, the influence function also depends on the integral operator T defining an ill-posed nonparametric IV model.

As a consequence of Theorem 4.1, we obtain the following Donsker-type central limit theorem.

Corollary 4.1. *Suppose that assumptions of Theorem 4.1 are satisfied. Then*

$$\sqrt{n}(\hat{F}_{\hat{U}} - F_U) \rightsquigarrow \mathbb{H} \qquad \text{in} \qquad L_\infty(\mathbf{R}),$$

where $\mathbb{H}$ is a tight centered Gaussian process with uniformly continuous sample paths and the covariance function

$$\begin{aligned}(u,u') \mapsto\ & F_U(u \wedge u') - F_U(u)F_U(u') \\ & + \mathbb{E}\left[U^2\left[T(T^*T)^{-1}f_{UZ}(u,.)\right](W)\left[T(T^*T)^{-1}f_{UZ}(u',.)\right](W)\right] \\ & + \mathbb{E}\left[\mathbf{1}_{(-\infty,u]}(U)U\left[T(T^*T)^{-1}f_{UZ}(u',.)\right](W)\right] \\ & + \mathbb{E}\left[\mathbf{1}_{(-\infty,u']}(U)U\left[T(T^*T)^{-1}f_{UZ}(u,.)\right](W)\right].\end{aligned}$$

Although the nonparametric IV residuals are coming from the estimation of an ill-posed inverse problem, the empirical distribution still converges to the true distribution at the square-root n speed. The estimation of residuals, however, affects the asymptotic distribution in a more complicated way than in the case of the nonparametric conditional mean regression.

5 Monte Carlo experiments

To evaluate the finite-sample performance of our separability test, we simulate samples as follows

$$Y = \sin(Z + \theta\varepsilon) + \varepsilon,$$

where

$$\begin{pmatrix} Z \\ W \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 & 0.3 \\ 0.9 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix} \right)$$

The degree of separability is governed by $\theta \in \mathbf{R}$. The separable model corresponds to $\theta = 0$, while any $\theta \neq 0$ corresponds to the alternative nonseparable model. Note that under H_1, the nonparametric IV regression does not estimate consistently the nonseparable structural function $(z, e) \mapsto \sin(z + \theta e)$, which depends on unobservables. The nonparametric IV regression estimates instead the function $z \mapsto \varphi(z)$ solving the functional equation $\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W]$. The difference between the two functions is precisely what gives the power to our test.

We set the number of Monte Carlo replications and the number of bootstrap replications to $1,000$ through all our experiments. We use the sixth-order Epanechnikov kernel to compute $\hat{r}$ and $\hat{T}$ as in Eq. A.1. The bandwidth parameter is computed using Silverman's rule of thumb, while the regularization parameter is computed using the data-driven method of Centorrino (2016). Lastly, we use the adaptive rule of Bickel and Sakov (2008) to estimate the size of the subsample. The rule consists of choosing $\hat{m}_j = \arg\min_{m_j \in M} \sup_x |F_m^*(x) - F_{m+1}^*(x)|$, where $M = \{\lfloor q^j n \rfloor : j = 0, 1, 2, \ldots, 5\}$, F_m^* denotes the empirical distribution of the simulated statistics using a subsample of size m, $\lfloor a \rfloor$ is the greatest integer less or equal to a, and $q = 0.5$.

We look at the distributions of the Kolmogorov-Smirnov statistics, computed as

$$T_{n,\infty} = \sup_{u,w} |\mathbb{G}_n(u, w)|$$

and at the Cramér-von Mises statistics, computed as

$$T_{n,2} = \iint |\mathbb{G}_n(u,w)|^2 \mathrm{d}\hat{F}_{\hat{U}W}(u,w).$$

Figure 1 shows the distribution of the test statistics under the null hypothesis and the two alternative hypotheses for different sample sizes. The two distributions are sufficiently distinct once the alternative hypothesis is sufficiently separated from the null hypothesis.

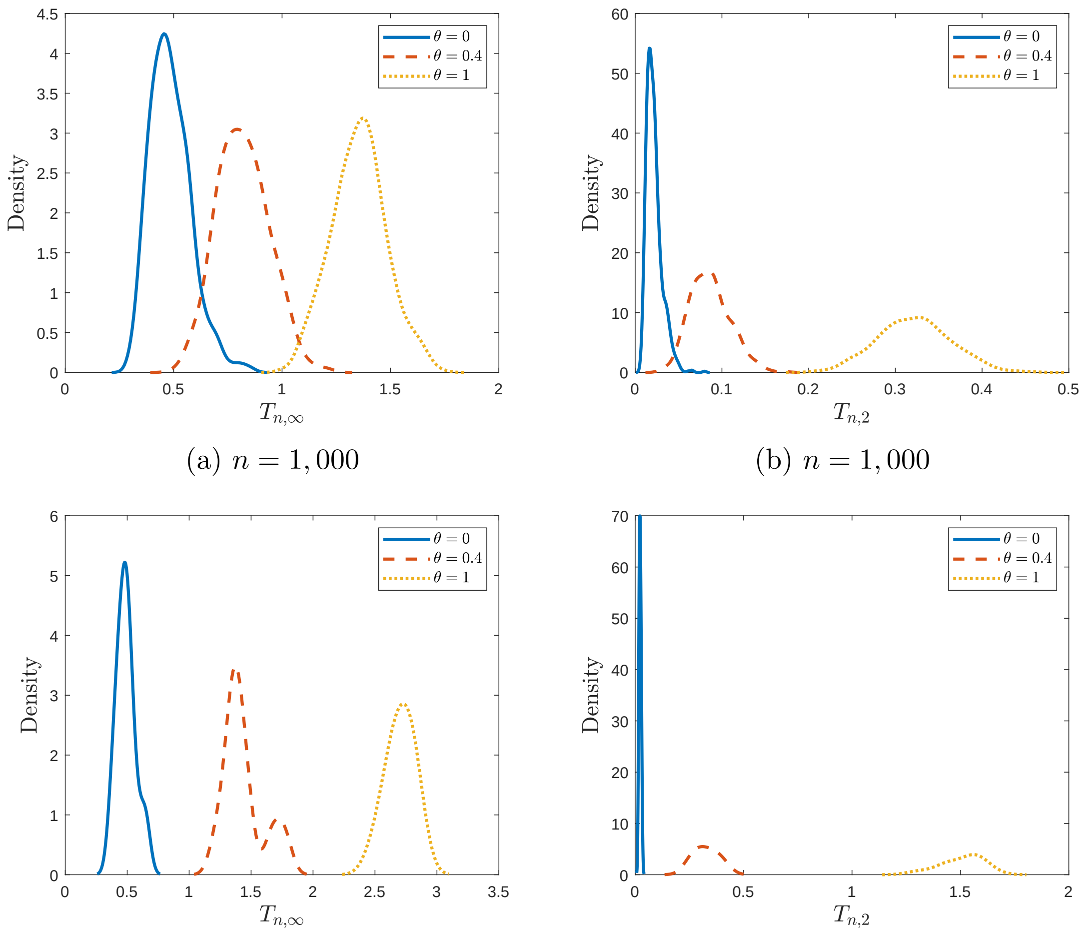
In Figure 2, we plot the power curves when the level of the test is fixed at 5%. The test controls the size, and the power of the test increases once alternative hypotheses become more distant from the null hypothesis. The Cramér-von Mises test seems to have a higher power for the class of considered alternatives. The same figure also indicates that the test is consistent in the sense that its power becomes closer to one as the sample size increases under H_1.

In Figure 3, we explore the performance of the bootstrap. We plot the exact finite sample distribution of both test statistics and the distribution of bootstrapped statistics under H_0. In panels (a) and (b), we plot the distribution of the naive bootstrap, drawing a sample of size n randomly with replacements from $(Y_i, Z_i, W_i)_{i=1}^n$. In panels (c) and (d), we plot the distribution of the m out of n bootstrap.[8] The naive bootstrap mimics neither the distribution of the Kolmogorov-Smirnov statistics nor the distribution of the Cramér-von Mises statistics. The m out of n bootstrap, on the other hand, seems to mimic relatively well the finite sample distributions of both statistics.

6 Testing separability of Engel curves

Engel curves are fundamental for the analysis of consumers' behavior and have implications for the aggregate economic outcomes. An Engel curve describes the relationship between the demand for a particular commodity and the household's budget. Interesting applications of the estimated Engel curves include a measurement of welfare losses associated with tax distortions in Banks, Blundell, and Lewbel (1997), an estimation of the growth and the inflation in Nakamura, Steinsson, and Liu (2016), or an estimation of the income inequality across countries in Almås (2012). Following Blundell, Chen, and Kristensen (2007), modern empirical practice focuses on the separable nonparametric IV approach to the estimation of Engel curves.

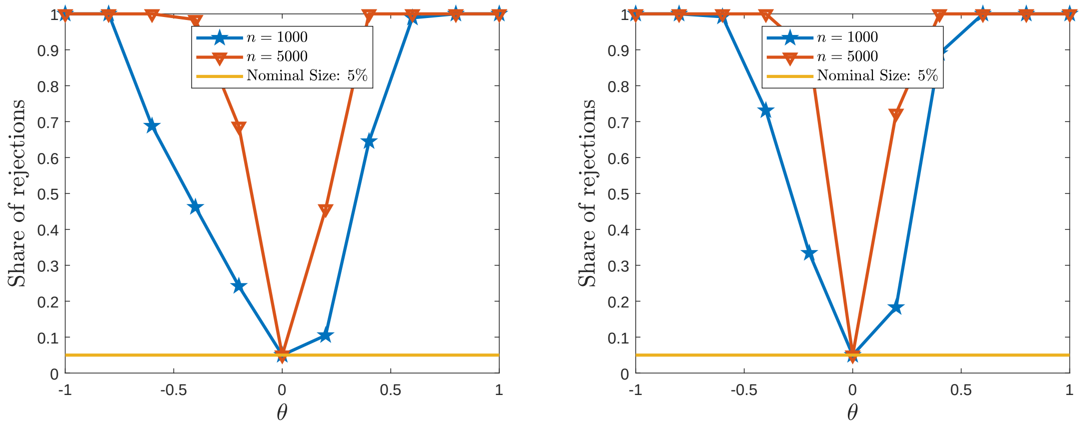
[8] In our experience, the difference between the m out of n bootstrap and the subsampling is negligible for all practical purposes.



(a) $n = 1,000$ (b) $n = 1,000$

(c) $n = 5,000$ (d) $n = 5,000$

Figure 1: Finite-sample distribution under H_0 and two alternatives.



(a) $T_{n,\infty}$

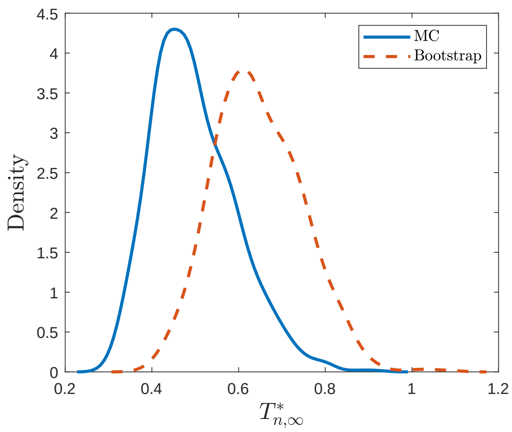
(b) $T_{n,2}$

Figure 2: Power of the test.

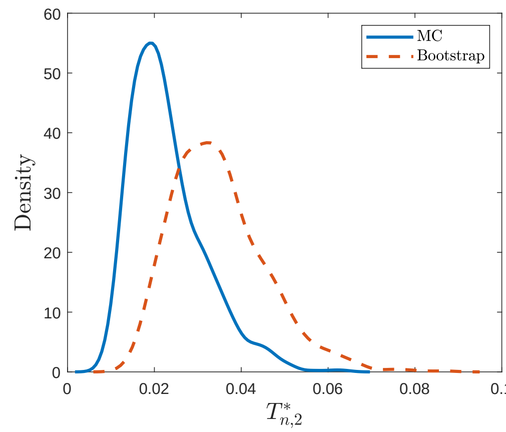
We draw a dataset from the 2015 US Consumer Expenditure Survey; see Babii (2020) for the estimated Engel curves with the uniform confidence bands using this dataset. We restrict our attention to married couples with a positive income during the last 12 months. The dependent variable is a share of expenditures on a particular commodity while the endogenous regressor is a natural logarithm of the total expenditures. Following Blundell, Chen, and Kristensen (2007), we instrument the expenditures using the gross incomes.

In Table 1, we compute the value of the test statistics and the quantiles of order $0.9, 0.95$, and 0.99 of the m out of n bootstrapped distribution, where m is selected adaptively; see Section 5 for more details on the practical implementation of the test. We report results for both the Kolmogorov-Smirnov (KS) and the Cramér-von Mises (CvM) tests. Remarkably, the null hypothesis is rejected in all instances at a 10% significance level and for all commodities, except for the Insurance (KS test) and the Entertainment (CvM test) at a 5% significance level. This suggests that a substantial heterogeneity in unobservables might be present in Engel curves.[9]

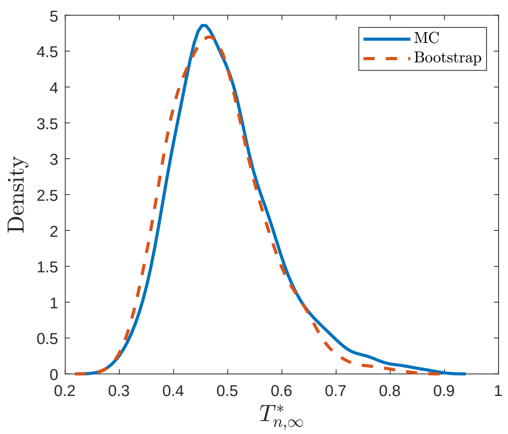
[9]Note that we assume throughout the paper that the nonseparable model is correct in the sense that the instrumental variable is independent of the unobservables. Under the model misspecification our empirical result suggests that the conventional approach to nonseparable modeling should be reconsidered; see, e.g., Masten and Poirier (2016) for the recent advances.



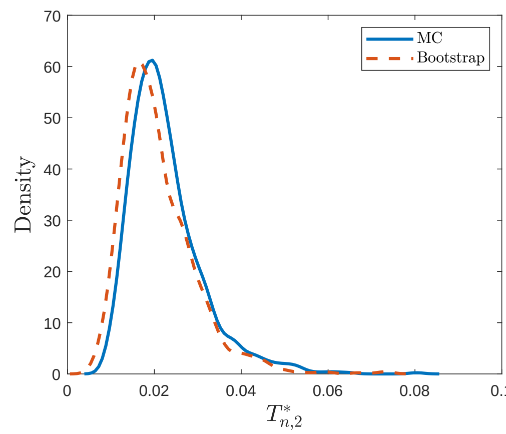
(a) Naive bootstrap: $m_n = n$



(b) Naive bootstrap: $m_n = n$



(c) Adaptive choice of m_n



(d) Adaptive choice of m_n

Figure 3: Distribution of the test statistics and the distribution statistics obtained from m out of n bootstrap under H_0. Sample size: $n = 1,000$.

Table 1: Testing separability of Engel curves

	Kolmogorov-Smirnov				Cramér-von Mises			
	$T_{n,\infty}$	0.9	0.95	0.99	$T_{n,2}$	0.9	0.95	0.99
Food home	1.895	1.166	1.247	1.473	0.719	0.258	0.309	0.430
Food away	0.847	0.644	0.704	0.811	0.086	0.049	0.057	0.078
Clothing	1.010	0.847	0.921	1.086	0.159	0.084	0.102	0.153
Tobacco	1.176	0.722	0.784	0.883	0.266	0.054	0.065	0.087
Alcohol	1.290	0.708	0.760	0.851	0.351	0.060	0.070	0.102
Trips	0.988	0.675	0.748	0.880	0.202	0.053	0.065	0.098
Entertainment	0.808	0.711	0.780	0.946	0.065	0.056	0.070	0.105
Gas and oil	2.731	0.927	0.989	1.110	1.410	0.803	0.849	0.927
Personal care	0.989	0.794	0.881	1.052	0.127	0.083	0.107	0.149
Health	1.899	0.802	0.867	0.979	0.597	0.086	0.101	0.133
Insurance	1.003	0.967	1.047	1.246	0.221	0.129	0.159	0.228
Reading	1.405	0.694	0.761	0.872	0.365	0.059	0.069	0.104
Transportation	1.198	0.770	0.841	0.965	0.264	0.073	0.088	0.126

7 Conclusion

This paper offers a new perspective on the separability of unobservables in economic models with endogeneity. Starting from the nonseparable model where the instrumental variable is independent of unobservables, our first contribution is to develop a novel fully nonparametric test of the separability of unobservables. The test is based on the estimation of a separable nonparametric IV regression and the verification of the independence restriction imposed by a more general nonseparable IV model.

To obtain a large sample approximation to the distribution of our test statistics, we develop novel uniform asymptotic expansions of the empirical distribution function of nonparametric IV residuals. We show that, despite the uncertainty coming from an ill-posed inverse nonparametric IV regression, the empirical distribution function of residuals and the residual-based independence empirical process still satisfy the Donsker central limit theorem. In contrast to the nonparametric regression without endogeneity, we find that the parameter uncertainty affects the asymptotic distribution of the residual-based independence test, which is nonstandard. In our Monte Carlo experiments, we find that the bootstrap fails in approximating the distribution of our test statistics under the null hypothesis; hence we rely on the m out of n bootstrap (or subsampling) procedure to compute its critical values.

Using the 2015 US Consumer Expenditure Survey data, we find that the test rejects the separability of Engel curves for most of the commodities at the commonly accepted significance levels. This indicates that there might be heterogeneity in unobservables and that the nonseparable modeling of Engel curves may be beneficial.

The paper offers several directions for future research. First, it might be interesting to test the separability of unobservables in other structural relations that are commonly estimated using the additively separable models in the empirical practice, such as the production function, the labor supply function, or the wage equation. Second, given the plethora of the residual-based specification tests for regression models without endogeneity, our results could also be used to develop similar tests for econometric models with endogeneity; see, e.g., Dette, Neumeyer, and Keilegom (2007) for a heteroskedasticity test, Pardo-Fernández, Van Keilegom, and González-Manteiga (2007) for a test of the equality of regression curves, and Escanciano, Pardo-Fernández, and Van Keilegom (2018) for more general semiparametric specification tests.

References

Akritas, M. G., and I. Van Keilegom (2001): "Non-parametric estimation of the residual distribution," *Scandinavian Journal of Statistics*, 28(3), 549–567.

Almås, I. (2012): "International income inequality: Measuring PPP bias by estimating Engel curves for food," *American Economic Review*, 102(2), 1093–1117.

Babii, A. (2020): "Honest confidence sets in nonparametric IV regression and other ill-posed models," *Econometric Theory (forthcoming).*

Babii, A., and J.-P. Florens (2017): "Is completeness necessary? Penalized estimation in non-identified models," *arXiv preprint arXiv:1709.03473.*

Babii, A., and R. Kumar (2019): "Isotonic regression discontinuity designs," *arXiv preprint arXiv:1908.05752.*

Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel curves and consumer demand," *Review of Economics and Statistics*, 79(4), 527–539.

Bickel, P. J., and A. Sakov (2008): "On the choice of m in the m out of n bootstrap and confidence bounds for extrema," *Statistica Sinica*, 18(3), 967–985.

Blum, J. R., J. Kiefer, and M. Rosenblatt (1961): "Distribution free tests of independence based on the sample distribution function," *Annals of Mathematical Statistics*, 32(2), 485–498.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75(6), 1613–1669.

Breunig, C. (2020): "Specification testing in nonparametric instrumental quantile regression," *Econometric Theory (forthcoming).*

Brown, B. W., and M. B. Walker (1989): "The random utility hypothesis and inference in demand systems," *Econometrica*, 57(4), 815–829.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of Econometrics*, 6B, 5633–5751.

——— (2014): "Asymptotic normal inference in linear inverse problems," *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics.*

Centorrino, S. (2016): "Data-driven selection of the regularization parameter in additive nonparametric instrumental regressions," *Working paper, Stony Brook University*.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *Econometrics Journal*, 21(1), C1–C68.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014): "Gaussian approximation of suprema of empirical processes," *Annals of Statistics*, 42(4), 1564–1597.

——— (2016): "Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings," *Stochastic Processes and Their Applications*, 126(12), 3632–3651.

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139(1), 4–14.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric instrumental regression," *Econometrica*, 79(5), 1541–1565.

Dette, H., N. Neumeyer, and I. V. Keilegom (2007): "A new test for the parametric form of the variance function in non-parametric regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5), 903–917.

D'Haultfœuille, X., and P. Février (2015): "Identification of nonseparable triangular models with discrete instruments," *Econometrica*, 83(3), 1199–1210.

Dunker, F., J.-P. Florens, T. Hohage, J. Johannes, and E. Mammen (2014): "Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression," *Journal of Econometrics*, 178(3), 444–455.

Durbin, J. (1973): "Weak convergence of the sample distribution function when parameters are estimated," *Annals of Statistics*, 1(2), 279–290.

Einmahl, J. H., and I. Van Keilegom (2008): "Specification tests in nonparametric regression," *Journal of Econometrics*, 143(1), 88–102.

Engl, H. W., M. Hanke, and A. Neubauer (1996): *Regularization of Inverse Problems*, vol. 375. Kluwer Academic Pub.

Escanciano, J. C., J. C. Pardo-Fernández, and I. Van Keilegom (2018): "Asymptotic distribution-free tests for semiparametric regressions with dependent data," *Annals of Statistics*, 46(3), 1167–1196.

Fève, F., J.-P. Florens, and I. Van Keilegom (2018): "Estimation of conditional ranks and tests of exogeneity in nonparametric nonseparable models," *Journal of Business & Economic Statistics*, 36(2), 334–345.

Florens, J.-P. (2003): "Inverse problems and structural econometrics," *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, 2, 46–85.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects," *Econometrica*, 76(5), 1191–1206.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2011): "Identification and estimation by penalization in nonparametric instrumental regression," *Econometric Theory*, 27(03), 472–496.

Gagliardini, P., and O. Scaillet (2012): "Tikhonov regularization for nonparametric instrumental variable estimators," *Journal of Econometrics*, 167(1), 61–75.

——— (2017): "A specification test for nonparametric instrumental variable regression," *Annals of Economics and Statistics/Annales d'Economie et de Statistique*, (128), 151–202.

Giné, E., and R. Nickl (2016): "Mathematical foundations of infinite-dimensional statistical models," *Cambridge Series in Statistical and Probabilistic Mathematics*.

Hall, P., and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics*, 33(6), 2904–2929.

Heckman, J. J. (2001): "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture," *Journal of Political Economy*, 109(4), 673–748.

Hoderlein, S., L. Su, H. White, and T. T. Yang (2016): "Testing for monotonicity in unobservables under unconfoundedness," *Journal of Econometrics*, 193(1), 183–202.

Hoderlein, S., and A. Vanhems (2010): "Welfare analysis using nonseparable models," *Cemmap working paper*.

Horowitz, J. L., and S. Lee (2007): "Nonparametric instrumental variables estimation of a quantile regression model," *Econometrica*, 75(4), 1191–1208.

Imbens, G. W. (2007): "Nonadditive models with endogenous regressors," *Econometric Society Monographs*, 43.

Krein, S. G., and Y. I. Petunin (1966): "Scales of Banach spaces," *Russian Mathematical Surveys*, 21(2), 85–159.

Loynes, R. (1980): "The empirical distribution function of residuals from generalised regression," *Annals of Statistics*, 8(2), 285–298.

Lu, X., and H. White (2014): "Testing for separability in structural equations," *Journal of Econometrics*, 182(1), 14–26.

Mammen, E. (1996): "Empirical process of residuals for high-dimensional linear models," *Annals of Statistics*, 24(1), 307–335.

Masten, M., and A. Poirier (2016): "Partial independence in nonseparable models," *Duke Working Paper*.

Nakamura, E., J. Steinsson, and M. Liu (2016): "Are Chinese growth and inflation too smooth? Evidence from Engel curves," *American Economic Journal: Macroeconomics*, 8(3), 113–44.

Neumeyer, N., and I. Van Keilegom (2019): "Bootstrap of residual processes in regression: To smooth or not to smooth?," *Biometrika*, 106(2), 385–400.

Newey, W. K., and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71(5), 1565–1578.

Nickl, R., and B. M. Pötscher (2007): "Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov-and Sobolev-type," *Journal of Theoretical Probability*, 20(2), 177–199.

Pardo-Fernández, J. C., I. Van Keilegom, and W. González-Manteiga (2007): "Testing for the equality of k regression curves," *Statistica Sinica*, 17(3), 1115–1137.

Politis, D. N., J. P. Romano, and M. Wolf (2001): "On the asymptotic theory of subsampling," *Statistica Sinica*, 11(4), 1105–1124.

Simar, L., A. Vanhems, and I. Van Keilegom (2016): "Unobserved heterogeneity and endogeneity in nonparametric frontier estimation," *Journal of Econometrics*, 190(2), 360–373.

Su, L., Y. Tu, and A. Ullah (2015): "Testing additive separability of error term in nonparametric structural models," *Econometric Reviews*, 34(6-10), 1057–1088.

Tikhonov, A. N. (1963): "On the solution of ill-posed problems and the method of regularization," *Doklady Akademii Nauk SSSR*, 151(3), 501–504.

Torgovitsky, A. (2015): "Identification of nonseparable models using instruments with small support," *Econometrica*, 83(3), 1185–1197.

Van Der Vaart, A. W., and J. A. Wellner (2000): *Weak convergence*. Springer.

APPENDIX

A.1 Tikhonov regularization in Sobolev scales

The asymptotic theory developed in this paper relies heavily on the empirical process theory. Our first insight is that the regularization in Sobolev spaces provides a natural link between the empirical process theory and the theory of the regularization of ill-posed inverse problems. Consequently, we focus on the Tikhonov-regularized estimator penalized by the Sobolev-norm. The idea of the penalization by derivatives in the ill-posed inverse literature dates back to the original work of Tikhonov (1963); see Florens, Johannes, and Van Bellegem (2011), Gagliardini and Scaillet (2012), and Carrasco, Florens, and Renault (2014) for the recent developments in econometrics. Another attractive feature of the regularization in Sobolev spaces is that it allows us to overcome the saturation effect of the classical Tikhonov regularization, cf., Darolles, Fan, Florens, and Renault (2011).

Let $(L_2(\mathbf{R}^p), \|.\|)$ denote the space of functions square-integrable with respect to the Lebesgue measure. Let $\langle x \rangle^s \triangleq (1+|x|^2)^{s/2}$ be a polynomial weight function with $s \in \mathbf{R}$, where $|.|$ is a Euclidean norm on $\mathbf{R}^p$. Consider the operator $L^s f = F^{-1}(\langle . \rangle^s F f)$ defined for all f such that $\|\langle . \rangle^2 F f\| < \infty$, where F is a Fourier transform on $L_2(\mathbf{R}^p)$ with scaling $(2\pi)^{-p/2}$. The self-adjoint operator L generates a Hilbert scale of Sobolev spaces

$$H^s(\mathbf{R}^p) = \left\{ f \in L_2(\mathbf{R}^p) : \ \|f\|_s \triangleq \|L^s f\| < \infty \right\};$$

see Krein and Petunin (1966) for more details on Banach and Hilbert scales.

Consider an ill-posed inverse problem $T\varphi = r$, where $\varphi \in L_2(\mathbf{R}^p)$ is a structural function of interest and $T : L_2(\mathbf{R}^p) \to L_2(\mathbf{R}^q)$ is an operator. We impose the following assumption on T and φ.

Assumption A.1.1. *For some $a, b > 0$*

(i) Operator smoothing: $\|T\phi\| \sim \|\phi\|_{-a}$ for all $\phi \in L_2(\mathbf{R}^p)$.

(ii) Parameter smoothness: $\varphi \in H^b(\mathbf{R}^p)$.

Assumption A.1.1 (i) restricts smoothing properties of the operator T. Roughly speaking, the action of the operator T increases the Sobolev smoothness by a. The number a is called the degree of ill-posedness, since the more the operator T smooths out features of the function, the harder it is to recover it. Assumption A.1.1 (ii) restricts the smoothness of the structural function φ and is standard in the theory of nonparametric estimation.

Our first result is a risk bound in the Sobolev norm for the Tikhonov-regularized estimator in Eq. 1, where $\hat{T}$ and $\hat{r}$ are some estimators of T and r such that $\hat{T}^*\hat{T}$ is a bounded linear self-adjoint operator. The problem in Eq. 1 admits the closed-form solution

$$\hat{\varphi} = L^{-s}(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\hat{r},$$

where $\hat{T}_s = \hat{T}L^{-s}$ and $\hat{T}_s^*$ is the adjoint operator to $\hat{T}_s$. The following result can be understood as the extension of Carrasco, Florens, and Renault (2014), Proposition 3.1 to the case of unknown operator.

Theorem A.1. *Suppose that Assumption A.1.1 is satisfied with $s \geq (b-a)/2$ and that $\|\hat{T} - T\|^2 \lesssim_P \alpha_n$. Then for every $c \in [0, s]$*

$$\|\hat{\varphi} - \varphi\|_c^2 \lesssim_P \alpha_n^{-\frac{a+c}{a+s}} \left\|\hat{r} - \hat{T}\varphi\right\|^2 + \alpha_n^{\frac{b-c}{a+s}}.$$

Proof. Decompose

$$\hat{\varphi} - \varphi = I_n + II_n + III_n + IV_n + V_n,$$

with

$$\begin{aligned}
I_n &= L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*(\hat{r} - \hat{T}\varphi) \\
II_n &= L^{-s}(\alpha_n I + T_s^*T_s)^{-1}(\hat{T}_s^* - T_s^*)(\hat{r} - \hat{T}\varphi) \\
III_n &= L^{-s}\left[(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1} - (\alpha_n I + T_s^*T_s)^{-1}\right]\hat{T}_s^*(\hat{r} - \hat{T}\varphi) \\
IV_n &= L^{-s}(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\hat{T}_s L^s\varphi - L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s L^s\varphi \\
V_n &= L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s L^s\varphi - \varphi.
\end{aligned}$$

For the first term, by Engl, Hanke, and Neubauer (1996), Corollary 8.22

$$\begin{aligned}
\|I_n\|_c^2 &= \left\|(\alpha_n I + T_s^*T_s)^{-1}T_s^*(\hat{r} - \hat{T}\varphi)\right\|_{c-s}^2 \\
&\lesssim \left\|(T_s^*T_s)^{\frac{s-c}{2(a+s)}}(\alpha_n I + T_s^*T_s)^{-1}T_s^*\right\|^2 \left\|(\hat{r} - \hat{T}\varphi)\right\|^2 \\
&\lesssim \sup_\lambda \left|\frac{\lambda^{\frac{2s+a-c}{2(a+s)}}}{\alpha_n + \lambda}\right|^2 \\
&\lesssim \alpha_n^{-\frac{a+c}{a+s}} \left\|\hat{r} - \hat{T}\varphi\right\|^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}\|II_n\|_c^2 &= \left\|(\alpha_n I + T_s^*T_s)^{-1}(\hat{T}_s^* - T_s^*)(\hat{r} - \hat{T}\varphi)\right\|_{c-s}^2 \\ &\lesssim \left\|(T_s^*T_s)^{\frac{s-c}{2(a+s)}}(\alpha_n I + T_s^*T_s)^{-1}\right\|^2 \|\hat{T}^* - T^*\|^2 \left\|\hat{r} - \hat{T}\varphi\right\|^2 \\ &\lesssim_P \alpha_n^{-\frac{2a+s+c}{a+s}} \alpha_n \left\|\hat{r} - \hat{T}\varphi\right\|^2 \\ &\lesssim \alpha_n^{-\frac{a+c}{a+s}} \left\|\hat{r} - \hat{T}\varphi\right\|^2.\end{aligned}$$

Since $L^s\varphi \in H^{b-s}$ and $s \geq (b-a)/2$, by Engl, Hanke, and Neubauer (1996), Corollary 8.22, there exists a function $\psi \in L_2(\mathbf{R}^p)$ such that $L^s\varphi = (T_s^*T_s)^{\frac{b-s}{2(a+s)}}\psi$. Then

$$\begin{aligned}\|V_n\|_c^2 &= \left\|(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_sL^s\varphi - L^s\varphi\right\|_{c-s} \\ &= \left\|\alpha_n(\alpha_n I + T_s^*T_s)^{-1}L^s\varphi\right\|_{c-s}^2 \\ &\lesssim \left\|\alpha_n(T_s^*T_s)^{\frac{s-c}{2(a+s)}}(\alpha_n I + T_s^*T_s)^{-1}(T_s^*T_s)^{\frac{b-s}{2(a+s)}}\psi\right\|^2 \\ &\leq \sup_\lambda \left|\frac{\alpha_n \lambda^{\frac{b-c}{2(a+s)}}}{\alpha_n + \lambda}\right|^2 \|\psi\|^2 \\ &= O\left(\alpha_n^{\frac{b-c}{a+s}}\right).\end{aligned}$$

Next, decompose

$$\begin{aligned}\|III_n\|_c^2 &= \left\|\left[(\alpha_n I + T_s^*T_s)^{-1} - (\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\right]\hat{T}_s^*(\hat{r} - \hat{T}\varphi)\right\|_{c-s}^2 \\ &= \left\|(\alpha_n I + T_s^*T_s)^{-1}(\hat{T}_s^*\hat{T}_s - T_s^*T_s)(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*(\hat{r} - \hat{T}\varphi)\right\|_{c-s}^2 \\ &\lesssim R_{1n} + R_{2n}\end{aligned}$$

with

$$\begin{aligned}R_{1n} &= \left\|(T_s^*T_s)^{\frac{s-c}{2(a+s)}}(\alpha_n I + T_s^*T_s)^{-1}T_s^*(\hat{T}_s - T_s)(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*(\hat{r} - \hat{T}\varphi)\right\|^2 \\ &\lesssim \left\|(T_s^*T_s)^{\frac{s-c}{2(a+s)}}(\alpha_n I + T_s^*T_s)^{-1}T_s^*\right\|^2 \left\|\hat{T}_s - T_s\right\|^2 \alpha_n^{-1} \left\|\hat{r} - \hat{T}\varphi\right\|^2 \\ &\lesssim_P \alpha_n^{-\frac{a+c}{a+s}} \left\|\hat{r} - \hat{T}\varphi\right\|^2\end{aligned}$$

and

$$
\begin{aligned}
R_{2n} &= \left\| (T_s^* T_s)^{\frac{s-c}{2(a+s)}} (\alpha_n I + T_s^* T_s)^{-1} (\hat{T}_s^* - T_s^*) \hat{T}_s (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{r} - \hat{T}\varphi) \right\|^2 \\
&\leq \left\| (T_s^* T_s)^{\frac{s-c}{2(a+s)}} (\alpha_n I + T_s^* T_s)^{-1} \right\|^2 \left\| \hat{T}_s^* - T_s^* \right\|^2 \left\| \hat{r} - \hat{T}\varphi \right\|^2 \\
&\lesssim_P \alpha_n^{-\frac{2a+c+s}{a+s}} \alpha_n \left\| \hat{r} - \hat{T}\varphi \right\|^2 \\
&\lesssim_P \alpha_n^{-\frac{a+c}{a+s}} \left\| \hat{r} - \hat{T}\varphi \right\|^2 .
\end{aligned}
$$

Lastly, decompose

$$
\begin{aligned}
\|IV_n\|_c^2 &= \left\| \alpha_n \left[(\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} - (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \right] L^s \varphi \right\|_{c-s}^2 \\
&\lesssim \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \left(\hat{T}_s^* \hat{T}_s - T_s^* T_s \right) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|_{c-s}^2 \\
&\leq 2S_{1n} + 2S_{2n}
\end{aligned}
$$

with

$$
\begin{aligned}
S_{1n} &= \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* \left(\hat{T}_s - T_s \right) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|_{c-s}^2 \\
&\lesssim \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* \right\|^2 \left\| \hat{T} - T \right\|^2 \left\| \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|_{-s}^2 \\
&\lesssim \left\| \alpha_n (T_s^* T_s)^{\frac{s}{2(a+s)}} (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{\frac{b-s}{2(a+s)}} \psi \right\|^2 \\
&\lesssim \sup_\lambda \left| \frac{\alpha_n \lambda^{\frac{b}{2(a+s)}}}{\alpha_n + \lambda} \right|^2 \\
&\lesssim \alpha_n^{\frac{b}{a+s}}
\end{aligned}
$$

and

$$
\begin{aligned}
S_{2n} &= \left\| (\alpha_n I + \hat{T}^* \hat{T})^{-1} \left(\hat{T}^* - T^* \right) \alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|_{c-s}^2 \\
&\leq \left\| (T_s^* T_s)^{-\frac{c-s}{2(a+s)}} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \right\|^2 \left\| \hat{T}^* - T^* \right\|^2 \left\| \alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{\frac{b-s}{2(a+s)}} \psi \right\|^2 \\
&\lesssim_P \alpha_n^{-\frac{2a+s+c}{a+s}} \left\| \hat{T} - T \right\|^2 \alpha_n^{\frac{b+a}{a+s}} \\
&\lesssim_P \alpha_n^{\frac{b-c}{a+s}} .
\end{aligned}
$$

The result follows from combining all estimates together. $\square$

It is worth emphasizing that our risk bound is not specific to the nonparametric IV regression. This result is of independent interest since it can be specialized to obtain convergence rates of derivatives of arbitrary nonparametric estimators of $\hat{T}$ and $\hat{r}$ that have known L_2 convergence rates, including the machine learning estimators that become increasingly popular.

Next, we specialize the generic risk bound in Theorem A.1 to the nonparametric IV regression with r and T estimated as

$$\begin{aligned} \hat{r}(w) &= \frac{1}{nh_n^q}\sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right) \\ (\hat{T}\phi)(w) &= \int \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z, \quad \phi \in L_2(\mathbf{R}^p) \\ \hat{f}_{ZW}(z,w) &= \frac{1}{nh_n^{p+q}}\sum_{i=1}^n K_z\left(h_n^{-1}(Z_i - z)\right)K_w\left(h_n^{-1}(W_i - w)\right), \end{aligned} \tag{A.1}$$

where $K_z : \mathbf{R}^p \to \mathbf{R}$ and $K_w : \mathbf{R}^q \to \mathbf{R}$ are kernel functions and $h_n \to 0$ is a sequence of bandwidth parameters. We introduce several additional restrictions on the distribution of the data and the kernel functions.

Assumption A.1.2. *(i) $(Y_i, Z_i, W_i)_{i=1}^n$ are i.i.d. observations of $(Y, Z, W) \in \mathbf{R} \times \mathbf{R}^p \times \mathbf{R}^q$ admitting Lebesgue density such that $\mathbb{E}\|W\| < \infty$, $\mathbb{E}\|Z\| < \infty$, and $\mathbb{E}[U^2|W] \leq C < \infty$; (ii) $f_Z, f_W, f_{ZW}, f_{U|Z} \in L_\infty$ and $f_Z, f_{ZW} \in L_2$; (iii) $f_{ZW} \in H^t(\mathbf{R}^{p+q})$ for some $t > 0$; (iv) K_z and K_w products of a univariate right (or left) continuous kernel $K \in L_2(\mathbf{R}) \cap L_\infty(\mathbf{R})$ of bounded variation with $\int K(u)\mathrm{d}u = 1$, $\int |u|^l |K(u)|\mathrm{d}u < \infty$, and $\int u^k K(u)\mathrm{d}u = 0$ for $k \in \{1, \dots, l\}$ and $l \geq t$.*

Assumption A.1.2 (i)-(ii) features some mild restrictions on the distribution of the data. Assumption A.1.2 (iii)-(iv) involves a smoothness condition and conditions on the kernel functions that are standard in the non-parametric estimation based on the kernel smoothing.

The following risk bound will be used in the subsequent section to characterize the asymptotic distribution of our separability test and to control remainders in the uniform asymptotic expansions of the residual empirical process.

Corollary A.1.1. *Suppose that Assumptions A.1.1 and A.1.2 are satisfied, $\frac{1}{nh_n^{p+q}} \vee h_n^{2t} = O(\alpha_n)$, and $2s = b - a$. Then for every $c \in [0, s]$*

$$\|\hat{\varphi} - \varphi\|_c^2 = O_P\left(\alpha_n^{-\frac{2(a+c)}{b+a}}\left(\frac{1}{nh_n^q} + h_n^{2t}\right) + \alpha_n^{\frac{2(b-c)}{b+a}}\right).$$

Proof. By the Cauchy-Schwartz inequality

$$\begin{aligned}\left\|\hat{T} - T\right\|^2 &\leq \left\|\hat{f}_{ZW} - f_{ZW}\right\|^2 \\ &= O\left(\frac{1}{nh_n^{p+q}} + h_n^{2t}\right),\end{aligned}$$

where the second line follows from the well-known risk bound; see, e.g., Giné and Nickl (2016) under Assumption A.1.2 (i), (iii)-(iv). Therefore, by Theorem A.1

$$\|\hat{\varphi} - \varphi\|_c^2 \lesssim_P \alpha_n^{-\frac{a+c}{a+s}} \left\|\hat{r} - \hat{T}\varphi\right\|^2 + \alpha_n^{\frac{b-c}{a+s}}.$$

The proof of

$$\mathbb{E}\left\|\hat{r} - \hat{T}\varphi\right\|^2 = O\left(\frac{1}{nh_n^q} + h_n^{2t}\right).$$

can be found in Babii and Florens (2017). □

A.2 Proofs of main results

We prove first a supplementary lemma, which is used to establish the uniform asymptotic expansions in Theorem 3.1 and Theorem 4.1.

Lemma A.2.1. *Suppose that Assumptions 4.1, A.1.1, and A.1.2 are satisfied. Then*

$$\sup_u \left|\hat{F}_{\hat{U}}(u) - \hat{F}_U(u) - \Pr(U \leq u + \hat{\Delta}(Z)|\mathscr{X}) + F_U(u)\right| = o_P\left(n^{-1/2}\right), \qquad \text{(A.2)}$$

where $\hat{\Delta} = \hat{\varphi} - \varphi$ *and* $\mathscr{X} = (Y_i, Z_i, W_i)_{i=1}^{\infty}$.

Proof. The main idea of the proof is to embed the process inside the supremum into an empirical process indexed by u and a Sobolev ball containing $\hat{\Delta}$ with a probability tending to one. The latter property follows from the consistency of the nonparametric IV estimator in the Sobolev norm; see Corollary A.1.1. We first show that the process is Donsker, whence the supremum in Eq. A.2 is $O_P(n^{-1/2})$. Finally, the required $o_P(n^{-1/2})$ order will follow from the fact that the process is degenerate.

Let H_M^c be a ball of radius $M < \infty$ in the Sobolev space $H^c(\mathbf{R}^p)$. For $u \in \mathbf{R}$ and $\Delta \in H_M^c$, define

$$f_{u,\Delta}(U, Z) = \mathbf{1}_{(-\infty, u+\Delta(Z)]}(U)$$

and

$$\begin{aligned}\mathcal{G}_1 &= \{f_{u,\Delta}:\ u \in \mathbf{R}, \Delta \in H_M^c(\mathbf{R}^p)\}\\ \mathcal{G}_2 &= \{f_{u,0}:\ u \in \mathbf{R}\}\\ \mathcal{G} &= \mathcal{G}_1 - \mathcal{G}_2.\end{aligned}$$

Note that $\mathcal{G}_2$ is a classical Donsker class of indicator functions. If we can show that $\mathcal{G}_1$ is Donsker, then $\mathcal{G}$ will be Donsker as a sum of two Donsker classes; see Van Der Vaart and Wellner (2000), Theorem 2.10.6. To this end, we check that the bracketing entropy condition is satisfied for $\mathcal{G}_1$. By Nickl and Pötscher (2007), Corollary 4, under Assumption A.1.2 (i), the bracketing number of H_M^c satisfies

$$\log N_{[\,]}(\varepsilon, H_M^c, \|.\|_{L_Z^2}) \lesssim \begin{cases}\varepsilon^{-p/c} & c < (p+1)/2\\ \varepsilon^{-2p/(p+1)} & c \geq (p+1)/2\end{cases},$$

where $(L_Z^2, \|.\|_{L_Z^2})$ denotes the space of functions, square-integrable with respect to f_Z. Put $M_\varepsilon = N_{[\,]}(\varepsilon, H_M^c, \|.\|_{L_Z^2})$ and fix $u \in \mathbf{R}$. Let $\left[\underline{\Delta}_j, \overline{\Delta}_j\right]_{j=1}^{M_\varepsilon}$ be a collection of ε-brackets for H_M^c, i.e., for any $\Delta \in H_M^c$, there exists $1 \leq j \leq M_\varepsilon$ such that $\underline{\Delta}_j \leq \Delta \leq \overline{\Delta}_j$ and $\left\|\overline{\Delta}_j - \underline{\Delta}_j\right\|_{L_Z^2} \leq \varepsilon$, and whence

$$\mathbf{1}_{\left(-\infty, u+\underline{\Delta}_j\right]} \leq \mathbf{1}_{(-\infty, u+\Delta]} \leq \mathbf{1}_{\left(-\infty, u+\overline{\Delta}_j\right]}.$$

Now for each $1 \leq j \leq M_\varepsilon$, partition the real line into intervals defined by grids of points $-\infty = \underline{u}_{j,1} < \underline{u}_{j,2} < \cdots < \underline{u}_{j,M_{1\varepsilon}} = \infty$ and $-\infty = \overline{u}_{j,1} < \overline{u}_{j,2} < \cdots < \overline{u}_{j,M_{2\varepsilon}} = \infty$, so that each segment has probabilities

$$\begin{aligned}&\Pr\left(U - \underline{\Delta}_j(Z) \leq \underline{u}_{j,k-1}\right) - \Pr\left(U - \underline{\Delta}_j(Z) \leq \underline{u}_{j,k}\right) \leq \varepsilon^2/2, \qquad 2 \leq k \leq \frac{2}{\varepsilon^2} \triangleq M_{1\varepsilon},\\ &\Pr\left(U - \overline{\Delta}_j(Z) \leq \overline{u}_{j,k-1}\right) - \Pr\left(U - \overline{\Delta}_j(Z) \leq \overline{u}_{j,k}\right) \leq \varepsilon^2/2, \qquad 2 \leq k \leq \frac{2}{\varepsilon^2} \triangleq M_{2\varepsilon}.\end{aligned}$$

Denote the largest $\underline{u}_{j,k}$ such that $\underline{u}_{j,k} \leq u$ by $\underline{u}_j^*$ and the smallest $\overline{u}_{j,k}$ such that $u \leq \overline{u}_{jk}$ by $\overline{u}_j^*$. Consider the following family of brackets

$$\left[\mathbf{1}_{\left(-\infty, \underline{u}_j^* + \underline{\Delta}_j\right]}, \mathbf{1}_{\left(-\infty, \overline{u}_j^* + \overline{\Delta}_j\right]}\right]_{j=1}^{M_\varepsilon}.$$

Under Assumption A.1.2 (ii)

$$\begin{aligned}\left\|\mathbf{1}_{(-\infty,\overline{u}_j^*+\overline{\Delta}_j]}-\mathbf{1}_{(-\infty,\underline{u}_j^*+\underline{\Delta}_j]}\right\|_{L_Z^2}^2 &= \Pr\left(\underline{u}_j^*+\underline{\Delta}_j(Z)\leq U\leq \overline{u}_j^*+\overline{\Delta}_j(Z)\right)\\ &\leq \Pr\left(u+\underline{\Delta}_j(Z)\leq U\leq u+\overline{\Delta}_j(Z)\right)+\varepsilon^2\\ &= \int\left\{\int_{u+\underline{\Delta}_j(z)}^{u+\overline{\Delta}_j(z)} f_{U|Z}(u|z)\mathrm{d}u\right\} f_Z(z)\mathrm{d}z+\varepsilon^2\\ &\leq \left\|\overline{\Delta}_j-\underline{\Delta}_j\right\|_{L_Z^2}\|f_{U|Z}\|_\infty+\varepsilon^2=O\left(\varepsilon^2\right).\end{aligned}$$

Therefore, we constructed brackets of size $O(\varepsilon)$, covering $\mathcal{G}_1$, and we have used at most $O\left(\varepsilon^{-2}M_\varepsilon\right)$ such brackets. Since $c > p/2$, the bracketing integral converges[10]

$$\int_0^1\sqrt{\log N_{[\,]}(\varepsilon,\mathcal{G},\|.\|_{L_Z^2})}\mathrm{d}\varepsilon<\infty.$$

Therefore, the empirical process

$$\sqrt{n}(P_n-P)g,\qquad g\in\mathcal{G}$$

is asymptotically equicontinuous; see Van Der Vaart and Wellner (2000), Theorem 1.5.7, i.e., for any $\varepsilon > 0$

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\Pr{}^*\left(\sup_{f,g\in\mathcal{G}:\,\rho(f-g)<\delta}|\sqrt{n}(P_n-P)(f-g)|>\varepsilon\right)=0,\tag{A.3}$$

where $\Pr^*$ denotes the outer probability measure. Now, we show that for every $u\in\mathbf{R}$

$$\rho^2(\hat{f}_u)=\mathbb{E}[\hat{f}_u^2]-(\mathbb{E}[\hat{f}_u])^2=o_P(1)$$

with $\hat{f}_u=\mathbf{1}_{(-\infty,u+\hat{\Delta}(Z))]}(U)-\mathbf{1}_{(-\infty,u]}(U)$, where the expectation is computed with respect to (U,Z) only. Indeed,

$$\begin{aligned}\mathbb{E}[\hat{f}_u] &= \Pr(u\leq U\leq u+\hat{\Delta}(Z)|\mathscr{X})\\ &= \int\int_u^{u+\hat{\Delta}(z)} f_{U|Z}(v|z)\mathrm{d}v f_Z(z)\mathrm{d}z\\ &\leq \|f_{U|Z}\|_\infty\|f_Z\|\|\hat{\Delta}\|\\ &= o_P(1),\end{aligned}$$

[10]Note that for $\varepsilon > 1$, $N_{[\,]}(\varepsilon,\mathcal{G},\|.\|_{L_Z^2})=1$, since a single bracket $[0,1]$ contains all $g\in\mathcal{G}$.

where the third line follows by the Cauchy-Schwartz inequality and Corollary A.1.1 under Assumptions 4.1, A.1.1, and A.1.2. Similarly,

$$\begin{aligned}\mathbb{E}[\hat{f}_u^2] &= \Pr(U \leq u + \hat{\Delta}(Z)|\mathscr{X}) + \Pr(U \leq u) - 2\Pr(U \leq (u + \hat{\Delta}(Z)) \wedge u|\mathscr{X}) \\ &\leq \int \int_u^{u+\hat{\Delta}(z)} f_{U|Z}(v|z)\mathrm{d}v f_Z(z)\mathrm{d}z \\ &\lesssim \|\hat{\Delta}\| \\ &= o_P(1).\end{aligned}$$

Lastly, let $\|\hat{\nu}_n\|_\infty$ denote the supremum in Eq A.2. Then

$$\Pr^*(\sqrt{n}\|\hat{\nu}_n\|_\infty > \varepsilon) \leq \Pr^*\left(\sqrt{n}\|\hat{\nu}_n\|_\infty > \varepsilon, \rho(\hat{f}_u) < \delta, \hat{\Delta} \in H_M^c\right) + \Pr^*\left(\rho(\hat{f}_u) \geq \delta\right) + \Pr^*\left(\hat{\Delta} \notin H_M^c\right),$$

where the second probability tends to zero as we have just shown and the last probability tends to zero since under the maintained assumptions, by Corollary A.1.1, $\|\hat{\varphi} - \varphi\|_c = o_P(1)$. Therefore, it follows from the asymptotic equicontinuity in Eq. A.3 that

$$\limsup_{n\to\infty} \Pr^*(\sqrt{n}\|\hat{\nu}_n\|_\infty > \varepsilon) = 0,$$

which concludes the proof. □

Proof of Theorem 4.1. By Lemma A.2.1, the following expansion holds uniformly in $u \in \mathbf{R}$

$$\sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)) = \sqrt{n}(\hat{F}_U(u) - F_U(u)) + \sqrt{n}\left(\Pr\left(U \leq u + \hat{\Delta}(Z)|\mathscr{X}\right) - F_U(u)\right) + o_P(1).$$

By Taylor's theorem, there exists some $\tau \in [0, 1]$ such that

$$\begin{aligned}&\sqrt{n}\left(\Pr\left(U \leq u + \hat{\Delta}(Z)|\mathscr{X}\right) - \Pr(U \leq u)\right) \\ &= \sqrt{n}\int\left\{\int_{-\infty}^{u+\hat{\Delta}(z)} f_{UZ}(v, z)\mathrm{d}v - \int_{-\infty}^{u} f_{UZ}(v, z)\mathrm{d}v\right\}\mathrm{d}z \\ &= \sqrt{n}\int\left\{f_{UZ}(u, z)\hat{\Delta}(z) + \frac{1}{2}\partial_u f_{UZ}(u + \tau\hat{\Delta}(z), z)\hat{\Delta}^2(z)\right\}\mathrm{d}z \\ &= \sqrt{n}\langle\hat{\varphi} - \varphi, f_{UZ}(u, .)\rangle + \sqrt{n}\frac{1}{2}\int \partial_u f_{UZ}(u + \tau\hat{\Delta}(z), z)\hat{\Delta}^2(z)\mathrm{d}z \\ &\triangleq T_{1n}(u) + T_{2n}(u).\end{aligned}$$

Under Assumption 4.1

$$\begin{aligned}\|T_{2n}\|_\infty &\le \|\partial_u f_{UZ}\|_\infty \sqrt{n}\|\hat{\varphi}-\varphi\|^2 \\ &= o_P(1).\end{aligned}$$

Next, similarly to the proof of Theorem A.1, decompose

$$\begin{aligned}T_{1n}(u) &= \sqrt{n}\langle\hat{\varphi}-\varphi, f_{UZ}(u,.)\rangle \\ &\triangleq I_n(u) + II_n(u) + III_n(u)\end{aligned}$$

with

$$\begin{aligned}I_n(u) &= \sqrt{n}\left\langle L^{-s}(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*(\hat{r}-\hat{T}\varphi), f_{UZ}(u,.)\right\rangle \\ II_n(u) &= \sqrt{n}\left\langle L^{-s}(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\hat{T}_s L^s\varphi - L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s L^s\varphi, f_{UZ}(u,.)\right\rangle \\ III_n(u) &= \sqrt{n}\left\langle L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s L^s\varphi - \varphi, f_{UZ}(u,.)\right\rangle.\end{aligned}$$

We show below that II_n and III_n are negligible. First,

$$\begin{aligned}\|III_n\|_\infty &= \sqrt{n}\sup_u\left|\left\langle L^{-(a+s)}\left[(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s - I\right]L^s\varphi, L^a f_{UZ}(u,.)\right\rangle\right| \\ &\lesssim \sqrt{n}\left\|\left[(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_s - I\right]L^s\varphi\right\|_{-(a+s)} \\ &\lesssim \sqrt{n}\left\|(T_s^*T_s)^{1/2}\alpha_n(\alpha_n I + T_s^*T_s)^{-1}(T_s^*T_s)^{\frac{b-s}{2(a+s)}}\psi\right\| \\ &\lesssim \sqrt{n}\sup_\lambda\left|\frac{\alpha_n\lambda^{\frac{b-s}{2(a+s)}+\frac{1}{2}}}{\alpha_n+\lambda}\right|\|\psi\| \\ &= \sqrt{n}\sup_\lambda\left|\frac{\alpha_n\lambda}{\alpha_n+\lambda}\right|\|\psi\| \\ &= O\left(\sqrt{n}\alpha_n\right) \\ &= o(1),\end{aligned}$$

where the first line follows since L is self-adjoint and $L^a f_{UZ}(u,.)$ exists, the second by the Cauchy-Schwartz inequality and Assumption 4.1 (i), the third by Engl, Hanke, and Neubauer (1996), Corollary 8.22. for some $\psi \in L_2(\mathbf{R}^p)$, the fourth by the isometry of the functional calculus, the fifth by $2s = b - a$, and the last since $n\alpha_n^2 \to 0$ under Assumption 4.1 (ii).

Next,

$$\begin{aligned}
\|II_n\|_\infty &= \sqrt{n}\sup_u \left|\left\langle L^{-s}\alpha_n\left[(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}-(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\right]L^s\varphi, f_{UZ}(u,.)\right\rangle\right| \\
&= \sqrt{n}\sup_u \left|\left\langle L^{-s}(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\left(\hat{T}_s^*\hat{T}_s-T_s^*T_s\right)\alpha_n(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi, f_{UZ}(u,.)\right\rangle\right| \\
&\lesssim \sqrt{n}\left\|(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\left(\hat{T}_s^*\hat{T}_s-T_s^*T_s\right)\alpha_n(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\|_{-(a+s)} \\
&\leq 2S_{1n}+2S_{2n}
\end{aligned}$$

with

$$\begin{aligned}
S_{1n} &= \sqrt{n}\left\|(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\left(\hat{T}_s-T_s\right)\alpha_n(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\|_{-(a+s)} \\
&\lesssim \sqrt{n}\left\|T_s(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\left(\hat{T}_s-T_s\right)\alpha_n(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\| \\
&\lesssim_P \sqrt{n}\alpha_n^{1/2}\left\|\alpha_n(\alpha_n I+T_s^*T_s)^{-1}(T_s^*T_s)^{\frac{b-s}{2(a+s)}}\psi\right\| \\
&\lesssim_P \sqrt{n}\alpha_n \\
&= o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
S_{2n} &= \sqrt{n}\left\|(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\left(\hat{T}_s^*-T_s^*\right)\alpha_n T_s(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\|_{-(a+s)} \\
&\lesssim \sqrt{n}\left\|T_s(\alpha_n I+\hat{T}_s^*\hat{T}_s)^{-1}\left(\hat{T}_s^*-T_s^*\right)\alpha_n T_s(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\| \\
&\lesssim_P \sqrt{n}\left\|\alpha_n T_s(\alpha_n I+T_s^*T_s)^{-1}L^s\varphi\right\| \\
&\lesssim_P \sqrt{n}\alpha_n \\
&= o_P(1),
\end{aligned}$$

where the second line follows by Engl, Hanke, and Neubauer (1996), Corollary 8.22, the third since $\|\hat{T}-T\|\lesssim_P \alpha_n^{1/2}$, and the last by computations similar to the proof of Theorem A.1 and $2s=b-a$.

Next, decompose

$$\begin{aligned}
I_n(u) &= \sqrt{n}\left\langle(\alpha_n I+\hat{T}^*\hat{T})^{-1}\hat{T}^*(\hat{r}-\hat{T}\varphi), f_{UZ}(u,.)\right\rangle \\
&= \sqrt{n}\left\langle\hat{r}-\hat{T}\varphi, T(\alpha_n I+T^*T)^{-1}f_{UZ}(u,.)\right\rangle+R_{1n}(u)+R_{2n}(u)
\end{aligned}$$

with

$$\begin{aligned}
R_{1n}(u) &= \sqrt{n}\left\langle\hat{r}-\hat{T}\varphi, (\hat{T}-T)(\alpha_n I+T^*T)^{-1}f_{UZ}(u,.)\right\rangle \\
R_{2n}(u) &= \sqrt{n}\left\langle\hat{r}-\hat{T}\varphi, \hat{T}\left[(\alpha_n I+\hat{T}^*\hat{T})^{-1}-(\alpha_n I+T^*T)^{-1}\right]f_{UZ}(u,.)\right\rangle.
\end{aligned}$$

By the Cauchy-Schwartz inequality, we bound the first remainder term as

$$\begin{aligned}
\|R_{1n}\|_\infty &\leq \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T} - T\|\|(\alpha_n I + T^*T)^{-1}(T^*T)^{1/2}\| \sup_u \|(T^*T)^{-1/2} f_{UZ}(u, .)\| \\
&\lesssim_P \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T} - T\|\alpha_n^{-1/2} \sup_u \|f_{UZ}(u, .)\|_a \\
&\lesssim_P \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T} - T\|\alpha_n^{-1/2} \\
&\lesssim_P \sqrt{n}\left(\frac{1}{\sqrt{nh_n^q}} + h_n^t\right)\left(\frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t\right)\alpha_n^{-1/2} \\
&= o_P(1),
\end{aligned}$$

where the second line follows by Engl, Hanke, and Neubauer (1996), Corollary 8.22., the third under Assumption 4.1 (i), the fourth from the proof of Corollary A.1.1, and the last under Assumption 4.1 (ii).

The second remainder term is decomposed as $R_{2n}(u) = -R_{2n}^I(u) - R_{2n}^{II}(u)$ with

$$\begin{aligned}
R_{2n}^I(u) &= \sqrt{n}\left\langle \hat{r} - \hat{T}\varphi, \hat{T}(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*(\hat{T} - T)(\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\right\rangle \\
R_{2n}^{II}(u) &= \sqrt{n}\left\langle \hat{r} - \hat{T}\varphi, \hat{T}(\alpha_n I + \hat{T}^*\hat{T})^{-1}(\hat{T}^* - T^*)T(\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\right\rangle,
\end{aligned}$$

where we bound in the same way as $\|R_{1n}\|_\infty$

$$\begin{aligned}
\|R_{2n}^I\|_\infty &\leq \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T}(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\|\|\hat{T} - T\|\|(\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\| \\
&\lesssim \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T} - T\|\alpha_n^{-1/2} \\
&= o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
\|R_{2n}^{II}\|_\infty &\leq \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\|\hat{T}(\alpha_n I + \hat{T}^*\hat{T})^{-1}\|\|\hat{T}^* - T^*\|\|T(\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\| \\
&\lesssim \sqrt{n}\|\hat{r} - \hat{T}\varphi\|\alpha_n^{-1/2}\|\hat{T} - T\| \\
&= o_P(1).
\end{aligned}$$

Therefore, uniformly over $u \in \mathbf{R}$

$$I_n(u) = \sqrt{n}\left\langle T^*(\hat{r} - \hat{T}\varphi), (\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\right\rangle + o_P(1).$$

Note that

$$(\hat{r} - \hat{T}\varphi)(w) = \frac{1}{n}\sum_{i=1}^n (Y_i - [\varphi * K_z](Z_i))h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right)$$

with $[\varphi * K_z](z) \triangleq \int \varphi(v) h_n^{-p} K_z\left(h_n^{-1}(z-v)\right) \mathrm{d}v$ and $Y_i = \varphi(Z_i) + U_i$. Note also that

$$T^*\left[h_n^{-q} K_w\left(h_n^{-1}(W_i - .)\right)\right](z) = [f_{ZW} * K_w](z, W_i)$$
$$\left\langle f_{ZW}(., w), (T^*T)^{-1} f_{UZ}(u, .)\right\rangle = [T(T^*T)^{-1} f_{UZ}(u, .)](w)$$

with $[f_{ZW} * K_w](z, w) \triangleq \int f_{ZW}(z, v) h_n^{-q} K_w\left(h_n^{-1}(w-v)\right) \mathrm{d}u$. Using these observations, decompose further

$$I_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i [T(T^*T)^{-1} f_{UZ}(u, .)](W_i) + Q_{1n} + Q_{2n} + Q_{3n} + o_P(1)$$

with

$$Q_{1n}(u) = \left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^n [\varphi - \varphi * K_z](Z_i)[f_{ZW} * K_w](., W_i), (\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\right\rangle$$
$$Q_{2n}(u) = \left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \left\{[f_{ZW} * K_w](., W_i) - f_{ZW}(., W_i)\right\}, (\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\right\rangle$$
$$Q_{3n}(u) = \left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i f_{ZW}(., W_i), \left[(T^*T)^{-1} - (\alpha_n I + T^*T)^{-1}\right] f_{UZ}(u, .)\right\rangle.$$

We show that the first remainder term is negligible by the Cauchy-Schwartz and the triangle inequalities under Assumption 4.1 (ii)

$$\begin{aligned}
\|Q_{1n}\|_\infty &\leq \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n [\varphi - \varphi * K_z](Z_i)[f_{ZW} * K_w](., W_i)\right\| \sup_u \|(\alpha_n I + T^*T)^{-1} f_{UZ}(u, .)\| \\
&\lesssim \frac{1}{\sqrt{n}} \sum_{i=1}^n |[\varphi - \varphi * K_z](Z_i)| \, \|[f_{ZW} * K_w](., W_i)\| \, \alpha_n^{-1/2} \\
&\lesssim_P \sqrt{n} h_n^b \alpha_n^{-1/2} \\
&= o_P(1),
\end{aligned}$$

where the second line follows under Assumption 4.1 (i), the third by

$$\begin{aligned}
&\mathbb{E}\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n |[\varphi - \varphi * K_z](Z_i) \, \|[f_{ZW} * K_w](., W_i)\| \, | \right| = \\
&= \sqrt{n} \mathbb{E} \, |[\varphi - \varphi * K_z](Z) \, \|[f_{ZW} * K_w](., W)\|| \\
&\leq \sqrt{n} \|\varphi - \varphi * K_z\| \|f_{ZW} * K_w\| \\
&= O(\sqrt{n} h_n^b),
\end{aligned}$$

where the second line follows under the i.i.d. assumption, the third by the Cauchy-Schwartz inequality and since f_Z and f_W are uniformly bounded under Assumption A.1.2 (ii), and the last by the standard bias computations under Assumptions A.1.1 (ii) and A.1.2 (iv), Young's inequality, and Assumption A.1.2 (ii) and (iv).

Similarly, by the Cauchy-Schwartz inequality

$$\begin{aligned}
\mathbb{E}\|Q_{2n}\|_\infty^2 &\lesssim \mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i\left\{[f_{ZW}*K_w](.,W_i)-f_{ZW}(.,W_i)\right\}\right\|^2\alpha_n^{-1/2} \\
&= \mathbb{E}\left\|U\left\{[f_{ZW}*K_w](.,W)-f_{ZW}(.,W)\right\}\right\|^2\alpha_n^{-1/2} \\
&\lesssim \mathbb{E}\|[f_{ZW}-f_{ZW}*K_w](.,W)\|^2\alpha_n^{-1/2} \\
&\lesssim \|f_{ZW}-f_{ZW}*K_w\|\alpha_n^{-1/2} \\
&= O(h_n^t\alpha_n^{-1/2}) \\
&= o(1),
\end{aligned}$$

where the second line follows under the i.i.d. assumption, the third since $\mathbb{E}[U|W]\leq C$ under Assumption A.1.2 (i), the fourth since f_W is uniformly bounded under Assumption A.1.2 (ii), the fifth by the standard bias computations under Assumptions A.1.1 (ii) and A.1.2 (iv), and the last since $h_n^{2t}/\alpha_n\to 0$ under Assumption 4.1 (ii).

Lastly, by the Cauchy-Schwartz inequality

$$\begin{aligned}
\|Q_{3n}\|_\infty &= \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i f_{ZW}(.,W_i)\right\|\sup_u\left\|\left[(T^*T)^{-1}-(\alpha_n I+T^*T)^{-1}\right]f_{UZ}(u,.)\right\| \\
&\lesssim_P \left\|\alpha_n(\alpha_n I+T^*T)^{-1}(T^*T)^{\kappa/2a-1}\right\|\sup_u\|f_{UZ}(u,.)\|_\kappa \\
&\lesssim \sup_\lambda\left|\frac{\alpha_n\lambda^{\kappa/2a-1}}{\alpha_n+\lambda}\right| \\
&= O(\alpha_n^{(\kappa/2a-1)\wedge 1}) \\
&= o(1),
\end{aligned}$$

where the second inequality follows under the i.i.d. assumption and Assumptions A.1.2 (i)-(ii), the third under Assumption 4.1 (i), and the last since $\kappa>2a$ under the same assumption.

Combining all estimates together, we obtain uniformly in $u\in\mathbf{R}$

$$\sqrt{n}(\hat{F}_{\hat{U}}(u)-F_U(u))=\sqrt{n}(\hat{F}_U(u)-F_U(u))+\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i[T(T^*T)^{-1}f_{UZ}(u,.)](W_i)+o_P(1).$$

□

Proof of Corollary 4.1. The process given in Theorem 4.1 is an empirical process indexed by the following class of functions

$$\mathcal{F} = \left\{(v, w) \mapsto \mathbf{1}_{\{v \leq u\}} + v\left(T(T^*T)^{-1} f_{UZ}(u, .)\right)(w),\ u \in \mathbf{R}\right\},$$

which is a sum of the classical Donsker class of indicator functions and the class

$$\mathcal{H} = \left\{(v, w) \mapsto v\left(T(T^*T)^{-1} f_{UZ}(u, .)\right)(w),\ u \in \mathbf{R}\right\}.$$

By Van Der Vaart and Wellner (2000), Theorem 2.10.6, it enough to show that $\mathcal{H}$ is Donsker. The former statement follows from the fact that $f_{ZW} \in H^t, t > q/2$, whence under Assumption 4.1 (i), there exists some $M < \infty$, such that $\|h\|_t \leq M$ for all $h \in \mathcal{H}$. Under Assumption A.1.2 (i), this shows that the class $\mathcal{H}$ is Donsker; see, e.g., Nickl and Pötscher (2007), Corollaries 4 and 5. □

Next, we look at the independence empirical process. The following assumption mirrors Assumption 4.1.

Assumption A.2.1. *(i) $\|\partial_u f_{UZ}\|_\infty < \infty$ and $\sup_u \|\partial_u f_{UZ}(u, .)\|_\kappa < \infty$ with $\kappa > 2a$; (ii) $\left\|\int_{\{v \leq .\}} \partial_u f_{UZW}(., .v)\mathrm{d}v\right\|_\infty < \infty$ and $\sup_{u,w} \left\|\int^w f_{UZW}(u, ., v)\mathrm{d}v\right\|_\kappa < \infty$ with $\kappa > 2a$; (iii) $h_n \to 0$ and $\alpha_n \to 0$ are such that $\alpha_n^{\frac{4a}{b+a}} nh_n^{2q} \to \infty$, $\alpha_n^{\frac{2(a+c)}{b+a}} nh_n^q \to \infty$, $\alpha_n^{-\frac{4a}{b+a} \vee 1} nh_n^{4t} \to 0$, $\alpha_n^2 n \to 0$, $\alpha_n nh_n^{p+2q} \to \infty$ $nh_n^{2b}/\alpha_n \to 0$, and $h_n^{2t-p-q}/\alpha_n \to 0$ with $c > p/2$, $t > q/2$, $2s = b - a$, $b \geq a$ and a, b, t, p, q are as in Assumptions A.1.1 and A.1.2.*

Assumption A.2.1 (ii) is an additional smoothness condition compared to Assumption 4.1. Similarly to Assumption A.2.1 (i), it requires the existence of a uniformly bounded derivative of a joint distribution/density and its Sobolev smoothness.

Lemma A.2.2. *Suppose that Assumptions A.1.1, A.1.2, and A.2.1 are satisfied. Then uniformly over $(u, w) \in \mathbf{R} \times \mathbf{R}^q$*

$$(\hat{F}_{\hat{U}}(u) - \hat{F}_U(u))\hat{F}_W(w) - \left(\Pr(U \leq u + \hat{\Delta}(Z)|\mathscr{X}) + F_U(u)\right) F_W(w) = o_P\left(n^{-1/2}\right)$$

and

$$\hat{F}_{\hat{U}W}(u, w) - \hat{F}_{UW}(u, w) - \Pr(U \leq u + \hat{\Delta}(Z), W \leq w|\mathscr{X}) + F_{UW}(u, w) = o_P\left(n^{-1/2}\right).$$

where $\hat{\Delta} = \hat{\varphi} - \varphi$ and $\mathscr{X} = (Y_i, Z_i, W_i)_{i=1}^\infty$, *Proof.* Note that the first expression and the expression in the statement of Lemma A.2.1 multiplied by F_W differ only by

$$(\hat{F}_{\hat{U}}(u) - F(u))(\hat{F}_W(w) - F_W(w)),$$

which is $O_P(n^{-1})$ by Corollary 4.1 and the maximal inequality. This proves the first statement since F_W is bounded by one.

The proof of the second statement is similar to the proof of Lemma A.2.1 and is, therefore, omitted. □

Proof of Theorem 3.1. By Lemma A.2.2, uniformly in (u, w)

$$\mathbb{G}_n(u, w) = T_{1n}(u, w) + T_{2n}(u, w) - T_{3n}(u, w) + o_P(1),$$

where

$$\begin{aligned}
T_{1n}(u, w) &= \sqrt{n}\left(\hat{F}_{UW}(u, w) - \hat{F}_U(u)\hat{F}_W(w)\right),\\
T_{2n}(u, w) &= \sqrt{n}\left(\Pr\left(U \le u + \hat{\Delta}(Z), W \le w | \mathscr{X}\right) - F_{UW}(u, w)\right),\\
T_{3n}(u, w) &= \sqrt{n}\left(\Pr\left(U \le u + \hat{\Delta}(Z) | \mathscr{X}\right) - F_U(u)\right) F_W(w).
\end{aligned}$$

The first term is a classical independence empirical process

$$\begin{aligned}
T_{1n}(u, w) = &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathbf{1}_{\{U_i \le u, W_i \le w\}} - \mathbf{1}_{\{U_i \le u\}}F_W(w) - \mathbf{1}_{\{W_i \le w\}}F_U(u) + F_U(u)F_W(w)\right\}\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathbf{1}_{\{W_i \le w\}} - F_W(w)\right\}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{1}_{\{U_i \le u\}} - F_U(u)\right\}\\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathbf{1}_{\{U_i \le u, W_i \le w\}} - \mathbf{1}_{\{U_i \le u\}}F_W(w) - \mathbf{1}_{\{W_i \le w\}}F_U(u) + F_U(u)F_W(w)\right\} + o_P(1),
\end{aligned}$$

where the second line follows by the maximal inequality.

Next, under Assumption A.2.1 (i), by Taylor's theorem, for some $\tau \in [0, 1]$

$$\begin{aligned}
T_{2n}(u, w) &= \sqrt{n}\int\int^{w}\left\{\int_{-\infty}^{u+\hat{\Delta}(z)} f_{UZW}(\tilde{u}, z, v)\mathrm{d}\tilde{u} - \int_{-\infty}^{u} f_{UZW}(\tilde{u}, z, v)\mathrm{d}\tilde{u}\right\}\mathrm{d}v\mathrm{d}z\\
&= \sqrt{n}\int\int^{w}\left\{f_{UZW}(u, z, \tilde{w})\hat{\Delta}(z) + \frac{1}{2}\partial_u f_{UZW}(u + \tau\hat{\Delta}(z), z, v)\hat{\Delta}^2(z)\right\}\mathrm{d}v\mathrm{d}z\\
&= \sqrt{n}\left\langle \hat{\varphi} - \varphi, \int^{w} f_{UZW}(u, ., v)\mathrm{d}v\right\rangle + \frac{\sqrt{n}}{2}\int\int^{w}\partial_u f_{UZW}(u + \tau\hat{\Delta}(z), z, v)\mathrm{d}v\hat{\Delta}^2(z)\mathrm{d}z\\
&\triangleq S_{1n}(u, w) + S_{2n}(u, w).
\end{aligned}$$

Under Assumptions A.2.1

$$\begin{aligned}\|S_{2n}\|_\infty &\le \sup_{w,u,z}\left|\int^w \partial_u f_{UZW}(u,z,v)\mathrm{d}v\right| \sqrt{n}\,\|\hat\varphi-\varphi\|^2\\ &= o_P(1)\end{aligned}$$

by Corollary A.1.1. Similarly,

$$T_{3n}(u,w) = \sqrt{n}\langle \hat\varphi - \varphi, f_{UZ}(u,.)\rangle F_W(w) + o_P(1).$$

Therefore, uniformly in $(u,w) \in \mathbf{R}\times\mathbf{R}^q$

$$\begin{aligned}&T_{2n}(u,w) - T_{3n}(u,w)\\ &= \sqrt{n}\int(\hat\varphi(z)-\varphi(z))\left\{\int^w f_{UZW}(u,z,\tilde w)\mathrm{d}\tilde w - f_{UZ}(u,z)F_W(w)\right\}\mathrm{d}z + o_P(1)\\ &= \sqrt{n}\int(\hat\varphi(z)-\varphi(z))g(u,w,.)\mathrm{d}z + o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i\left(T(T^*T)^{-1}g(u,w,.)\right)(W_i) + o_P(1),\end{aligned}$$

where the last line follows by the same argument as in the proof of Theorem 4.1 under Assumption A.2.1 (i). □

Proof of Corollary 3.2. The asymptotic distribution of $T_{n,\infty}$ under H_0 is readily obtained by the continuous mapping theorem; see Van Der Vaart and Wellner (2000), Theorem 1.3.6. For the Cramér-von Mises statistics, write

$$T_{n,2} = \iint \mathbb{G}^2(u,w)\mathrm{d}F_{UW}(u,w) + R_{1n} + R_{2n}$$

with

$$\begin{aligned}R_{1n} &= \iint\left\{\mathbb{G}_n^2(u,w) - \mathbb{G}^2(u,w)\right\}\mathrm{d}\hat F_{\hat U W}(u,w)\\ R_{2n} &= \iint \mathbb{G}^2(u,w)\mathrm{d}[\hat F_{\hat U W}(u,w) - F_{UW}(u,w)].\end{aligned}$$

By Corollary 3.1, under H_0, $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ and $\sqrt{n}(\hat F_{\hat U W}(u,w) - F_{UW}(u,w))$ also converges weakly by Corollary 3.1 and Theorem 4.1, whence by the Skorokhod construction

$$n^{-1/2}\sup_{u,w}|\mathbb{G}_n(u,w)| \xrightarrow{\text{a.s.}} 0 \qquad \text{and} \qquad \sup_{u,w}\left|\hat F_{\hat U W}(u,w) - F_{UW}(u,w)\right| \xrightarrow{\text{a.s.}} 0. \quad \text{(A.4)}$$

The first expression in Eq. A.4 implies that $R_{1n} \xrightarrow{\text{a.s.}} 0$. Since $\mathbb{G}$ has a.s. bounded and continuous trajectories, the second expression in Eq. A.4 in conjunction with the Helly-Bray theorem show that $R_{2n} \xrightarrow{\text{a.s.}} 0$. Therefore, the asymptotic distribution of the Cramér-von Mises test follows by the continuous mapping theorem.

Under the alternative hypothesis, by Theorem 3.1, the Glivenko-Cantelli theorem and a similar argument we obtain

$$n^{-1/2}T_{n,2} = \iint |n^{-1/2}\mathbb{G}_n(u,w)|^2 \mathrm{d}\hat{F}_{\hat{U}W}(u,w) \xrightarrow{\text{a.s.}} 2d_2 > 0$$
$$n^{-1/2}T_{n,\infty} = \sup_{u,w} |n^{-1/2}\mathbb{G}_n(u,w)| \xrightarrow{\text{a.s.}} 2d_\infty > 0.$$

Therefore, by Slutsky's theorem $T_{n,2} \xrightarrow{\text{a.s.}} \infty$ and $T_{n,\infty} \xrightarrow{\text{a.s.}} \infty$, which proves the second statement. $\square$