

# Are Unobservables Separable?

Andrii Babii\*  
*UNC Chapel Hill*

Jean-Pierre Florens†  
*Toulouse School of Economics*

March 31, 2021

## 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, which may be of independent interest. 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, Engel curves, heterogeneity in unobservables, distribution of nonparametric IV residuals.

---

\*Department of Economics, University of North Carolina–Chapel Hill - Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. Email: [babii.andrii@gmail.com](mailto:babii.andrii@gmail.com).

†Department of Economics, Toulouse School of Economics - 1, Esplanade de l'Université, 31080 Toulouse Cedex 06, France.

# 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 interactions between the two are allowed for. However, the economic theory and considerations often lead to nonseparable models. Prominent examples are demand functions, where the 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 are heterogeneous in unobservables.

In response to these empirical challenges, there is a growing literature studying the nonparametric identification of nonseparable models with endogeneity; see Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Imbens and Newey (2009), Torgovitsky (2015), and D'Haultfouille and Février (2015) among many others. It is well-understood that the fully nonparametric estimation of a 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).

Since a fully nonparametric estimation of a nonseparable model is more challenging and since separable models rule out the heterogeneity of marginal effects in 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 could be well-justified.

Despite the significant efforts focused on understanding the identification and the estimation of nonseparable IV models and the widespread use of separable IV models in 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 inde-*

*pendence 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 the monotonicity test of [Hoderlein, Su, White, and Yang \(2016\)](#), the endogeneity test of [Fève, Florens, and Van Keilegom \(2018\)](#), and the specification test for the quantile IV regression of [Breunig \(2020\)](#).

In this paper, we design a novel fully nonparametric separability test. Our test is based on the independence condition of the nonseparable model and does not rely on additional identifying restrictions, such as the monotonicity in unobservable. The test is based on the insight that the structural function in the separable model can be estimated using the nonparametric IV approach; see [Florens \(2003\)](#), [Newey and Powell \(2003\)](#), [Hall and Horowitz \(2005\)](#), [Blundell, Chen, and Kristensen \(2007\)](#), and [Darolles, Fan, Florens, and Renault \(2011\)](#). If the separable model is correct, then the nonparametric IV residuals should approximate unobservables that are 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 a linear *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 challenging, and it is remarkable that the empirical distribution of nonparametric regression residuals still converges weakly as was shown in [Akritas and](#)

Van Keilegom (2001). 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 practice.

In this paper, we show that the empirical distribution function of nonparametric IV residuals converges weakly to a Gaussian process at a parametric rate, even though residuals are obtained from the nonparametrically estimated ill-posed inverse problem. To the best of our knowledge, this is the first result on the distribution of the nonparametric IV residuals, which can be used to develop various residual-based specification tests and is of independent interest. Building on this result, we obtain the large sample approximation to the distributions of independence separability tests. The distributions of residual-based independence tests are non-standard and not amenable to standard bootstrap approximations. Therefore, we suggest using the  $m$  out of  $n$  bootstrap or subsampling to compute the critical values.

Our results are based on the insight that the Tikhonov regularization in Sobolev spaces, considered in Florens, Johannes, and Van Bellegem (2011), Gagliardini and Scaillet (2012), Carrasco, Florens, and Renault (2014), and Gagliardini and Scaillet (2017), among others, provides a natural link between the modern empirical process theory and the theory of ill-posed inverse problems. In regards to this literature, we obtain new results for the Tikhonov regularization with a Sobolev penalty that can be applied to generic ill-posed inverse problems, including various nonparametric IV estimators, e.g., based on kernel smoothing. In particular, the Tikhonov regularization with a Sobolev penalty achieves sufficiently fast convergence rates for the semiparametric theory. In contrast, the simple one-step Tikhonov regularization without Sobolev penalization suffers from the well-known saturation effects; see Darolles, Fan, Florens, and Renault (2011).

The paper is organized as follows. In Section 2, we present two motivating examples, where economic considerations lead to nonseparable models with endogeneity and discuss a testable implication of separability. In Section 3, we characterize the large sample ap-

proximation to the distribution of the residual-based Kolmogorov-Smirnov and Cramér-von Mises independence tests and introduce a resampling procedure to compute the critical values. We also study the behavior of these tests under the fixed and the local alternative hypotheses. We report on a Monte Carlo study in Section 4 which provides insights about the validity of our asymptotic approximations in finite samples. In Section 5, 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 6. All technical details, auxiliary results, and proofs are collected in the Appendix and the Supplementary Material.

## 2 Separability of unobservables

### 2.1 Motivating examples

The instrumental variable models with additively separable unobservables constitute a workhorse of modern empirical practice. However, the additive separability of unobservables is a restrictive modeling assumption that essentially rules out the heterogeneity of estimated causal structural effects in unobservables; see, e.g., Heckman (2001) or Imbens (2010). Indeed, the structural economic models typically lead to nonseparable unobservables as illustrated below.

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

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

where  $U(\cdot, \cdot)$  is a utility function,  $Q$  is a vector of demanded quantities,  $\varepsilon$  is an individual preference variable, unobserved by the econometrician,  $P$  is a vector of prices, and  $I$  is the income. The solution to this optimization problem leads to the nonseparable demand functions  $Q_j = \Phi(P, I, \varepsilon)$  for each good  $j = 1, \dots, J$  as shown in Brown and Walker (1989) and Lewbel (2001); see also Hoderlein and Vanhems (2018) 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  $\phi$  such that  $Y = \phi(Z, \varepsilon) - U$ , where  $Y$  is an output,  $Z$  are observed inputs,  $\varepsilon$  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 A testable implication

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

$$Y = \Phi(Z, \varepsilon), \quad \varepsilon \perp\!\!\!\perp W, \quad (1)$$

where  $Y \in \mathbf{R}$  is outcome,  $Z \in \mathbf{R}^p$  are regressors,  $\varepsilon \in \mathbf{R}$  is unobservable,  $W \in \mathbf{R}^q$  is a vector of instrumental variables, and  $\Phi : \mathbf{R}^p \times \mathbf{R} \rightarrow \mathbf{R}$  is a structural function. We assume that  $W$  are valid instrumental variables satisfying the exclusion restriction,  $\varepsilon \perp\!\!\!\perp W$ , and the relevance condition,  $W \not\perp\!\!\!\perp Z$ . Note that the independence exclusion restriction is a commonly used identifying condition for nonseparable models; see Chernozhukov, Fernandez-Val, Newey, Stouli, and Vella (2020), Blundell, Horowitz, and Parey (2017), Torgovitsky (2017), Torgovitsky (2015), D’Haultfouille and Février (2015), Dunker, Florens, Hohage, Johannes, and Mammen (2014), Gagliardini and Scaillet (2012), and Horowitz and Lee (2007) for recent examples and applications, as well as Chernozhukov and Hansen (2013), Matzkin (2013), and Imbens (2010) for the review of earlier econometrics literature on the identification of nonseparable models.

The independence condition  $\varepsilon \perp\!\!\!\perp W$  does not rule-out the heteroskedasticity in the distribution of  $Y$  conditionally on  $Z$  or  $W$ , which is often observed in the empirical practice. It also does not rule-out the heteroskedasticity in the distribution of unobservables  $\varepsilon$  conditionally on covariates  $Z$ . However, it rules out the heteroskedasticity of unobservables conditionally on the instrumental variable, which could be less restrictive, since the instrumental variable is univariate in typical applications. This leads to an interesting trade-off between the heterogeneity of causal structural effects in unobservables allowed for in the nonseparable model and the heteroskedasticity of unobservables conditionally on the instrumental variable allowed for in the separable model.

To develop the separability test, several strategies can be adopted. For instance, one could nonparametrically estimate the nonseparable model and check whether the separability holds. 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 condition of the nonseparable model. This approach corresponds to the principle behind Rao's score test in the parametric setting and is the one adopted in this paper.

We say that the model in equation (1) has a *separable representation* if there exists measurable functions  $\psi : \mathbf{R}^p \rightarrow \mathbf{R}$  and  $g : \mathbf{R} \rightarrow \mathbf{R}$  such that

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

If the model has a separable representation, then the structural function can be estimated consistently using the nonparametric IV approach; see Darolles, Fan, Florens, and Renault (2011), Blundell, Chen, and Kristensen (2007), Horowitz and Lee (2007), and Newey and Powell (2003). The nonparametric IV regression function  $\varphi : \mathbf{R}^p \rightarrow \mathbf{R}$  solves the functional equation

$$r(w) \triangleq \mathbb{E}[Y|W = w]f_W(w) = \int \varphi(z)f_{ZW}(z, w)dz \triangleq (T\varphi)(w), \quad (2)$$

where  $T : L_2(\mathbf{R}^p) \rightarrow L_2(\mathbf{R}^q)$  is an integral operator. Let  $U \triangleq Y - \varphi(Z)$  be the nonparametric IV regression error. Note that even if the model is nonseparable, we still have  $\mathbb{E}[U|W] = 0$  with  $U = Y - \varphi(Z)$  for  $\varphi$  solving the functional equation (2). The following result provides a convenient for us testable implication of separability, provided that  $U$  is unambiguously defined, see Appendix for a formal proof.

**Proposition 2.1.** *Suppose that there exists a unique solution to equation (2). If the model in equation (1) admits a separable representation, then  $U \perp\!\!\!\perp W$ .*

It is worth mentioning that the independence between  $U$  and  $W$  is only a *testable implication* of additive separability of unobservables. However, when the model is nonseparable, we have  $U = \Phi(Z, \varepsilon) - \varphi(Z) \triangleq h(Z, \varepsilon)$ , for some non-degenerate function  $h$  of  $(Z, \varepsilon)$ , which in many cases is not independent of  $W$ , because  $Z \not\perp\!\!\!\perp W$  by the relevance condition. Therefore, the independence test between  $U$  and  $W$  will have power against

many interesting deviations from the separability. Note also that Proposition 2.1 relies on the injectivity of  $T$ , which is known as a completeness condition, see Newey and Powell (2003) and Babii and Florens (2020), and does not require that the nonseparable model is identified; see, e.g., Chernozhukov and Hansen (2005) and Chen, Chernozhukov, Lee, and Newey (2014). Lastly, note that the additive separability is different from the multiplicative separability when  $Y = \psi(Z)g(\varepsilon)$ . However, when  $Y, \psi(Z)$ , and  $g(\varepsilon)$  are positive, we obtain the additively separable model after taking logs.

### 3 Independence test

In this section, we introduce tests of the independence condition characterized in Proposition 2.1. Formally, we focus on testing

$$H_0 : U \perp\!\!\!\perp W \quad \text{vs.} \quad U \not\perp\!\!\!\perp W.$$

$H_0$  is testable, provided that the nuisance parameter  $\varphi$  in  $U = Y - \varphi(Z)$  is replaced by the appropriate estimator.

#### 3.1 Tikhonov regularization in Sobolev spaces

We focus on the Tikhonov-regularized estimator penalized by the Sobolev norm to estimate the nuisance parameter  $\varphi$ ; see Carrasco, Florens, and Renault (2014), Gagliardini and Scaillet (2012), and Florens, Johannes, and Van Bellegem (2011). The attractive feature of this estimator is that it does not suffer from the well-known saturation bias and can achieve a sufficiently fast convergence rate for our asymptotic theory and more generally for semiparametric applications; see Corollary A.1.1 in the Appendix.

Let  $(L_2(\mathbf{R}^p), \|\cdot\|)$  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  $x \in \mathbf{R}^p$  and  $|\cdot|$  is a Euclidean norm on  $\mathbf{R}^p$ . Consider the operator  $L^s f = F^{-1}(\langle \cdot \rangle^s F f)$  defined for all  $f$  such that  $\|\langle \cdot \rangle^2 F f\| < \infty$ , where  $F$  is a Fourier transform on  $L_2(\mathbf{R}^p)$  with scaling  $(2\pi)^{-p/2}$ . Then 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.

Let  $(\hat{T}, \hat{r})$  be the kernel estimators of  $(T, r)$  in equation (2) computed as

$$\begin{aligned} \hat{r}(w) &= \frac{1}{nh_n^q} \sum_{i=1}^n Y_i K_w(h_n^{-1}(W_i - w)), & (\hat{T}\phi)(w) &= \int \phi(z) \hat{f}_{ZW}(z, w) dz, \\ \hat{f}_{ZW}(z, w) &= \frac{1}{nh_n^{p+q}} \sum_{i=1}^n K_z(h_n^{-1}(Z_i - z)) K_w(h_n^{-1}(W_i - w)), \end{aligned} \quad (3)$$

where  $K_z : \mathbf{R}^p \rightarrow \mathbf{R}$  and  $K_w : \mathbf{R}^q \rightarrow \mathbf{R}$  are kernel functions and  $h_n \rightarrow 0$  is a sequence of bandwidth parameters.

We estimate  $\varphi$  using the Tikhonov-regularized estimator penalized by the Sobolev norm with  $s \geq 0$

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

where  $\hat{T}$  and  $\hat{r}$  are as described above. It is easy to see that this problem has a 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  $\hat{T}_s$ .

### 3.2 Distribution of statistics

Let  $\hat{U}_i = Y_i - \hat{\varphi}(Z_i)$  be the nonparametric IV residuals and let

$$\hat{F}_{\hat{U}W}(u, w) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i \leq u, W_i \leq w\}}, \quad \hat{F}_{\hat{U}}(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i \leq u\}}, \quad \hat{F}_W(w) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{W_i \leq w\}} \quad (4)$$

be the empirical distribution functions. To test  $H_0$ , we focus on the following residual-based independence empirical process

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

Note that this process involves residuals  $\hat{U}_i$  instead of the true regression errors  $U_i$ , hence, its asymptotic behavior can be significantly different from the asymptotic behavior of classical independence empirical processes; see [van der Vaart and Wellner \(1996\)](#), Chapter 3.8. In particular, the estimation of the nuisance component  $\varphi$  may affect the asymptotic distribution of the independence empirical process.

To understand the behavior of  $\mathbb{G}_n$ , we introduce several assumptions.

**Assumption 3.1.** For some  $a, b > 0$

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

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

Assumption 3.1 (i) describes the smoothing property of the operator  $T$ . Roughly speaking, the action of  $T$  increases the Sobolev smoothness by  $a$ , which is called the degree of ill-posedness. Intuitively, the more  $T$  smooths out features of  $\varphi$ , the harder it is to recover it from the equation (2). Condition (ii) describes the smoothness of the structural function  $\varphi$  and is a standard smoothness restriction in the nonparametric literature.

**Assumption 3.2.** (i)  $(Y_i, Z_i, W_i)_{i=1}^n$  are i.i.d. observations of  $(Y, Z, W)$  with  $\mathbb{E}|Y|^2 < \infty$ ,  $\mathbb{E}\|W\| < \infty$ ,  $\mathbb{E}\|Z\| < \infty$ , and  $\mathbb{E}[U^2|W] \leq C < \infty$ ; (ii) the distribution of  $(U, Z, W)$  is absolutely continuous with respect to the Lebesgue measure with densities  $f_Z, f_W, f_{ZW}, f_{U|Z} \in L_\infty$  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 continuous kernel  $K \in L_2(\mathbf{R}) \cap L_\infty(\mathbf{R})$  of bounded variation with  $\int K(u)du = 1$ ,  $\int |u|^l |K(u)|du < \infty$ , and  $\int u^k K(u)du = 0$  for  $k \in \{1, \dots, l\}$  and  $l \geq t$ .

Assumption 3.2 describes several mild conditions on the distribution of the data and the kernel functions that are largely standard for kernel estimators; see also Darolles, Fan, Florens, and Renault (2011), Appendix B for a discussion of generalized boundary kernels that can be used when supports are bounded. To introduce the next assumption, let  $\partial_u$  be a partial derivative with respect to the variable  $u$ , let  $\|\cdot\|_\infty$  denote the uniform norm, and put  $x \vee y = \max\{x, y\}$  and  $x \wedge y = \min\{x, y\}$ .

**Assumption 3.3.** (i)  $\|\partial_u f_{UZ}\|_\infty < \infty$  and  $\sup_u \|f_{UZ}(u, \cdot)\|_\kappa < \infty$  with  $\kappa > 2a \vee (a + q/2)$ ; (ii)  $\left\| \int_{\{v \leq \cdot\}} \partial_u f_{UZW}(\cdot, v)dv \right\|_\infty < \infty$  and  $\sup_{u,w} \left\| \int^w f_{UZW}(u, \cdot, v)dv \right\|_\kappa < \infty$  with  $\kappa > 2a \vee (a + q/2)$ ;

Assumption 3.3 imposes some relatively mild smoothness conditions on the distribution of the data.

**Assumption 3.4.**  $h_n \rightarrow 0$  and  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$  are such that (i)  $nh_n^q \alpha_n^{2(a+c)/(a+b)} \rightarrow \infty$ ,  $nh_n^{p+q} \alpha_n \rightarrow \infty$ , and  $h_n^{2t}/\alpha_n \rightarrow 0$ ; (ii)  $\sqrt{n} \alpha_n^{2b/(a+b)} \rightarrow 0$ ,  $\sqrt{n} h_n^q \alpha_n^{2a/(a+b)} \rightarrow \infty$ , and

$\sqrt{nh_n^{2t}}/\alpha_n^{2a/(a+b)} \rightarrow 0$ ; (iii)  $n\alpha_n^2 \rightarrow 0$ ,  $nh_n^{2(b \wedge 2t)} \rightarrow 0$ , and  $nh_n^{p+2q} \rightarrow \infty$ ; where  $2s = b - a \geq 0$ ,  $b > c$ ,  $s \geq c > p/2$ ,  $t > (p + q)/2$ ,  $a, b, t, p, q$  are as in Assumptions 3.1 and 3.2.

Assumption 3.3 (i) provides a set of sufficient conditions for  $\|\hat{\varphi} - \varphi\|_c = o_P(1)$  with  $c > p/2$ , while condition (ii) states additional requirements for  $\|\hat{\varphi} - \varphi\| = o_P(n^{-1/4})$ ; see Corollary A.1.1 in the Appendix. The former condition is needed for the asymptotic equicontinuity argument, while the latter requires that the nuisance parameter  $\varphi$  is estimated at a sufficiently fast rate, which is often encountered in the semiparametric literature. Lastly, condition (iii) ensures that a certain uniform asymptotic expansion holds. To illustrate that conditions on tuning parameters are feasible, 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  $c_1, c_2 \in (0, 1)$ . Then (i) requires that  $c_1 + 2c_2(a + c)/(a + b) < 1$ ,  $2c_1 + c_2 < 1$ , and  $t > c_2/2c_1$ . For (ii), we additionally need  $c_2 > (a + b)/4b$ ,  $c_1 + 2c_2a/(a + b) < 0.5$ , and  $t > 1/4c_1 + c_2a/c_1(a + b)$ . Lastly, (iii) requires that  $c_2 > 0.5$ ,  $c_1 > 1/2(b \wedge 2t)$ , and  $c_1 < 1/3$ . Therefore, we require  $(c_1, c_2) \in \{(x, y) \in \mathbf{R}^2 : 0.5 < y < 1 - 2x, x \in (1/2(b \wedge 2t), 1/3)\}$ , which is non-empty provided that  $b \wedge 2t > 3/2$ . Given this choice, the following smoothness conditions are imposed in Assumption 3.4:  $t > [1/4c_1 + c_2a/c_1(a + b)] \wedge [c_2/2c_1]$  and  $b > [2c_2(a + c)/(1 - c_1) - a] \wedge [2c_2a/(0.5 - c_1) - a] \wedge [1/(c_2 - 0.25)]$ .

The following result describes a convenient for us approximation to the residual-based independence empirical process:

**Theorem 3.1.** *Suppose that Assumptions 3.1, 3.2, 3.3, and 3.4 are satisfied. Then*

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

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

$$\begin{aligned} \delta_{u,w}(U_i, W_i) &= U_i (T(T^*T)^{-1} \rho(u, \cdot, w)) (W_i), \\ \rho(u, z, w) &= \int^w f_{UZW}(u, z, \tilde{w}) d\tilde{w} - f_{UZ}(u, z) F_W(w). \end{aligned}$$

It is worth mentioning that Theorem 3.1 does not require  $U \perp\!\!\!\perp W$ . The proof of this result can be found in the Appendix and relies on the asymptotic equicontinuity arguments. Roughly speaking, we show that the consistency of the nonparametric IV estimator in the Sobolev norm together with the Donsker property of Sobolev balls imply that that

certain terms associated with residuals are asymptotically negligible. At the same time, the estimation of the nuisance component  $\varphi$  has a first-order asymptotic effect due to the  $\delta_{u,w}(U_i, W_i)$  term, while the higher-order terms are negligible provided that  $\|\hat{\varphi} - \varphi\| = o_P(n^{-1/4})$ . This rate condition is typically encountered for the semiparametric problems; see Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) and Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2016) for recent contributions, Andrews (1994) for earlier treatment, and Babii (2021), Section 3.3 for a related discussion in the setting of ill-posed inverse problems.

It is worth mentioning that, in some cases, the estimation of nuisance parameters does not have any first-order asymptotic effect, which is known as the Neyman orthogonality property in the semiparametric literature. In particular, this is the case for the independence empirical process based on the nonparametric conditional mean regression residuals; see Einmahl and Van Keilegom (2008). Interestingly, if we had  $W \perp\!\!\!\perp (U, Z)$ , then  $\rho = 0$ , and the estimation of  $\varphi$  would not have any first-order asymptotic effect.

Theorem 3.1 can be readily used to construct the residual-based Cramér-von Mises and Kolmogorov-Smirnov statistics

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

To understand the behavior of the two statistics under the null and the alternative hypotheses, consider a centered version of the process in Theorem 3.1

$$\mathbb{H}_n(u, w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_{u,w}(U_i, W_i) - \mathbb{E}[h_{u,w}(U_i, W_i)],$$

where  $h_{u,w}(U, W) = \mathbb{1}_{\{U \leq u, W \leq w\}} - \mathbb{1}_{\{U \leq u\}}F_W(w) - \mathbb{1}_{\{W \leq w\}}F_U(u) + F_{UW}(u, w) + \delta_{u,w}(U, W)$ .

The following Donsker-type central limit theorem holds:

**Proposition 3.1.** *Suppose that assumptions of Theorem 3.1 are satisfied. Then*

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

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

$$(u, w, u', w') \mapsto \mathbb{E}[(h_{u,w}(U, W) - \mathbb{E}[h_{u,w}(U, W)])(h_{u',w'}(U, W) - \mathbb{E}[h_{u',w'}(U, W)])].$$

Note that under the null hypothesis  $H_0 : U \perp\!\!\!\perp W$ , we have  $\mathbb{E}[h_{u,w}(U, W)] = 0$  and the covariance function of  $\mathbb{H}$  simplifies to

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

For the alternative hypothesis,  $H_1 : U \not\perp\!\!\!\perp W$ , put

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

Consider also a sequence of local alternative hypotheses

$$H_{1,n} : F_{UW}(u, w) = F_U(u)F_W(w) + n^{-1/2}H(u, w), \quad \forall u, w,$$

where the function  $H$  is such that  $F_{UW}$  is a proper CDF. There exist several ways to construct such local alternatives with prespecified marginal distributions  $F_U$  and  $F_W$ . For instance, the Morgenstern's family is  $F_{UW}(u, w) = F_U(u)F_W(w) + aF_U(u)F_W(w)(1 - F_U(u))(1 - F_W(w))$  with  $a \in [-1, 1]$ ; see [Devroye \(1986\)](#), Chapter XI, Theorem 3.2. The following corollary describes the behavior of the independence test under the null and fixed/local alternative hypotheses:

**Corollary 3.1.** *Suppose that assumptions of Theorem 3.1 are satisfied. Then under  $H_0$*

$$T_{2,n} \rightsquigarrow \iint |\mathbb{H}(u, w)|^2 dF_{UW}(u, w) \quad \text{and} \quad T_{\infty,n} \rightsquigarrow \sup_{u,w} |\mathbb{H}(u, w)|,$$

while under  $H_1$ , we have  $T_{2,n}, T_{\infty,n} \xrightarrow{\text{a.s.}} \infty$ , provided that  $d_2, d_\infty > 0$ . Moreover, under  $H_{1,n}$

$$T_{2,n} \rightsquigarrow \iint |\mathbb{H}(u, w) + 2H(u, w)|^2 dF_{UW}(u, w) \quad \text{and} \quad T_{\infty,n} \rightsquigarrow \sup_{u,w} |\mathbb{H}(u, w) + 2H(u, w)|.$$

Corollary 3.1 shows that the residual-based independence test can detect parametric local alternatives. The asymptotic distributions under  $H_0$  are 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 could be an attractive alternative for simulating the critical values of the test. Interestingly, the naive nonparametric and the multiplier bootstraps do not work.

### 3.3 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 A.1 and Corollary A.2.1. 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 (2021) 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. Note also 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 4, we find in Monte Carlo experiments that the standard nonparametric bootstrap does not work.

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 such that  $m_n \rightarrow \infty$  and  $m_n/n \rightarrow 0$  and as  $n \rightarrow \infty$ .
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. Alternatively, compute p-values

as  $1 - F_n^*(T_n)$ , where  $T_n$  is the statistics computed from  $(Y_i, Z_i, W_i)_{i=1}^n$  and  $F_n^*$  is the empirical distribution function of bootstrapped statistics.

An attractive feature of 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., in [Bickel and Sakov \(2008\)](#).

## 4 Monte Carlo experiments

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

$$Y = \varphi(Z) + \theta ZU + U, \quad \begin{pmatrix} Z \\ W \\ U \end{pmatrix} \sim_{i.i.d.} N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 & 0.3 \\ 0.4 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix} \right).$$

We set  $\varphi(x) = \cos(x)$  and consider samples of size  $n = 500$  and  $n = 1,000$  observations; see Supplementary Material for additional simulation results. Note that the degree of separability of unobservables 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. It is worth mentioning that under  $H_1$ , the nonparametric IV regression does not estimate consistently the nonseparable structural function  $(z, u) \mapsto \cos(z) + \theta zu$ , which depends on unobservables. The nonparametric IV regression estimates instead the function  $z \mapsto \phi(z)$  solving the functional equation  $\mathbb{E}[Y|W] = \mathbb{E}[\phi(Z)|W]$ . The difference between the two functions is precisely what gives the power to the test.

We set the number of Monte Carlo replications and the number of bootstrap replications to 1,000 through all our experiments. We also discretize all continuous quantities on the grid of 100 equidistant points in  $[-4, 4]$ . The estimates  $\hat{r}$  and  $\hat{T}$  in equation (3) are obtained using the sixth-order Epanechnikov kernel. The corresponding bandwidth parameters are computed using Silverman's rule of thumb:  $h_z = 3.53\hat{\sigma}_z n^{-1/13}$  and  $h_w = 3.53\hat{\sigma}_w n^{-1/13}$ , where  $\hat{\sigma}_z$  and  $\hat{\sigma}_w$  are sample standard deviations of observed  $Z$  and  $W$ . This choice satisfies Assumption 3.4 and requires that the regularization parameter is  $\alpha_n \sim n^{-c_2}$  with  $c_2 \in (0.5, 11/13)$ . To satisfy this requirement, we set  $\alpha_n = n^{-4/5}$ .

We look at the distributions of Kolmogorov-Smirnov and Cramér-von Mises statistics, computed respectively as

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

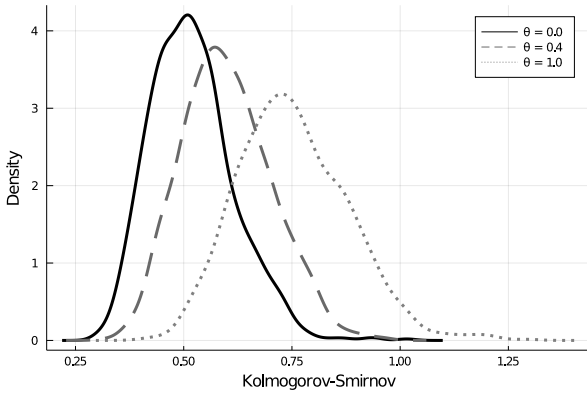
where  $\mathbb{G}_n(u, w) = \sqrt{n}(\hat{F}_{\hat{U}W}(u, w) - \hat{F}_{\hat{U}}(u)\hat{F}_W(w))$  and the empirical distribution functions are computed as in equation (4). Lastly, we use the adaptive rule of [Bickel and Sakov \(2008\)](#) to estimate the size of the subsample. The rule consists of choosing  $\hat{m}_j = \arg \min_{j \geq 0} \sup_x |F_j^*(x) - F_{j+1}^*(x)|$ , where  $F_j^*$  is the empirical distribution of the simulated statistics using a subsample of size  $m_j = [q^j n]$ ,  $j = 0, 1, 2, \dots, 4$ ,  $[a]$  is integer part of  $a$ , and  $q = 0.5$ .

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 becomes more separated from the null hypothesis.

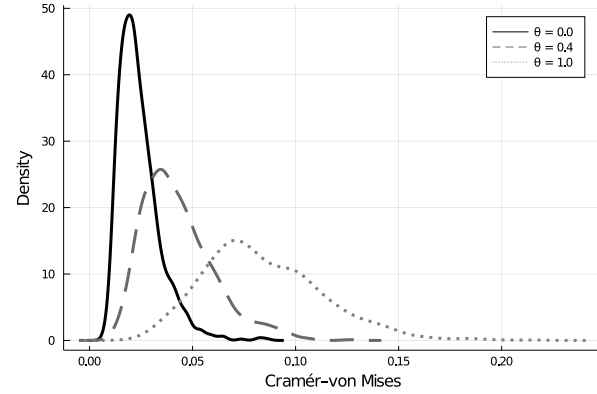
We plot in Figure 2 the power curves when the level of the test is fixed at 5%. The power of the test increases once alternative hypotheses become more distant from the null hypothesis. The Cramér-von Mises test seems to have higher power for the class of considered alternatives. We can also see that the figure illustrates the consistency of the test in the sense that its power becomes closer to one as the sample size increases under the alternative hypotheses.

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. The naive bootstrap fails and does not mimic the distribution of the Kolmogorov-Smirnov/Cramér-von Mises statistics. The distribution of the  $m$  out of  $n$  bootstrap, on the other hand, is close to the finite sample distributions of both statistics. We also observe that the adaptive choice seems to work slightly better for the Cramér-von Mises statistics.

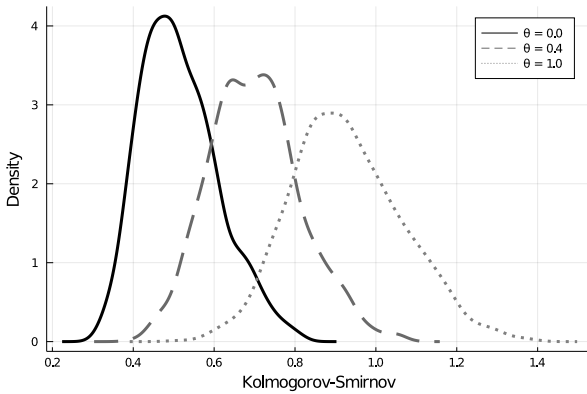




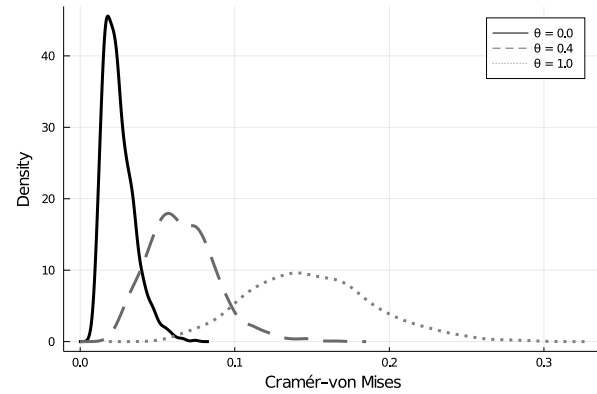
(a) Sample size:  $n = 500$



(b) Sample size:  $n = 500$

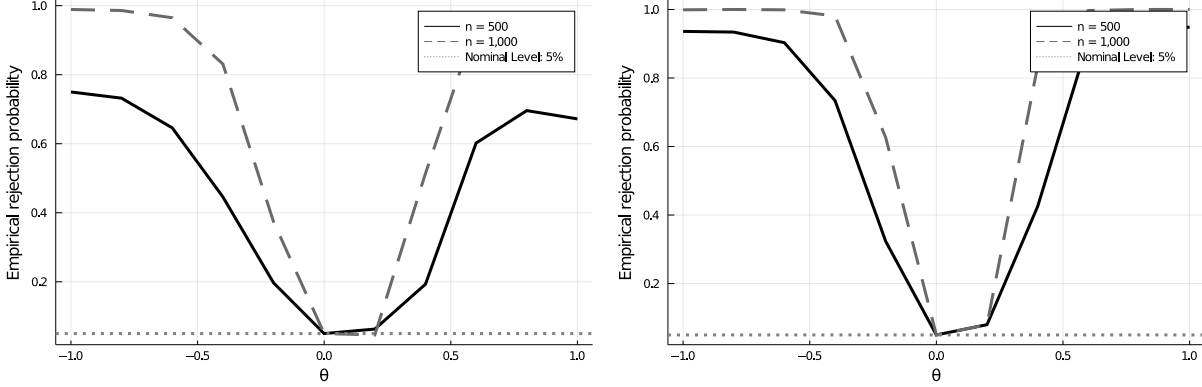


(c) Sample size:  $n = 1,000$



(d) Sample size:  $n = 1,000$

Figure 1: Finite-sample distribution of the test. The figure shows density estimates for the Kolmogorov-Smirnov and Cramér-von Mises statistics under  $H_0$ ,  $\theta = 0$  (solid line), and two alternative hypotheses:  $\theta = 0.4$  (dashed line) and  $\theta = 1$  (dotted line).



(a) Kolmogorov-Smirnov test

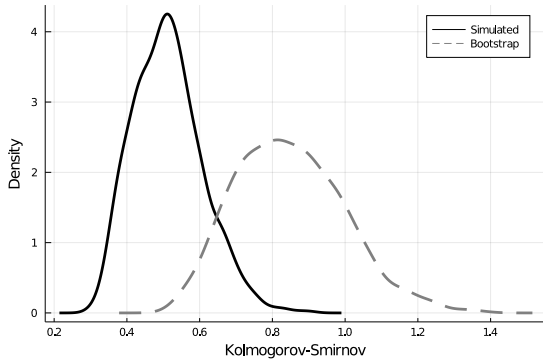
(b) Cramér-von Mises test

Figure 2: Power curves. The figure shows empirical rejection probabilities as a function of degree of separability  $\theta$  for samples of size  $n = 500$  (solid line) and  $n = 1,000$  (dashed line). The value  $\theta = 0$  corresponds to the separable model, while  $\theta \neq 0$  are deviations from separability. The nominal level of the test is set at 5%.

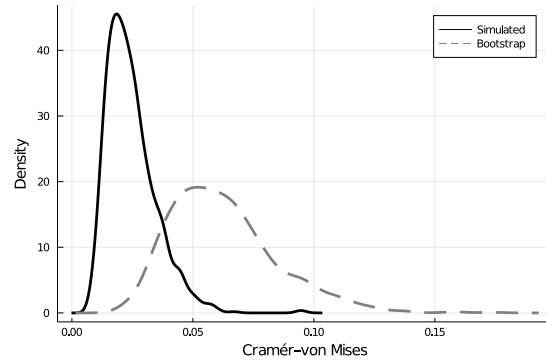
## 5 Are Engel curves separable?

Engel curves are fundamental for the analysis of consumers' behavior and have implications for the aggregate economic outcomes. The 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 \(2012\)](#). The nonparametric IV approach to the estimation of Engel curves is pioneered in the seminal paper [Blundell, Chen, and Kristensen \(2007\)](#) who focus on the estimation of Engel curves in the UK.

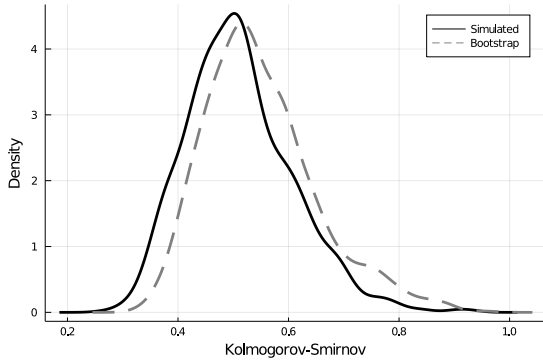
We draw a dataset from the 2015 US Consumer Expenditure Survey; see [Babii \(2020\)](#) for the estimated Engel curves with the uniform confidence bands using this dataset. We restrict our attention to married couples with a positive income during the last 12 months, yielding 10,055 observations. The dependent variable is a share of expenditures on a particular commodity while the endogenous regressor is a natural logarithm of the total expenditures. We instrument the expenditures using the gross income. In particular, [Blundell,](#)



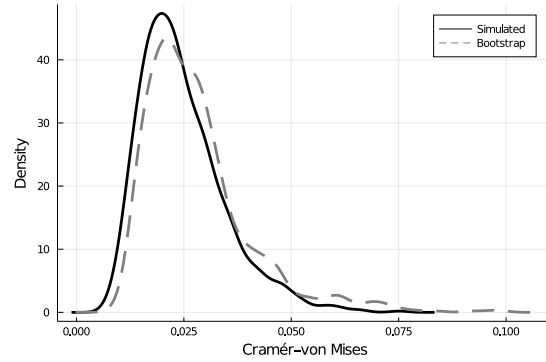
(a) Naive bootstrap:  $m_n = n$



(b) Naive bootstrap:  $m_n = n$



(c) Adaptive choice of  $m_n$



(d) Adaptive choice of  $m_n$

Figure 3: Naive bootstrap vs.  $m$  out of  $n$  bootstrap. The figure shows density estimates of simulated and bootstrapped Kolmogorov-Smirnov and Cramér-von Mises statistics under  $H_0$ . We compare the naive ( $n$  out of  $n$ ) bootstrap and the  $m$  out of  $n$  bootstrap with the adaptive choice of  $m$  for the sample of size 1,000 observations.

Table 1: Testing separability of Engel curves. The table shows  $m$  out of  $n$  bootstrap p-values of Kolmogorov-Smirnov and Cramér-von Mises tests for 13 commodities.

Commodity	KS	CvM	Commodity	KS	CvM
Food home	0.00	0.00	Gas and oil	0.00	0.00
Food away	0.00	0.00	Personal care	0.00	0.00
Clothing	0.00	0.00	Health	0.00	0.00
Tobacco	0.00	0.00	Insurance	0.00	0.00
Alcohol	0.00	0.00	Reading	0.00	0.53
Trips	0.00	0.00	Transportation	0.01	0.03
Entertainment	0.08	0.00			

Chen, and Kristensen (2007) point out that the gross income will be exogenous for consumption expenditures assuming that heterogeneity in earnings is not related to unobserved preferences over consumption; see also Chen and Christensen (2018) and Babii (2020).

In Table 1, we report the  $m$  out of  $n$  bootstrap p-values, with the adaptive choice of  $m$ ; see Section 4 for more details on the practical implementation of tests. We report results for both the Kolmogorov-Smirnov (KS) and the Cramér-von Mises (CvM) tests. Remarkably, the 5% level tests reject the separability for all commodities with the exception for the Entertainment (KS) and Reading (CvM). Moreover, the 1% level tests reject separability in all cases, except for Reading (CvM) and Transportation (CvM). This suggests that Engel curves for these commodities may exhibit substantial heterogeneities in unobservables.

## 6 Conclusions

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 separability test. The test is based on the estimation of a separable nonparametric IV regression and the verification of the independence restriction imposed by the nonseparable IV model. To obtain a large sample approximation to the distribution of our test

statistics, we develop a novel uniform asymptotic expansions of the empirical distribution function of nonparametric IV residuals and obtain new results for the Tikhonov regularization in Sobolev spaces. 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 tests, which are highly nonstandard. In our Monte Carlo experiments, we find that the bootstrap fails in approximating the distribution of the 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 1% level test rejects the separability of Engel curves for most of the commodities. This indicates that the Engel curves may be heterogeneous in unobservables and that the nonseparable modeling of Engel curves may be useful, see, e.g., [Blundell, Horowitz, and Parey \(2017\)](#) for the estimation of nonseparable demand functions.

The paper offers several other 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, the demand function, or the wage equation. Second, given the plethora of 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 [Pardo-Fernández, Van Keilegom, and González-Manteiga \(2007\)](#) and [Escanciano, Pardo-Fernández, and Van Keilegom \(2018\)](#).

## Acknowledgement

This work was supported by the French National Research Agency under Grant ANR-19-CE40-0013-01/ExtremReg project. 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.

# APPENDIX: ADDITIONAL RESULTS AND PROOFS

**Notation.** For two sequences  $(a_n)_{n \in \mathbf{N}}$  and  $(b_n)_{n \in \mathbf{N}}$ , we denote  $a_n \lesssim b_n$  if  $a_n = O(b_n)$  and  $a_n \sim b_n$  if both  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$ . For two sequences of random variables  $(X_n)_{n \in \mathbf{N}}$  and  $(Y_n)_{n \in \mathbf{N}}$ , we denote  $X_n \lesssim_P Y_n$  for  $X_n = O_P(Y_n)$ . For a bounded linear operator  $T : \mathcal{X} \rightarrow \mathcal{Y}$  on normed spaces, we use  $\|T\|_{\text{op}} = \inf\{c \geq 0 : \|Tx\| \leq c\|x\|, \forall x \in \mathcal{X}\}$  to denote its operator norm, where with some abuse of notation, we use  $\|\cdot\|$  to denote the norm of both spaces.

## A.1 Tikhonov regularization in Sobolev spaces

This section discusses convergence rates for the Tikhonov-regularized estimator in Sobolev spaces. The following result extends Carrasco, Florens, and Renault (2014), Proposition 3.1 to the case of the unknown operator.

**Theorem A.1.** *Suppose that Assumption 3.1 is satisfied,  $\|\hat{T} - T\|_{\text{op}}^2 \lesssim_P \alpha_n$ , and  $2s \geq b - a$ . Then for every  $c \in [0, s]$*

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

It is worth emphasizing that this result is not specific to the nonparametric IV regression and can be applied to a generic ill-posed inverse problem  $T\varphi = r$ , where  $(T, r)$  is estimated with  $(\hat{T}, \hat{r})$ . Moreover, in the case of nonparametric IV regression, it can be easily applied to nonparametric/machine learning estimators  $(\hat{T}, \hat{r})$  other than the kernel smoothing. Next, we specialize the generic result of Theorem A.1 to the nonparametric IV regression with  $(T, r)$  estimated via kernel smoothing, see equation (3).

**Corollary A.1.1.** *Suppose that Assumptions 3.1 and 3.2 are satisfied,  $\frac{1}{nh_n^{p+q}} \forall h_n^{2t} = O(\alpha_n)$ , and  $2s \geq b - a$ . Then for every  $c \in [0, s]$*

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

## A.2 Distribution of nonparametric IV residuals

In this section, we present results on the weak convergence of the empirical distribution of nonparametric IV residuals. These results are used to obtain the large sample approximation to the distribution of independence tests and are of independent interest.

**Theorem A.1.** *Suppose that Assumptions 3.1, 3.2, and 3.3 (i), and 3.4 are satisfied. Then*

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

uniformly over  $u \in \mathbf{R}$ .

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

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

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

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

By Lemma B.1.1 in the Supplementary Material,

$$T_{1n}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i [T(T^*T)^{-1} f_{UZ}(u, \cdot)](W_i) + o_P(1),$$

while under Assumptions 3.3 (i) and 3.4

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

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

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

□

As a consequence of Theorem A.1, we obtain the following Donsker-type central limit theorem for the empirical distribution of nonparametric IV residuals.

**Corollary A.2.1.** *Suppose that assumptions of Theorem A.1 are satisfied. Then*

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

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

$$(u, u') \mapsto F_U(u \wedge u') - F_U(u)F_U(u') + \mathbb{E} [U^2 [T(T^*T)^{-1}f_{UZ}(u, \cdot)](W) [T(T^*T)^{-1}f_{UZ}(u', \cdot)](W)] \\ + \mathbb{E} [\mathbb{1}_{\{U \leq u\}}U [T(T^*T)^{-1}f_{UZ}(u', \cdot)](W) + \mathbb{1}_{\{U \leq u'\}}U [T(T^*T)^{-1}f_{UZ}(u, \cdot)](W)].$$

*Proof.* The process given in Theorem A.1 is an empirical process indexed by the following class of functions  $\mathcal{F} = \{(v, w) \mapsto \mathbb{1}_{\{v \leq u\}} + v (T(T^*T)^{-1}f_{UZ}(u, \cdot))(w), u \in \mathbf{R}\}$ , which is a sum of the Donsker class and  $\mathcal{H} = \{(v, w) \mapsto v (T(T^*T)^{-1}f_{UZ}(u, \cdot))(w), u \in \mathbf{R}\}$ . By van der Vaart and Wellner (1996), Example 2.10.5, it enough to show that  $\mathcal{H}$  is Donsker. The former statement follows from the fact that under Assumption 3.1 in the Supplementary Material by Engl, Hanke, and Neubauer (2000), since for  $\kappa - a > q/2$

$$\sup_{u \in \mathbf{R}} \|T(T^*T)^{-1}f_{UZ}(u, \cdot)\|_{\kappa-a} \lesssim \sup_{u \in \mathbf{R}} \|f_{UZ}(u, \cdot)\|_\kappa \leq M < \infty,$$

where the last inequality follows under Assumption 3.3 (i). Therefore,  $\mathcal{H} \subset \{(v, w) \mapsto vg(w) : g \in H_M^{\kappa-a}\}$ , where  $H_M^{\kappa-a}$  is a Sobolev ball of radius  $M$ . Since  $\kappa > a + q/2$ , this shows that the class  $\mathcal{H}$  is Donsker; see Nickl and Pötscher (2007), Corollaries 4 and 5. The covariance function simplifies since  $\mathbb{E}[U|W] = 0$ .  $\square$

## A.3 Proofs of main results

In this section we provide proofs of main results of the paper.

*Proof of Proposition 2.1.* Since  $T : L_2(\mathbf{R}^p) \rightarrow L_2(\mathbf{R}^q)$  is injective, the nonparametric IV regression  $\varphi \in L_2(\mathbf{R}^p)$  is unique. Therefore,  $U = Y - \varphi(Z)$  is a well-defined unique random variable. If the model in equation (1) admits a separable representation, then since  $\varepsilon \perp\!\!\!\perp W$

$$\begin{aligned} \mathbb{E}[Y|W] &= \mathbb{E}[\psi(Z) + g(\varepsilon)|W] \\ &= \mathbb{E}[\psi(Z) + \mathbb{E}g(\varepsilon)|W]. \end{aligned}$$



Therefore,  $\varphi(Z) = \psi(Z) + \mathbb{E}g(\varepsilon)$  by the injectivity of  $T$ , and whence  $U = g(\varepsilon) - \mathbb{E}g(\varepsilon)$ . This shows that  $U \perp\!\!\!\perp W$  because  $\varepsilon \perp\!\!\!\perp W$ .  $\square$

*Proof of Theorem 3.1.* By Lemma A.4.2, 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

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

The first term is a classical independence empirical process

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

where the second line follows by the maximal inequality.

Next, under Assumption 3.3 (i), by Taylor's theorem, for some  $\tau \in [0, 1]$

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

Under Assumptions 3.3 by Corollary A.1.1

$$\|S_{2n}\|_\infty \leq \sup_{w, u, z} \left| \int^w \partial_u f_{UZW}(u, z, \tilde{w}) d\tilde{w} \right| \sqrt{n} \|\hat{\varphi} - \varphi\|^2 = o_P(1)$$

Similarly, we have uniformly in  $(u, w)$

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

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

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

where the last line follows by the same argument as in the proof of Theorem A.1 under Assumption 3.3 (i).  $\square$

*Proof of Proposition 3.1.*  $\mathbb{H}_n$  is an empirical process indexed by the class of functions

$$\mathcal{F} = \{(v, w) \mapsto \mathbb{1}_{\{v \leq \tilde{v}, w \leq \tilde{w}\}} - \mathbb{1}_{\{v \leq \tilde{v}\}} F_W(\tilde{w}) - \mathbb{1}_{\{w \leq \tilde{w}\}} F_U(\tilde{v}) + F_{UW}(\tilde{v}, \tilde{w}) + \delta_{\tilde{v}, \tilde{w}}(v, w) : (\tilde{v}, \tilde{w}) \in \mathbf{R}^{1+q}\}.$$

By van der Vaart and Wellner (1996), Example 2.10.7 it suffices to show that each of the functions in the sum constitutes a Donsker class. To that end, recall first that the indicator functions are classical examples of Donsker classes. Therefore, all terms in  $\mathcal{F}$ , but the last one, are either Donsker or can be factored as Donsker classes and a deterministic bounded function not depending on the argument of the indicator function. Lastly, under Assumptions 3.1 (i) by Engl, Hanke, and Neubauer (2000), Corollary 8.22

$$\|T(T^*T)^{-1}g(v, w, \cdot)\|_{\kappa-a} \lesssim \sup_{(v,w) \in \mathbf{R}^{1+q}} \|g(v, w, \cdot)\|_{\kappa} \leq M < \infty,$$

where the latter follows under Assumption 3.3 (ii). Therefore, we obtain that  $\{(v, w) \mapsto v(T(T^*T)^{-1}g(\tilde{v}, \tilde{w}, \cdot))(w) : \tilde{v} \in \mathbf{R}, \tilde{w} \in \mathbf{R}^q\} \subset \{(v, w) \mapsto vg(w) : g \in H_M^{\kappa-a}\}$ , where  $H_M^{\kappa-a}$  is a Sobolev ball of radius  $M$ . Since  $\kappa > a + q/2$ , this shows that  $\mathcal{F}$  is Donsker; see Nickl and Pötscher (2007), Corollaries 4 and 5.  $\square$

*Proof of Corollary 3.1.* Since under  $H_0$ ,  $\mathbb{G}_n \rightsquigarrow \mathbb{H}$  by Proposition 3.1, the asymptotic distribution of  $T_{\infty, n}$  under  $H_0$  is readily obtained by the continuous mapping theorem; see van der Vaart and Wellner (1996), Theorem 1.3.6. For the Cramér-von Mises statistics, write

$$T_{2, n} = \iint \mathbb{H}^2(u, w) dF_{UW}(u, w) + R_{1n} + R_{2n}$$

with

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

By Proposition 3.1, under  $H_0$ ,  $\mathbb{G}_n \rightsquigarrow \mathbb{H}$  and  $\sqrt{n}(\hat{F}_{\hat{U}W}(u, w) - F_{UW}(u, w))$  also converges weakly by Proposition 3.1 and Theorem A.2.1, whence by the Skorokhod construction

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

The first expression in Eq. A.1 implies that  $R_{1n} \xrightarrow{\text{a.s.}} 0$ . Since  $\mathbb{H}$  has a.s. bounded and continuous trajectories, the second expression in Eq. A.1 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 fixed alternative hypothesis, since  $\mathbb{E}[U|W] = 0$ , by Theorem 3.1, the Glivenko-Cantelli theorem, and a similar argument we obtain

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

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

$$\mathbb{E}[h_{u,w}(U, W)] = 2(F_{UW}(u, w) - F_U(u)F_W(w)) = 2n^{-1/2}H(u, w).$$

Therefore, by Corollary 3.1 and continuous mapping theorem

$$\begin{aligned} T_{\infty,n} &= \sup_{u,w} |\mathbb{G}_n(u, w)| \\ &= \sup_{u,w} |\mathbb{G}_n(u, w) - \sqrt{n}\mathbb{E}[h_{u,w}(U, W)] + 2H(u, w)| \\ &\rightsquigarrow \sup_{u,w} |\mathbb{H}(u, w) + 2H(u, w)|. \end{aligned}$$

For the Cramér-von Mises statistics, write

$$T_{2,n} = \iint |\mathbb{H}(u, w) + 2H(u, w)|^2 dF_{UW}(u, w) + S_{1n} + S_{2n},$$

where

$$S_{1n} = \iint \left\{ \left| \mathbb{G}_n(u, w) - \sqrt{n} \mathbb{E}[h_{u,w}(U, W)] + 2H(u, w) \right|^2 - \left| \mathbb{H}(u, w) + 2H(u, w) \right|^2 \right\} d\hat{F}_{\hat{U}W}(u, w)$$

$$S_{2n} = \iint \left| \mathbb{H}(u, w) + 2H(u, w) \right|^2 d \left[ \hat{F}_{\hat{U}W}(u, w) - F_{UW}(u, w) \right].$$

Therefore, the result follows by Proposition 3.1 and the same argument as under  $H_0$  with the only difference that now we have the bias  $2H$  in the limiting distribution.  $\square$

## A.4 Auxiliary technical results

In this section, we provide several auxiliary technical results.

**Lemma A.4.1.** *Suppose that Assumption 3.1, 3.2, 3.3, and 3.4. Then*

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

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

*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. 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  $f_{u,\Delta}(U, Z) = \mathbb{1}_{(-\infty, u+\Delta(Z)]}(U)$ ,  $\mathcal{G}_1 = \{f_{u,\Delta} : u \in \mathbf{R}, \Delta \in H_M^c(\mathbf{R}^p)\}$ ,  $\mathcal{G}_2 = \{f_{u,0} : u \in \mathbf{R}\}$ , and  $\mathcal{G} = \mathcal{G}_1 - \mathcal{G}_2$ . 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 (1996), 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 the bracketing number of  $H_M^c$  satisfies  $\log N_{[\cdot]}(\varepsilon, H_M^c, \|\cdot\|_{L_Z^2}) \lesssim \varepsilon^{-p/c}$ , where  $(L_Z^2, \|\cdot\|_{L_Z^2})$  denotes the space of functions, square-integrable with respect to  $f_Z$ . Put  $M_\varepsilon = N_{[\cdot]}(\varepsilon, H_M^c, \|\cdot\|_{L_Z^2})$  and fix  $u \in \mathbf{R}$ . Let  $[\underline{\Delta}_j, \overline{\Delta}_j]_{j=1}^{M_\varepsilon}$  be a collection of  $\varepsilon$ -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  $\|\overline{\Delta}_j - \underline{\Delta}_j\|_{L_Z^2} \leq \varepsilon$ , and whence  $\mathbb{1}_{(-\infty, u+\underline{\Delta}_j]} \leq \mathbb{1}_{(-\infty, u+\Delta]} \leq \mathbb{1}_{(-\infty, u+\overline{\Delta}_j]}$ .

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

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

Denote the largest  $\underline{u}_{j,k}$  such that  $\underline{u}_{j,k} \leq u$  by  $\underline{u}_j^*$  and the smallest  $\bar{u}_{j,k}$  such that  $u \leq \bar{u}_{j,k}$  by  $\bar{u}_j^*$ . Consider the following family of brackets  $\left[ \mathbb{1}_{(-\infty, \underline{u}_j^* + \underline{\Delta}_j]}, \mathbb{1}_{(-\infty, \bar{u}_j^* + \bar{\Delta}_j]} \right]_{j=1}^{M_\varepsilon}$ . Under Assumption 3.2 (ii)

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

Therefore, we constructed brackets of size  $O(\varepsilon)$ , covering  $\mathcal{G}_1$ , and we have used at most  $O(\varepsilon^{-2} M_\varepsilon)$  such brackets. Since  $c > p/2$ , we have  $\int_0^1 \sqrt{\log N_{[]}(\varepsilon, \mathcal{G}, \|\cdot\|_{L_Z^2})} d\varepsilon < \infty$ . This shows that the empirical process  $\sqrt{n}(P_n - P)g, g \in \mathcal{G}$  is Donsker, hence, asymptotically equicontinuous; see [van der Vaart and Wellner \(1996\)](#), Theorem 1.5.7. Then for any  $\varepsilon > 0$

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

where  $\Pr^*$  denotes the outer probability measure.

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

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

where the third line follows by the Cauchy-Schwartz inequality and Corollary A.1.1 under Assumptions 3.1, 3.2, and 3.4. Similarly,

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

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

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

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  $\limsup_{n \rightarrow \infty} \Pr^*(\sqrt{n}\|\hat{\nu}_n\|_\infty > \varepsilon) = 0$ , which concludes the proof.  $\square$

**Lemma A.4.2.** *Suppose that Assumptions 3.1, 3.2, 3.3, and 3.4 are satisfied. Then uniformly over  $(u, w) \in \mathbf{R} \times \mathbf{R}^q$*

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

and

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

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

*Proof.* Note that the first expression and the expression in the statement of Lemma A.4.1 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 A.2.1 and the classical Donsker central limit theorem. By Lemma A.4.1, we obtain the first statement since  $F_W$  is uniformly bounded by one.

The proof of the second statement is similar to the proof of Lemma A.4.1 and is omitted.  $\square$

## B.1 Additional proofs and auxiliary results

This section contains proofs of several results from the main part of the paper as well as several auxiliary result.

*Proof of Theorem A.1.* 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

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

where the second line follows by [Engl, Hanke, and Neubauer \(2000\)](#), Corollary 8.22 with  $\nu = (s - c)/(a + s) \leq 1$ ; the third line by the definition of operator norm; the fourth line by the isometry of functional calculus; and the last since  $\sup_{\lambda} |\lambda^d/(\alpha_n + \lambda)| \lesssim \alpha_n^{d-1}$  for all  $d \in [0, 1]$ .

Similarly, since for bounded linear operators  $A$  and  $B$ ,  $\|AB\|_{\text{op}} \leq \|A\|_{\text{op}}\|B\|_{\text{op}}$ ,

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

Next, since  $L^s \varphi \in H^{b-s}$  and  $s \geq (b-a)/2$ , by [Engl, Hanke, and Neubauer \(2000\)](#), Corollary 8.22, there exists  $\psi \in L_2$  such that  $L^s \varphi = (T_s^* T_s)^{\frac{b-s}{2(a+s)}} \psi$ . Therefore,

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

Next, decompose

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

with

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



and

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

Similarly, decompose

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

with  $S_{1n}$  and  $S_{2n}$  defined below. In particular,

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

where the last two lines follow by [Engl, Hanke, and Neubauer \(2000\)](#), Corollary 8.22 with  $\nu = s/(a+s) \leq 1$  and previous computations. Similarly,

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

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

*Proof of Corollary A.1.1.* By the Cauchy-Schwartz inequality

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

where the second line follows from the well-known risk bound; see, e.g., [Giné and Nickl \(2015\)](#), p. 403-404 under Assumption 3.2. 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

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

under Assumption 3.2 can be found in [Babii and Florens \(2020\)](#).  $\square$

**Lemma B.1.1.** *Suppose that Assumptions 3.1, 3.2, 3.3, and 3.4, and are satisfied. Then*

$$\langle \hat{\varphi} - \varphi, f_{UZ}(u, \cdot) \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i [T(T^*T)^{-1} f_{UZ}(u, \cdot)](W_i) + o_P(1)$$

*Proof.* Similarly to the proof of Theorem A.1, decompose

$$\sqrt{n} \langle \hat{\varphi} - \varphi, f_{UZ}(u, \cdot) \rangle \triangleq I_n(u) + II_n(u) + III_n(u)$$

with

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

We show below that  $\|II_n + III_n + IV_n + V_n\|_\infty = o_P(1)$ . To that end, first since  $T_s = TL^{-s}$

$$\begin{aligned}\|II_n\|_\infty &= \sqrt{n} \sup_u \left\langle (\alpha_n I + T^* T)^{-1} (\hat{T}^* - T^*) (\hat{r} - \hat{T}\varphi), f_{UZ}(u, \cdot) \right\rangle \\ &\leq \sqrt{n} \left\| (\hat{T}^* - T^*) (\hat{r} - \hat{T}\varphi) \right\| \sup_u \left\| (\alpha_n I + T^* T)^{-1} f_{UZ}(u, \cdot) \right\| \\ &\lesssim \sqrt{n} \|\hat{T}^* - T^*\|_{\text{op}} \left\| \hat{r} - \hat{T}\varphi \right\| \left\| (\alpha_n I + T^* T)^{-1} T^* T \right\|_{\text{op}} \\ &\lesssim_P \sqrt{n} \left( \frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t \right) \left( \frac{1}{\sqrt{nh_n^q}} + h_n^t \right) = o_P(1),\end{aligned}$$

where the third line follows under Assumptions 3.1 (i) and 3.3 (i); and the fourth by arguments as in the proof of Corollary A.1.1 under Assumptions 3.2 and 3.4 (ii).

Second,

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

where the first equality follows since  $L$  is self-adjoint; the second line by the Cauchy-Schwartz inequality since  $\sup_u \|L^a f_{UZ}(u, \cdot)\| < \infty$  under Assumption 3.3 (i) and by Assumption 3.1 (i); the third since  $L^s \varphi = (T_s^* T_s)^{\frac{b-s}{2(a+s)}} \psi$  for some  $\psi \in L_2$  by Engl, Hanke, and Neubauer (2000), Corollary 8.22; the fourth by the isometry of the functional calculus; and the last since  $2s = b - a$  and since  $n\alpha_n^2 \rightarrow 0$  under Assumption 3.4 (iii).

Next, decompose  $III_n(u) = R_{1n}(u) + R_{2n}(u)$  with

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

By the Cauchy-Schwartz inequality and previous computations

$$\begin{aligned}
\|R_{1n}\|_\infty &\leq \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \left\| \hat{T}(\alpha_n I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \right\|_{\text{op}} \|\hat{T} - T\|_{\text{op}} \sup_u \|(\alpha_n I + T^* T)^{-1} f_{UZ}(u, \cdot)\| \\
&\lesssim \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \|\hat{T} - T\|_{\text{op}}
\end{aligned}$$

and

$$\begin{aligned}
\|R_{2n}\|_\infty &\leq \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \left\| \hat{T}(\alpha_n I + \hat{T}^* \hat{T})^{-1} \right\|_{\text{op}} \|\hat{T}^* - T^*\|_{\text{op}} \sup_u \|T(\alpha_n I + T^* T)^{-1} f_{UZ}(u, \cdot)\| \\
&\lesssim \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \alpha_n^{-1/2} \|\hat{T} - T\|_{\text{op}} \|T(\alpha_n I + T^* T)^{-1} (T^* T)^{\kappa/2a}\|_{\text{op}} \\
&\lesssim \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \|\hat{T} - T\|_{\text{op}} \alpha_n^{\kappa/2a-1}.
\end{aligned}$$

Therefore, under Assumption 3.4 since  $\kappa > 2a$

$$\begin{aligned} \|III_n\|_\infty &\lesssim \sqrt{n} \left\| \hat{r} - \hat{T}\varphi \right\| \alpha_n^{-1/2} \|\hat{T} - T\|_{\text{op}} \\ &\quad \sqrt{n} \left( \frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t \right) \left( \frac{1}{\sqrt{nh_n^q}} + h_n^t \right) = o_P(1). \end{aligned}$$

Similarly, decompose  $IV_n(u) = S_{1n}(u) + S_{2n}(u)$  with

$$\begin{aligned} S_{1n}(u) &= \sqrt{n} \left\langle L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{T}_s - T_s) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi, f_{UZ}(u, \cdot) \right\rangle, \\ S_{2n}(u) &= \sqrt{n} \left\langle L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} (\hat{T}_s^* - T_s^*) T_s \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi, f_{UZ}(u, \cdot) \right\rangle. \end{aligned}$$

Likewise, by the Cauchy-Schwartz inequality and previous computations

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

and

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

where we use  $\|\hat{T} - T\|_{\text{op}} \lesssim_P \alpha_n^{1/2}$  and  $2s = b - a$ ; see also the proof of Theorem A.1.

Therefore,  $\|IV_n\|_\infty = o_P(1)$ .

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

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

Next, note that

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

with  $[\varphi * K_z](z) \triangleq \int \varphi(v) h_n^{-p} K_z(h_n^{-1}(z - v)) dv$ , whence

$$T^*(\hat{r} - \hat{T}\varphi) = \frac{1}{n} \sum_{i=1}^n (Y_i - [\varphi * K_z](Z_i)) [f_{ZW} * K_w](\cdot, W_i)$$

with  $[f_{ZW} * K_w](z, w) \triangleq \int f_{ZW}(z, v) h_n^{-q} K_w(h_n^{-1}(w - v)) dv$ . Using this observation, decompose equation (B.1) further

$$I_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \left\langle f_{ZW}(\cdot, W_i), (T^* T)^{-1} f_{UZ}(u, \cdot) \right\rangle + Q_{1n} + Q_{2n} + Q_{3n} + o_P(1)$$

with

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

By the Cauchy-Schwartz inequality

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

where the second line follows by triangle inequality and Assumption 3.3 (i); the third by Assumption 3.2 (i), Cauchy-Schwartz inequality, and since  $f_Z$  and  $f_W$  are uniformly bounded under Assumption 3.2 (ii); and the last by the standard bias computations under Assumptions 3.1 (ii) and 3.2, and Young's inequality under Assumption 3.2 (ii) and (iv).

Similarly, by the Cauchy-Schwartz inequality and Assumption 3.1 (i)

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

where the second line follows under the i.i.d. assumption; the third since  $\mathbb{E}[U|W] \leq C$  under Assumption 3.2 (i); the fourth since  $f_W$  is uniformly bounded under Assumption 3.2 (ii); and the last by the standard bias computations under Assumptions 3.1 (ii) and 3.2 (iv).

Lastly, by the Cauchy-Schwartz inequality

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

where the second inequality follows under Assumptions 3.2 (i); the third line under Assumptions 3.1, 3.2 (i)-(ii), and 3.3 (i); and the last by the isometry of functional calculus.

Combining these estimates under Assumptions 3.3 (i) and 3.4, we obtain the result

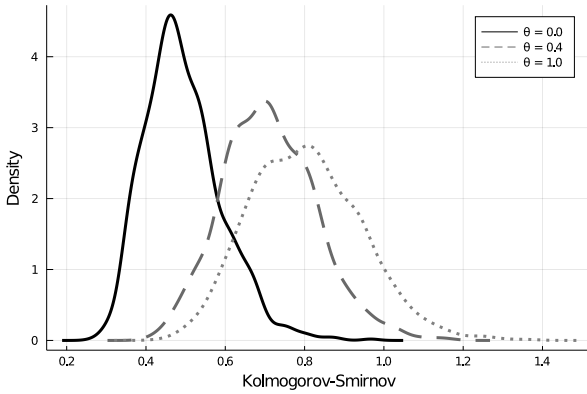
$$\begin{aligned}
II_n(u) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \langle f_{ZW}(\cdot, W_i), (T^*T)^{-1} f_{UZ}(u, \cdot) \rangle + O_P(\sqrt{n}h_n^b + h_n^t + \alpha_n^{(\kappa/2a-1)\wedge 1}) + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i [T(T^*T)^{-1} f_{UZ}(u, \cdot)](W_i) + o_P(1).
\end{aligned}$$

□

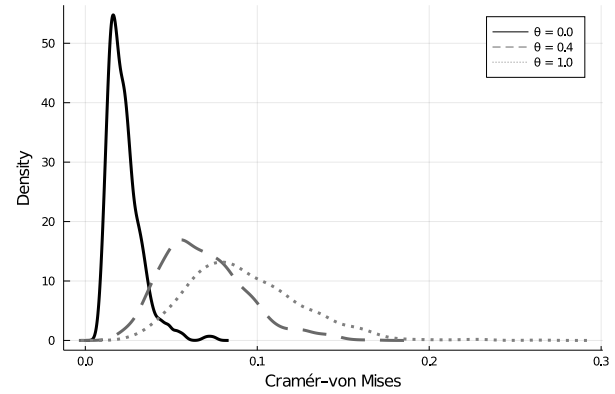
## B.2 Additional Monte Carlo experiments

In this section, we report results of additional Monte Carlo experiments when the structural function is  $\varphi(x) = \exp(-x^2/4)$ . The rest of the data-generating process is the same as in the main part of the paper.

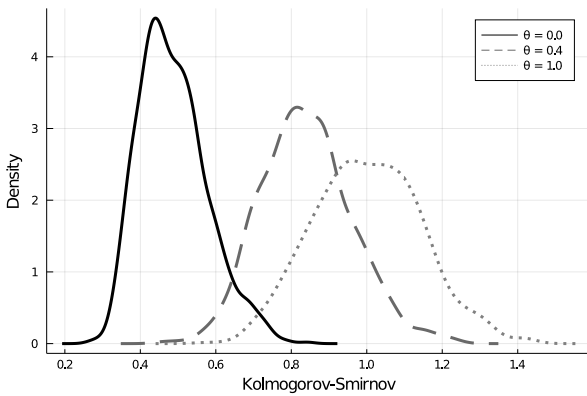
Figure B.1 shows the distribution of the test statistics under the null hypothesis and the two alternative hypotheses for different sample sizes. The two distributions are sufficiently distinct once the alternative hypothesis becomes more separated from the null hypothesis. We plot in Figure B.2 the power curves when the level of the test is fixed at 5%. The power of the test increases once alternative hypotheses become more distant from the null hypothesis and when the sample size is larger. The Cramér-von Mises test seems to have a higher power for the class of considered alternatives. Overall, the findings are largely similar to the findings of experiments presented in the main part of the paper.



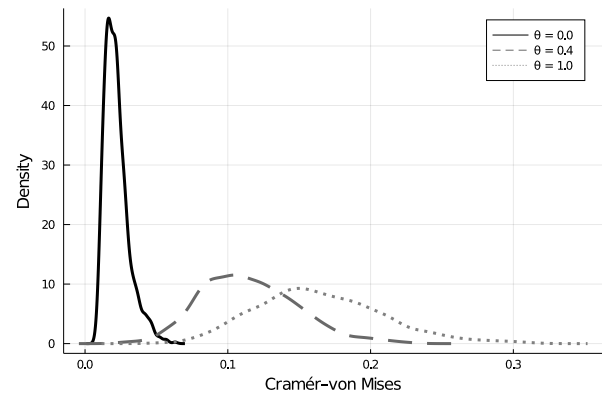
(a) Sample size:  $n = 500$



(b) Sample size:  $n = 500$

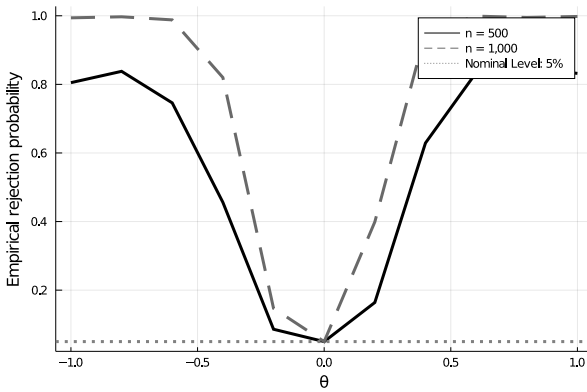


(c) Sample size:  $n = 1,000$

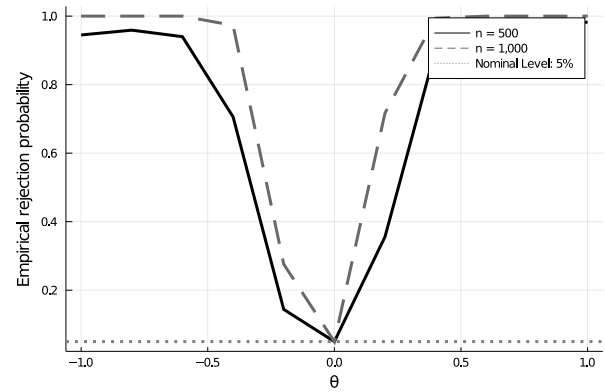


(d) Sample size:  $n = 1,000$

Figure B.1: Finite-sample distribution of the test – density estimates of the distribution of Kolmogorov-Smirnov and Cramér-von Mises statistics under  $H_0$ ,  $\theta = 0$  (solid line), and two alternative hypotheses,  $\theta = 0.4$  (dashed line) and  $\theta = 1$  (dotted line).



(a) Kolmogorov-Smirnov test



(b) Cramér-von Mises test

Figure B.2: Power curves. The figure shows empirical rejection probabilities as a function of degree of separability  $\theta$  for samples of size  $n = 500$  (solid line) and  $n = 1,000$  (dashed line). The value  $\theta = 0$  corresponds to the separable model, while  $\theta \neq 0$  are deviations from separability. The nominal level of the test is set at 5%.



## References

- AKRITAS, M., AND I. VAN KEILEGOM (2001): “Non-parametric estimation of the residual distribution,” *Scandinavian Journal of Statistics*, 28(3), 549–567.
- ALMÁS, I. (2012): “International income inequality: measuring ppp bias by estimating Engel curves for food,” *American Economic Review*, 102(2), 1093–1117.
- ANDREWS, D. (1994): “Chapter 37: Empirical process methods in econometrics,” *Handbook of Econometrics*, 4, 2247–2294.
- BABII, A. (2020): “Honest confidence sets in nonparametric IV regression and other ill-posed models,” *Econometric Theory*, 36(4), 658–706.
- BABII, A. (2021): “High-dimensional mixed-frequency IV regression,” *arXiv preprint arXiv:2003.13478*.
- BABII, A., AND J.-P. FLORENS (2020): “Is completeness necessary? Estimation in non-identified linear models,” *arXiv preprint arXiv:1709.03473*.
- BABII, A., AND R. KUMAR (2021): “Isotonic regression discontinuity designs,” *Journal of Econometrics (forthcoming)*.
- BANKS, J., R. BLUNDELL, AND A. LEWBEL (1997): “Quadratic engel curves and consumer demand,” *Review of Economics and Statistics*, 79(4), 527–538.
- BICKEL, P., 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.
- BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): “Semi-nonparametric IV estimation of shape-invariant Engel curves,” *Econometrica*, 75(6), 1613–1669.
- BLUNDELL, R., J. HOROWITZ, AND M. PAREY (2017): “Nonparametric estimation of a nonseparable demand function under the slusky inequality restriction,” *Review of Economics and Statistics*, 99(2), 291–304.

- BREUNIG, C. (2020): “Specification testing in nonparametric instrumental quantile regression,” *Econometric Theory*, 36(4), 583–625.
- 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): “Chapter 77: 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,” *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*.
- CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. NEWEY (2014): “Local identification of nonparametric and semiparametric models,” *Econometrica*, 82(2), 785–809.
- CHEN, X., AND T. CHRISTENSEN (2018): “Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression,” *Quantitative Economics*, 9(1), 39–84.
- 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., J. C. ESCANCIANO, H. ICHIMURA, W. K. NEWEY, AND J. M. ROBINS (2016): “Locally robust semiparametric estimation,” *arXiv preprint arXiv:1608.00033*.
- CHERNOZHUKOV, V., I. FERNNDEZ-VAL, W. NEWEY, S. STOULI, AND F. VELLA (2020): “Semiparametric estimation of structural functions in nonseparable triangular models,” *Quantitative Economics*, 11(2), 503–533.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “AN IV model of quantile treatment effects,” *Econometrica*, 73(1), 245–261.

- (2013): “Quantile models with endogeneity,” *Annual Review of Economics*, 5, 57–81.
- CHERNOZHUKOV, V., G. IMBENS, AND W. 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.
- DEVROYE, L. (1986): *Non-uniform random variate generation*. Springer-Verlag New York.
- D’HAULTFOUILLE, 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., 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 (2000): *Regularization of inverse problems*. Springer Netherlands.
- ESCANCIANO, J., J. 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 and Economic Statistics*, 36(2), 334–345.

- FLORENS, J.-P. (2003): “Inverse problems and structural econometrics: the example of instrumental variables,” *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume II*, pp. 284–311.
- FLORENS, J.-P., 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(3), 472–496.
- GAGLIARDINI, P., AND O. SCAILLET (2012): “Tikhonov regularization for nonparametric instrumental variable estimators,” *Journal of Econometrics*, 167(1), 61–75.
- GAGLIARDINI, P., AND O. SCAILLET (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 (2015): *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press.
- HALL, P., AND J. HOROWITZ (2005): “Nonparametric methods for inference in the presence of instrumental variables,” *Annals of Statistics*, 33(6), 2904–2929.
- HECKMAN, 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. YANG (2016): “Testing for monotonicity in unobservables under unconfoundedness,” *Journal of Econometrics*, 193(1), 183–202.
- HODERLEIN, S., AND A. VANHEMS (2018): “Estimating the distribution of welfare effects using quantiles,” *Journal of Applied Econometrics*, 33(1), 52–72.
- HOROWITZ, J., AND S. LEE (2007): “Nonparametric instrumental variables estimation of a quantile regression model,” *Econometrica*, 75(4), 1191–1208.

- IMBENS, G. (2010): “Nonadditive models with endogenous regressors,” *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume III*, pp. 17–46.
- IMBENS, G., AND W. NEWEY (2009): “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica*, 77(5), 1481–1512.
- KREIN, S. G., AND Y. I. PETUNIN (1966): “Scales of Banach spaces,” *Russian Mathematical Surveys*, 21(2), 85–159.
- LEWBEL, A. (2001): “Demand systems with and without errors,” 91(3), 611–618.
- 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.
- MATZKIN, R. (2013): “Nonparametric identification in structural economic models,” *Annual Review of Economics*, 5, 457–486.
- 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., AND J. POWELL (2003): “Instrumental variable estimation of nonparametric models,” *Econometrica*, 71(5), 1565–1578.
- NICKL, R., AND B. 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-FERNNDEZ, J., I. VAN KEILEGOM, AND W. GONZLEZ-MANTEIGA (2007): “Testing for the equality of  $k$  regression curves,” *Statistica Sinica*, 17(3), 1115–1137.
- POLITIS, D., J. 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.
- TORGOVITSKY, A. (2015): “Identification of nonseparable models using instruments with small support,” *Econometrica*, 83(3), 1185–1197.
- (2017): “Minimum distance from independence estimation of nonseparable instrumental variables models,” *Journal of Econometrics*, 199(1), 35–48.
- VAN DER VAART, A., AND J. WELLNER (1996): *Weak convergence and empirical processes: with applications to statistics*. Springer-Verlag New York.