Econometric Theory, **0**, 2024, 1–33.
doi:10.1017/S026646662400001X

ARE UNOBSERVABLES SEPARABLE?

Andrii Babii
University of North Carolina at Chapel Hill

Jean-pierre Florens
Toulouse School of Economics

It is common to assume in empirical research that observables and unobservables are additively separable, especially when the former are endogenous. This is 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 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 U.S. Consumer Expenditure Survey, we find that the test rejects the separability in Engel curves for some commodities.

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 because it is widely recognized that identification and estimation challenges arise when interactions between the two are allowed for. However, economic theory and intuition often lead to nonseparable models. Prominent examples are demand functions with price or income effects heterogeneous in unobserved preferences; production functions with elasticities heterogeneous in unobserved input choices; labor supply functions with heterogeneous wage effects; wage equations with returns to schooling heterogeneous in unobserved ability.[1]

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 et al. (2008), Imbens and Newey (2009), Torgovitsky (2015), and

This work was supported by the French National Research Agency under Grant ANR-19-CE40-0013-01/ExtremReg project. We thank the seminar participants at Yale University, and the editorial team for helpful comments. We also thank Ivan Canay, Xiaohong Chen, Tim Christensen, Elia Lapenta, Pascal Lavergne, Thierry Magnac, Nour Meddahi, Ingrid Van Keilegom, and Edward Vytlacil for insightful discussions. All remaining errors are ours. Address correspondence to Andrii Babii, Department of Economics, University of North Carolina at Chapel Hill—Gardner Hall, CB 3305, Chapel Hill, NC 27599-3305, USA; e-mail: babii.andrii@gmail.com.

[1] For treatment effect models, the applied researcher often has a nonseparable model in mind; see Imbens and Angrist (1994) or Heckman and Vytlacil (2001).

D'Haultfœuille and Février (2015), among others. It is well understood that a 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 et al. (2014).

Despite the significant efforts focused on understanding the identification and estimation of nonseparable IV models and the widespread use of separable IV models, 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 two notable exceptions that focus on separability tests under the *conditional independence restriction*. The conditional independence restriction is different from the *mean-independence restriction* imposed by the separable nonparametric IV model that we are interested in here. Other recently developed specification tests include the monotonicity test of Hoderlein et al. (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. The test is based on the independence condition of a nonseparable IV model. We build on an insight that the structural function in the separable IV 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), Darolles et al. (2011), and Chen and Christensen (2018), among others. If the separable model is correct, then the nonparametric IV residuals should approximate the unobserved error that should be independent of instrumental variables. This intuition suggests that it should be possible to detect the separability with the classical Kolmogorov–Smirnov (KS) or Cramér–von Mises (CvM) test. 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 nonparametric IV residuals satisfies the Donsker property.

Formalizing this intuition is far from trivial because 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. Moreover, the empirical distribution function of nonparametric IV residuals is a *non-smooth* function. The uniform central limit theorem for the empirical distribution of regression residuals in the parametric linear case is a classical problem in statistics (e.g., Durbin, 1973; Loynes, 1980; Mammen, 1996). The nonparametric extension is more challenging, and it is remarkable that the empirical distribution of nonparametric regression residuals also satisfies the uniform central limit theorem (see Akritas and Van Keilegom, 2001). The additively separable nonparametric IV regression differs from the problems discussed above in two important ways. First, its asymptotic properties depend on both the smoothness of the regression function and the smoothing properties of the conditional expectation operator. Second, the regression error is not independent of endogenous regressors.

In this paper, we show that the empirical distribution function of nonparametric IV residuals satisfies the uniform central limit theorem. To the best of our knowledge, this is the first result on the distribution of the nonparametric IV residuals. The result can be used to develop various residual-based specification tests and is of independent interest. Building on this result, we obtain a large sample approximation to the distributions of our test statistics. The distribution is non-standard, and the critical values can be estimated with resampling techniques.

Our results are based on an insight that the Tikhonov regularization in Sobolev spaces, considered in Florens, Johannes, and Van Bellegem (2011), Gagliardini and Scaillet (2012, 2017), and Carrasco et al. (2014), among others, provides a natural link between the modern empirical process theory and the theory of ill-posed inverse problems. In contrast to this literature, we obtain new results for the Tikhonov regularization with a Sobolev penalty.

The paper is organized as follows. In Section 2, we introduce the problem and discuss the main testable implication. In Section 3, we characterize a large sample approximation to the distribution of the residual-based KS and CvM independence tests. In Section 4, we report on a Monte Carlo study 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. Conclusions appear in Section 6. All technical details, auxiliary results, and proofs are collected in the Appendix.

2. SEPARABILITY OF UNOBSERVABLES

Let (Y,Z,W) be observed random variables satisfying the nonseparable model

$$Y=\Phi(Z,\varepsilon), \qquad \varepsilon \perp\!\!\!\perp W, \tag{1}$$

where $Y\in\mathbf{R}$ is an outcome, $Z\in\mathbf{R}^p$ are regressors, $\varepsilon\in\mathbf{R}$ is an unobservable, $W\in\mathbf{R}^q$ is a vector of instrumental variables, and $\Phi:\mathbf{R}^p\times\mathbf{R}\to\mathbf{R}$ is a structural function. We assume that W are valid instrumental variables satisfying the exclusion restriction, $\varepsilon \perp\!\!\!\perp W$, and the relevance condition, $W \not\!\perp\!\!\!\perp Z$. Note that the independence restriction is a commonly used identifying condition for nonseparable models; see Chernozhukov et al. (2020), Blundell, Horowitz, and Parey (2017), Torgovitsky (2017), Torgovitsky (2015), D'Haultfœuille and Février (2015), Dunker et al. (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 a review of earlier econometric literature on the identification of nonseparable models.

The independence condition $\varepsilon \perp\!\!\!\perp W$ does not rule out heteroskedasticity in the distribution of Y conditionally on Z or W, which is often observed in empirical practice. It does not rule out heteroskedasticity in the distribution of unobservables ε conditionally on covariates Z. However, it rules out heteroskedasticity of unobservables conditionally on the instrumental variable.

Testing separability can be done in several different ways. For instance, one could fit the nonseparable model and check whether the estimated function is separable. This approach corresponds to the principle behind the Wald test for parametric models. Alternatively, one could estimate the separable model and check the independence restriction. 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 exist measurable functions $\psi : \mathbf{R}^p \to \mathbf{R}$ and $g : \mathbf{R} \to \mathbf{R}$ such that

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

If the model has such representation, then ψ can be estimated consistently using the nonparametric IV approach; see Darolles et al. (2011), Blundell et al. (2007), Horowitz and Lee (2007), and Newey and Powell (2003). The nonparametric IV regression function $\varphi : \mathbf{R}^p \to \mathbf{R}$ solves the functional equation

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

where $T : L_2(\mathbf{R}^p) \to L_2(\mathbf{R}^q)$ 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 φ solving the functional equation (2).[2] The following result provides a convenient testable implication of separability, provided that U is unambiguously defined (see the 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* $U \perp\!\!\!\perp W$.

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,ε), 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; Babii and Florens, 2020), and does not require that the nonseparable model is identified (e.g., Chernozhukov and Hansen, 2005; Chen et al., 2014).

3. INDEPENDENCE TEST

In this section, we introduce tests of independence (see Proposition 2.1). Formally, we focus on testing

$$H_0 : U \perp\!\!\!\perp W \qquad \text{vs.} \qquad H_1 : U \not\perp\!\!\!\perp W.$$

The null hypothesis H_0 is testable, provided that the nuisance parameter φ in $U = Y - \varphi(Z)$ is replaced by a consistent estimator.

[2] Here and later, we assume that r belongs to the range of the operator T, which is a relatively mild restriction on the regression function.

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

3.1. Tikhonov Regularization in Sobolev Spaces

We focus on the Tikhonov regularization in Hilbert scales to estimate the nuisance parameter φ; see Carrasco et al. (2014), Gagliardini and Scaillet (2012), and Florens et al. (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 under sufficient smoothness assumptions and restrictions on the degree of ill-posedness. This makes it appropriate for our test and more generally for semiparametric applications (see Corollary A.1.1 in the Appendix).

We first recall the definition of Hilbert scales. A family of Hilbert spaces $(H^s, \langle .,. \rangle_s)_{s\in\mathbf{R}}$ is called a Hilbert scale if $H^t \subset H^s$ for every $t > s$ and the inclusion is a continuous embedding, i.e., $\|\phi\|_s \leq c\|\phi\|_t$ for every $\phi \in H^t$. Let $(L_2(\mathbf{R}^p), \|.\|)$ be a space of square-integrable functions with respect to the Lebesgue measure, and let $|.|$ be the euclidean norm on $\mathbf{R}^p$. Let also $\hat{f}(\xi) = (2\pi)^{-p/2}\int_{\mathbb{R}^p} e^{-iz^\top \xi} f(z)\mathrm{d}z$ be the Fourier transform of f. Then the Sobolev space[3]

$$H^s(\mathbf{R}^p) = \left\{ f \in L_2(\mathbf{R}^p) : \int_{\mathbf{R}^p} (1+|\xi|^2)^s |\hat{f}(\xi)|^2 d\xi < \infty \right\}$$

is a Hilbert space with inner product

$$\langle f, g\rangle_s = \int_{\mathbf{R}^p} (1+|\xi|^2)^s \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi$$

and the induced norm $\|.\|_s = \sqrt{\langle .,. \rangle_s}$. It is known that the inclusion $H^t(\mathbf{R}^p) \subset H^s(\mathbf{R}^p)$ is a continuous embedding for $t > s$, so that $(H^s(\mathbf{R}^p))_{s\in\mathbf{R}}$ is a Hilbert scale. Note that the Hilbert scale $(H^s(\mathbf{R}^p))_{s\in\mathbf{R}}$ is essentially generated by powers of the operator $L^s = F^{-1}(1+|.|^2)^{s/2}F$ with the inner product $\langle f, g\rangle_s = \langle L^s f, L^s g\rangle$, where $(Ff)(\xi) = \hat{f}(\xi)$ denotes the Fourier transform and F^{-1} its inverse; see Carrasco et al. (2014), Nair (2015), and Krein and Petunin (1966) for more details and examples.

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

$$\hat{r}(w) = \frac{1}{nh_n^q}\sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right), \qquad (\hat{T}\phi)(w) = \int \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z,$$

$$\hat{f}_{ZW}(z,w) = \frac{1}{nh_n^{p+q}}\sum_{i=1}^n K_z\left(h_n^{-1}(Z_i - z)\right) K_w\left(h_n^{-1}(W_i - w)\right), \tag{3}$$

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

[3]The definition of Sobolev spaces via the Fourier transform is equivalent to the one based on weak derivatives (see Evans, 2010, Sect. 5.8.5 for more details). The Sobolev space is usually defined for $s \geq 0$ and can be extended to $s \in \mathbf{R}$ using a concept of Gelfand triple (see Nair, 2015 for more details).

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

We estimate φ using the Tikhonov-regularized estimator penalized by a Sobolev norm

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

for some $s \geq 0$. 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 residual, and let

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

be the empirical distribution functions. We focus on the following residual-based independence empirical process

$$\mathbb{G}_n(u,w) = \sqrt{n}\left(\hat{F}_{\hat{U}W}(u,w) - \hat{F}_{\hat{U}}(u)\hat{F}_W(w)\right). \tag{5}$$

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, Chap. 3.8). In particular, the estimation of φ 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 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 the features of φ, the harder it is to recover T from equation (2). Condition (ii) describes the smoothness of the structural function φ and is standard in the nonparametric literature; see Chen (2007) or Giné and Nickl (2015).

Assumption 3.2. (i) $(Y_i, Z_i, W_i)_{i=1}^n$ are i.i.d. observations of (Y, Z, W) with $\mathbb{E}\|W\| < \infty$, $\mathbb{E}\|Z\| < \infty$, $\mathbb{E}\left[\varphi^2(Z)|W\right] \leq C$, and $\mathbb{E}\left[Y^2|W\right] \leq C$ for some $C < \infty$; (ii) the distribution of (Y, Z, W) is absolutely continuous with respect to the Lebesgue measure with densities $f_Z, f_W, f_{ZW}, f_{Y|Z} \in L_\infty$ 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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

continuous kernel $K \in L_2(\mathbf{R}) \cap L_\infty(\mathbf{R})$ of bounded variation with $\int K(u)\mathrm{d}u = 1$, $\int |u|^l |K(u)|\mathrm{d}u < \infty$, and $\int u^k K(u)\mathrm{d}u = 0$ for $k \in \{1, \ldots, l\}$ and $l \geq t$.

Assumption 3.2 is standard for kernel estimators; see also Darolles et al. (2011, Appendix B) for a discussion of generalized boundary kernels that can be used when supports are bounded. We could also allow for discrete regressors provided that the instrumental variables are continuous following the approach of Das (2005).

To introduce the next assumption, let ∂_u be a partial derivative with respect to the variable u, let $\|.\|_\infty$ 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 $\sup_u \|f_{UZ}(u, .)\|_\kappa < \infty$ with $\kappa > 2a \vee (a + q/2)$; (ii) $\left\| \int_{\{v \leq .\}} \partial_u f_{UZW}(., ., v)\mathrm{d}v \right\|_\infty < \infty$ and $\sup_{u,w} \left\| \int_{-\infty}^{w} f_{UZW}(u, ., v)\mathrm{d}v \right\|_\kappa < \infty$ with $\kappa > 2a \vee (a + q/2)$.

Note that since $f_{UZW}(u, z, w) = f_{YZW}(u + \varphi(z), z, w)$, Assumption 3.3 imposes equivalently several mild smoothness restrictions on f_{YZW} and φ.

Assumption 3.4. $h_n \to 0$ and $\alpha_n \to 0$ as $n \to \infty$ are such that (i) $nh_n^q \alpha_n^{2(a+c)/(a+b)} \to \infty$, $nh_n^{p+q}\alpha_n \to \infty$, and $h_n^{2t}/\alpha_n \to 0$; (ii) $\sqrt{n}\alpha_n^{2b/(a+b)} \to 0$, $\sqrt{n}h_n^q \alpha^{2a/(a+b)} \to \infty$, and $\sqrt{n}h_n^{2t}/\alpha_n^{2a/(a+b)} \to 0$; (iii) $n\alpha_n^2 \to 0$, $nh_n^{2(b \wedge 2t)} \to 0$, and $nh_n^{p+2q} \to \infty$; 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 3.4(i) provides a set of sufficient conditions for the Sobolev norm consistency, $\|\hat{\varphi} - \varphi\|_c = o_P(1)$ with $c > p/2$. Assumption 3.4(ii) ensures that $\|\hat{\varphi} - \varphi\| = o_P(n^{-1/4})$, which is a standard rate requirement in the semiparametric literature.[4] Our proof strategy relies on the consistency in the Sobolev norm and *does not* require uniform consistency (cf. Akritas and Van Keilegom, 2001). Lastly, condition (iii) ensures that a certain uniform asymptotic expansion holds.[5]

The following result describes a large sample approximation to the residual-based independence empirical process.

THEOREM 3.1. *Suppose that Assumptions 3.1–3.4 are satisfied. Then*

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

[4]This condition may be restrictive for the severely ill-posed inverse problems if the structural function is not sufficiently smooth.

[5]It is easy to see that when $p = q = c = 1$, there exist $c_1, c_2 \in (0, 1)$ such that Assumption 3.4 holds for $h_n \sim n^{-c_1}$ and $\alpha_n \sim n^{-c_2}$ provided that t and b are sufficiently large.

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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

$$\delta_{u,w}(U_i, W_i) = U_i \left(T(T^*T)^{-1}\rho(u,.,w)\right)(W_i),$$
$$\rho(u,z,w) = \int_{-\infty}^{w} f_{UZW}(u,z,\tilde{w})\mathrm{d}\tilde{w} - f_{UZ}(u,z)F_W(w).$$

Theorem 3.1 does not require $U \perp\!\!\!\perp W$. The proof of this result can be found in the Appendix and relies on 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 implies that certain terms associated with residuals are asymptotically negligible. The estimation of the nuisance component φ has a first-order asymptotic effect through 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 et al. (2018) and Chernozhukov et al. (2022) for recent contributions, Andrews (1994) for earlier treatment, and Babii (2022, Sect. 3.3) for a related discussion in the setting of ill-posed inverse problems.

Theorem 3.1 can be readily used to construct the residual-based CvM and KS statistics:

$$T_{2,n} = \iint |\mathbb{G}_n(u,w)|^2 \mathrm{d}\hat{F}_{\hat{U}W}(u,w) \qquad \text{and} \qquad T_{\infty,n} = \sup_{u,w} |\mathbb{G}_n(u,w)|. \tag{6}$$

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 $h_{u,w}(U,W) = \mathbb{1}_{\{U \le u, W \le w\}} - \mathbb{1}_{\{U \le u\}} F_W(w) - \mathbb{1}_{\{W \le w\}} F_U(u) + F_{UW}(u,w) + \delta_{u,w}(U,W)$. The following Donsker-type central limit theorem holds:

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

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

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

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

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

$$(u,w,u',w') \mapsto \mathbb{E}\Big[\big(\mathbb{1}_{\{U \le u, W \le w\}} - \mathbb{1}_{\{U \le u\}} F_W(w) - \mathbb{1}_{\{W \le w\}} F_U(u) + F_{UW}(u,w) + \delta_{u,w}(U,W)\big) \times$$

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

$$\times \Big(\mathbb{1}_{\{U\le u', W\le w'\}} - \mathbb{1}_{\{U\le u'\}}F_W(w') - \mathbb{1}_{\{W\le w'\}}F_U(u') + F_{UW}(u',w') + \delta_{u',w'}(U,W)\Big)\Big].$$

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

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

Consider also a sequence of local alternative hypotheses

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

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, $F_{UW}(u,w) = F_U(u)F_W(w) + aF_U(u)F_W(w)(1 - F_U(u))(1 - F_W(w))$ with $a \in [-1,1]$ (see Devroye, 1986, Chap. XI, Thm. 3.2).

The following corollary describes the behavior of our 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*,*

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

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* $H_{1,n}$*,*

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

Corollary 3.1 shows that the residual-based independence tests can detect parametric local alternatives. The asymptotic distributions under H_0 are not pivotal, in contrast to 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, we focus on resampling methods to compute the critical values.

We conclude this section with instructions on how to implement the test in practice:

1. Compute the nonparametric IV regression $\hat{\varphi}$ based on the kernel estimators in equation (3). One could use, e.g., the product of sixth-order Epanechnikov kernels for which the Silverman's rule of thumb bandwidth choices based on sample standard deviations are $h_z = 3.53\hat{\sigma}_z n^{-1/13}$ and $h_w = 3.53\hat{\sigma}_w n^{-1/13}$.
2. Compute the nonparametric IV residuals, $\hat{U}_i = Y_i - \varphi(Z_i)$ and the residual-based independence empirical process $\mathbb{G}_n$ in equations (4) and (5).

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

3. Compute either the KS or the CvM statistics, denoted $T_{2,n}$ and $T_{\infty,n}$ (see equation (6)). The integrals and suprema are evaluated on a discrete grid of 100 points.
4. Compute the bootstrap critical values as $1-\alpha$ empirical quantile of bootstrapped statistics $T^*_{2,n}$ or $T^*_{\infty,n}$, denoted $q^*_{1-\alpha}$, where the process $\mathbb{G}$ is replaced by its bootstrap counterpart $\mathbb{G}^*$ (see equation (7)). The bootstrap counterparts are obtained drawing, e.g., 5,000 samples of size n with replacements from $(Y_i, W_i, Z_i)_{i=1}^n$.
5. Reject separability if the observed statistics exceeds the critical value $q^*_{1-\alpha}$.

4. MONTE CARLO EXPERIMENTS

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

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

We set $\varphi(x) = \cos(x)$ and consider samples of size $n = 500$ and $n = 1,000$ observations (see Appendix A.5 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. Under H_1, the nonparametric IV regression does not estimate consistently the nonseparable structural function $(z,e) \mapsto \cos(z) + \theta ze + e$, which depends on unobservables. The nonparametric IV regression estimates instead a 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.

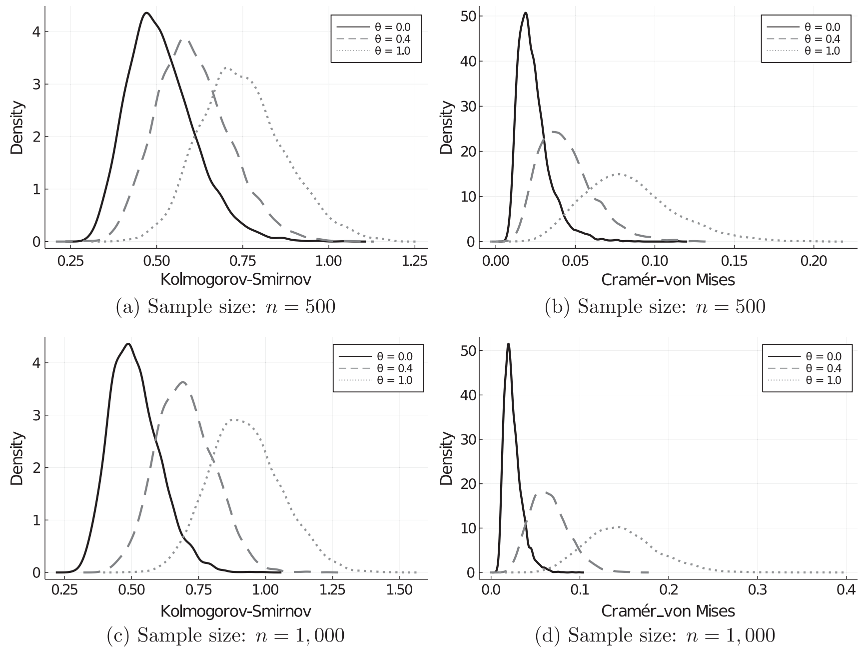
We set the number of Monte Carlo replications and the number of bootstrap replications to 1,000 through all our experiments. We also discretize all continuous quantities on the grid of 100 equidistant points in $[-4,4]$. The estimates $\hat{r}$ and $\hat{T}$ in equation (3) are obtained using the sixth-order 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 KS and CvM statistics, computed, respectively, as

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

where $\mathbb{G}_n(u,w) = \sqrt{n}(\hat{F}_{\hat{U}W}(u,w) - \hat{F}_{\hat{U}}(u)\hat{F}_W(w))$. Lastly, we compute the critical values using the nonparametric bootstrap, replacing the empirical process $\mathbb{G}_n$ by the bootstrapped process

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press



(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 distributions. The figure shows density estimates for the Kolmogorov–Smirnov and Cramér–von Mises statistics under H_0, $\theta = 0$ (solid line), and the two alternative hypotheses: $\theta = 0.4$ (dashed line) and $\theta = 1$ (dotted line).

$$\mathbb{G}^*_n(u,w) = \sqrt{n}\left((\hat{F}^*_{\hat{U}W}(u,w) - \hat{F}^*_{\hat{U}}(u)\hat{F}^*_W(w) - (\hat{F}_{\hat{U}W}(u,w) - \hat{F}_{\hat{U}}(u)\hat{F}_W(w))\right), \tag{7}$$

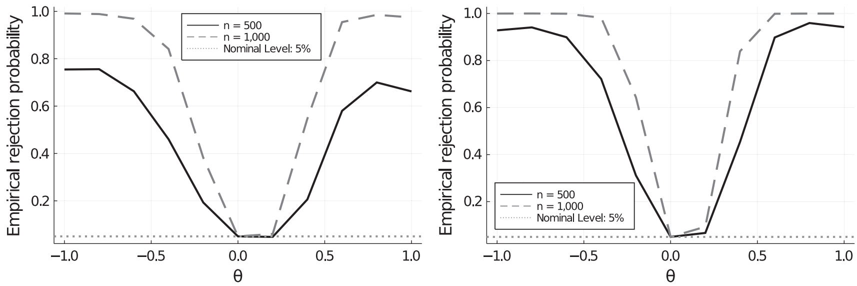
where $\hat{F}^*_{\hat{U},W}$, $\hat{F}^*_{\hat{U}}$, and $\hat{F}^*_W$ are computed via resampling $(Y_i, Z_i, W_i)^n_{i=1}$.[6]

Figure 1 shows the distribution of the test statistics under the null and the two alternative hypotheses for different sample sizes. The distributions under H_0 and H_1 are sufficiently distinct when there is a larger separation as measured by θ.

In Figure 2, we plot the power curves when the level is fixed at 5%. The power increases once alternative hypotheses become more distant from the null hypothesis. The Cramér–von Mises test seems to have a higher power for the class of considered alternatives. The figure also indicates that the tests are consistent.

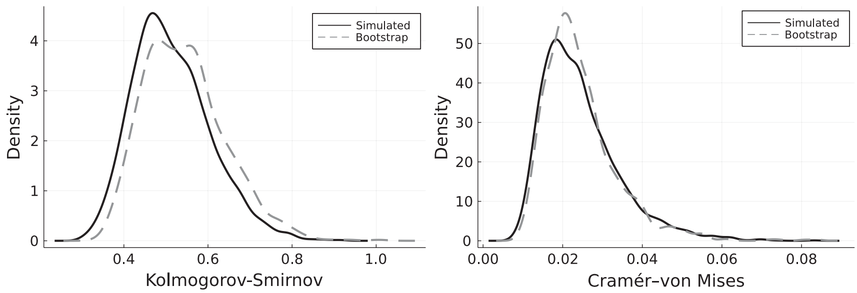
[6]The validity of the bootstrap could be justified in light of Neumeyer (2009) and Neumeyer and Van Keilegom (2019) (see also Chen, Linton, and Van Keilegom, 2003). For the empirical distribution function (EDF) of regression residuals, the arguments reduce to the following decomposition: $\sqrt{n}(\hat{F}^*_{\hat{U}} - \hat{F}_{\hat{U}}) = \sqrt{n}(\tilde{F}^*_{\hat{U}} - \tilde{F}_{\hat{U}}) + \sqrt{n}(\hat{F}^*_{\hat{U}} - \tilde{F}^*_{\hat{U}}) + \sqrt{n}(\tilde{F}_{\hat{U}} - \hat{F}_{\hat{U}})$, where $\tilde{F}^*_{\hat{U}}$ is the EDF of a sample bootstrapped from a smoothed EDF, denoted $\tilde{F}_{\hat{U}}$. Neumeyer (2009) shows that the first term estimates consistently the asymptotic distribution, while Neumeyer and Van Keilegom (2019) show that the last two terms are asymptotically negligible.

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press



(a) Kolmogorov-Smirnov test (b) Cramér-von Mises test

FIGURE 2. Power curves. The figure shows the empirical rejection probabilities as a function of the separability parameter θ 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 is set at 5%.



(a) Kolmogorov-Smirnov (b) Cramér-von Mises

FIGURE 3. Bootstrapped and simulated distributions. The figure shows density estimates of simulated and bootstrapped Kolmogorov–Smirnov and Cramér–von Mises statistics under H_0. Based on 1,000 bootstrap replications.

In Figure 3, we explore the bootstrap performance. We plot the exact finite-sample distribution of both test statistics and the distribution of bootstrapped statistics under H_0. The bootstrap seems to mimic the finite-sample distribution relatively well.

5. ARE ENGEL CURVES SEPARABLE?

Engel curves are fundamental for the analysis of consumers' behavior and have implications for aggregate economic outcomes. The Engel curve describes the relationship between demand for a particular commodity and a household's budget. Interesting applications of Engel curve estimations include measurement of welfare losses associated with tax distortions in Banks, Blundell, and Lewbel (1997), estimation of growth and inflation in Nakamura, Steinsson, and Liu (2016), or estimation of income inequality across countries in Almås (2012).

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

Table 1. Testing separability of Engel curves. The table shows the 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	0.35	0.33
Food away	0.11	0.02	Personal care	0.45	0.34
Clothing	0.20	0.17	Health	0.38	0.31
Tobacco	0.14	0.01	Insurance	0.38	0.20
Alcohol	0.36	0.36	Reading	0.39	0.41
Trips	0.20	0.02	Transportation	0.00	0.00
Entertainment	0.22	0.00			

The nonparametric IV approach to Engel curves is pioneered in the seminal paper of Blundell et al. (2007), who focus on the estimation of Engel curves for the United Kingdom.

We draw a dataset from the 2015 U.S. 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 et al. (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.

In Table 1, we report the bootstrap p-values for the KS and CvM tests for 13 different commodities. Our test provides some evidence that the Engel curves for Food, Transportation, and possibly Tobacco may be non-separable, and hence heterogeneous in unobservables.

6. CONCLUSIONS

This paper offers a new perspective on the separability of unobservables in econometric models with endogeneity. Starting from a 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 expansion 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 satisfy the Donsker central limit theorem. In contrast to nonparametric regression without endogeneity, we find that parameter uncertainty affects the asymptotic distribution of the residual-based independence tests, which are highly nonstandard.

Using the 2015 U.S. Consumer Expenditure Survey data, we find some evidence for non-separable Engel curves for some commodities. This indicates that some Engel curves may be heterogeneous in unobservables and that the nonseparable modeling of Engel curves may be useful (see, e.g., Blundell et al., 2017 for the estimation of nonseparable demand functions).

The paper offers several directions for future research. First, it might be interesting to test the separability of unobservables in other structural relations that are commonly estimated using the additively separable models in empirical practice, such as a production function, a labor supply function, a demand function, or a 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; Escanciano et al., 2018).

APPENDIX

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 $T : \mathcal{X} \to \mathcal{Y}$ 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 $\|.\|$ to denote the norm of both spaces.

A.1. Tikhonov Regularization in Hilbert Scales

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

THEOREM A.1.1. *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* $c \in [0, s]$,

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

Proof. Decompose

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

with

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

$$IV_n = L^{-s}(\alpha_n I + \hat{T}_s^*\hat{T}_s)^{-1}\hat{T}_s^*\hat{T}_s L^s\varphi - L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_sL^s\varphi,$$
$$V_n = L^{-s}(\alpha_n I + T_s^*T_s)^{-1}T_s^*T_sL^s\varphi - \varphi.$$

For the first term,

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

where the second line follows by Engl, Hanke, and Neubauer (2000, Cor. 8.22) with $\nu = (s-c)/(a+s) \le 1$; the third line by the definition of operator norm; the fourth line by the isometry of functional calculus (see Cavalier, 2011, Thm. 1.3); 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, $\|AB\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}}\|B\|_{\mathrm{op}}$,

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

Next, since $L^s\varphi \in H^{b-s}$ and $s \ge (b-a)/2$, by Engl et al. (2000, Cor. 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,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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

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

and

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

Similarly, decompose

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

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

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

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

where the last two lines follow by Engl et al. (2000, Cor. 8.22) with $\nu = s/(a+s) \leq 1$ and previous computations. Similarly,

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

The result follows from combining all estimates. □

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

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

Proof. By the Cauchy–Schwartz inequality,

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

where the second line follows from the well-known risk bound (see, e.g., Giné and Nickl, 2015, pp. 403–404 under Assumption 3.2). Therefore, by Theorem A.1.1,

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

Lastly, by Babii and Florens (2020, Prop. A.1.1),

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

under Assumption 3.2. □

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. For instance, they may ensure that we can estimate the structural quantile function $\Phi(z,e)=\varphi(z)+F_U^{-1}(e)$ in the separable IV model

$$Y=\varphi(Z)+F_U^{-1}(\varepsilon),$$

where $\varepsilon\sim U(0,1)$ without loss of generality (see Chernozhukov et al., 2020).

THEOREM A.2.1. *Suppose that Assumptions 3.1, 3.2, and 3.3(i), and 3.4 are satisfied. Then*

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

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

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

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

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

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

By Lemma A.4.3 in Appendix A.4,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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 $u \in \mathbf{R}$

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

□

As a consequence of Theorem A.2.1, 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.2.1 are satisfied. Then*

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

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

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

Proof. The process given in Theorem A.2.1 is an empirical process indexed by the following class of functions $\mathcal{F} = \left\{(v,w) \mapsto \mathbb{1}_{\{v\leq u\}} + v\left(T(T^*T)^{-1}f_{UZ}(u,.)\right)(w),\ u \in \mathbf{R}\right\}$, which is a sum of a Donsker class and $\mathcal{H} = \left\{(v,w) \mapsto v\left(T(T^*T)^{-1}f_{UZ}(u,.)\right)(w),\ u \in \mathbf{R}\right\}$. By van der Vaart and Wellner (1996, Exam. 2.10.5), it is enough to show that $\mathcal{H}$ is Donsker. The former statement follows from the fact that under Assumption 3.1 by Engl et al. (2000), since for $\kappa - a > q/2$,

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

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, Cors. 4 and 5)). The covariance simplifies since $\mathbb{E}[U|W] = 0$. □

A.3. Proofs of the Main Results

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

Proof of Proposition 2.1. Since φ is unique by assumption, $U = Y - \varphi(Z)$ is a well-defined unique random variable. If the model in equation (1) admits a separable representation, then since $\varepsilon \perp\!\!\!\perp W$,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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

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

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

The first term is a classical independence empirical process

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

where the second line follows by the maximal inequality.

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

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

Under Assumptions 3.3 by Corollary A.1.1,

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

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

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

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

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

where the last line follows by the same argument as in the proof of Theorem A.2.1 under Assumption 3.3(i). □

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

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

By van der Vaart and Wellner (1996, Exam. 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 et al. (2000, Cor. 8.22),

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

where the latter follows under Assumption 3.3(ii). Therefore, we obtain that $\{(v,w)\mapsto v(T(T^*T)^{-1}g(\tilde{v},\tilde{w},.))(w):\ \tilde{v}\in\mathbf{R},\tilde{w}\in\mathbf{R}^q\}\subset\{(v,w)\mapsto vg(w):\ g\in H_M^{\kappa-a}\}$, 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, Cors. 4 and 5)). □

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, Thm. 1.3.6)). For the CvM statistics, write

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

with

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

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; thus, by the Skorokhod construction,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

The first expression in equation (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 equation (A.1) in conjunction with the Helly–Bray theorem shows that $R_{2n} \xrightarrow{\text{a.s.}} 0$. Therefore, the asymptotic distribution of the CvM 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

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

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

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

Therefore, by Corollary 3.1 and the 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 CvM statistics, write

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

where

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

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. □

A.4. Additional Technical Results and Proofs

In this section, we provide several auxiliary technical results.

LEMMA A.4.1. *Suppose that Assumptions 3.1–3.4 are satisfied. Then*

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

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

Proof. The main idea of the proof is to embed the process inside the supremum into an empirical process indexed by u and a Sobolev ball containing $\hat{\Delta}$ with a probability tending

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

to one. We first show that the process is Donsker; thus, the supremum in equation (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 by van der Vaart and Wellner (1996, Thm. 2.10.6). To this end, we check that the bracketing entropy condition is satisfied for $\mathcal{G}_1$.

By Nickl and Pötscher (2007, Cor. 4), the bracketing number of H_M^c satisfies $\log N_{[\,]}(\varepsilon, H_M^c, \|.\|_{L_Z^2}) \lesssim \varepsilon^{-p/c}$, where $(L_Z^2, \|.\|_{L_Z^2})$ denotes the space of functions, square-integrable with respect to f_Z. Put $M_\varepsilon = N_{[\,]}(\varepsilon, H_M^c, \|.\|_{L_Z^2})$ and fix $u \in \mathbf{R}$. Let $\left[\underline{\Delta}_j, \overline{\Delta}_j\right]_{j=1}^{M_\varepsilon}$ be a collection of ε-brackets for H_M^c, i.e., for any $\Delta \in H_M^c$, there exists $1 \le j \le M_\varepsilon$ such that $\underline{\Delta}_j \le \Delta \le \overline{\Delta}_j$ and $\left\|\overline{\Delta}_j - \underline{\Delta}_j\right\|_{L_Z^2} \le \varepsilon$, and thus $\mathbb{1}_{\left(-\infty, u+\underline{\Delta}_j\right]} \le \mathbb{1}_{(-\infty, u+\Delta]} \le \mathbb{1}_{(-\infty, u+\overline{\Delta}_j]}$. Now, for each $1 \le j \le M_\varepsilon$, partition the real line into intervals defined by grids of points $-\infty = \underline{u}_{j,1} < \underline{u}_{j,2} < \cdots < \underline{u}_{j,M_{1\varepsilon}} = \infty$ and $-\infty = \overline{u}_{j,1} < \overline{u}_{j,2} < \cdots < \overline{u}_{j,M_{2\varepsilon}} = \infty$, so that each segment has probabilities

$$\Pr\left(U - \underline{\Delta}_j(Z) \le \underline{u}_{j,k}\right) - \Pr\left(U - \underline{\Delta}_j(Z) \le \underline{u}_{j,k-1}\right) \le \varepsilon^2/2, \qquad 2 \le k \le \frac{2}{\varepsilon^2} \triangleq M_{1\varepsilon},$$

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

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

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

Therefore, we constructed brackets of size $O(\varepsilon)$, covering $\mathcal{G}_1$, and we have used at most $O\left(\varepsilon^{-2} M_\varepsilon\right)$ such brackets. Since $c > p/2$, we have $\int_0^1 \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, \|.\|_{L_Z^2})}\mathrm{d}\varepsilon < \infty$. 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, Thm. 1.5.7). Then, for any $\varepsilon > 0$,

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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

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

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,

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

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,

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

Lastly, let $\|\hat{v}_n\|_\infty$ denote the supremum in equation (A.2). Then

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

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 equation (A.3) that $\limsup_{n\to\infty}\Pr^*(\sqrt{n}\|\hat{v}_n\|_\infty>\varepsilon)=0$, which concludes the proof. □

LEMMA A.4.2. *Suppose that Assumptions 3.1–3.4 are satisfied. Then, uniformly over* $(u,w)\in\boldsymbol{R}\times\boldsymbol{R}^q$,

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

and

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

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

Proof. Note that the first expression and the expression in the statement of Lemma A.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. □

LEMMA A.4.3. *Suppose that Assumptions 3.1–3.4 are satisfied. Then*

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

Proof. Similarly to the proof of Theorem A.1.1, decompose

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

with

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

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

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

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,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

$$= \sqrt{n} \sup_{\lambda} \left| \frac{\alpha_n \lambda}{\alpha_n + \lambda} \right|$$
$$\lesssim \sqrt{n}\alpha_n = o(1),$$

where the first equality follows since L is self-adjoint; the second line by the Cauchy–Schwartz inequality since $\sup_u \|L^a f_{UZ}(u,.)\| < \infty$ 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 et al. (2000, Cor. 8.22); the fourth by the isometry of the functional calculus (see Cavalier, 2011, Thm. 1.3); and the last since $2s = b - a$ and since $n\alpha_n^2 \to 0$ under Assumption 3.4(iii).

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

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

By the Cauchy–Schwartz inequality and previous computations

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

and

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

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

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

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

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

Likewise, by the Cauchy–Schwartz inequality and previous computations

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

and

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

where we use $\|\hat{T}-T\|_{\text{op}} \lesssim_P \alpha_n^{1/2}$ and $2s = b-a$ (see also the proof of Theorem A.1.1). Therefore, $\|IV_n\|_\infty = o_P(1)$.

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

$$I_n(u) = \sqrt{n}\left\langle T^*(\hat{r}-\hat{T}\varphi), (\alpha_n I + T^*T)^{-1}f_{UZ}(u,.)\right\rangle + o_P(1). \tag{A.4}$$

Next, note that

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

with $[\varphi * K_z](z) \triangleq \int \varphi(v)h_n^{-p}K_z\left(h_n^{-1}(z-v)\right)\mathrm{d}v$; thus,

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

with $[f_{ZW} * K_w](z,w) \triangleq \int f_{ZW}(z,v)h_n^{-q}K_w\left(h_n^{-1}(w-v)\right)\mathrm{d}v$. Using this observation, decompose equation (A.4) further

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

with

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

By the Cauchy–Schwartz inequality,

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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

where the second line follows by the 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),

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

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,

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

where the second inequality follows under Assumptions 3.2(i); the third line under Assumptions 3.1, 3.2(i) and (ii), and 3.3(i); and the last by the isometry of functional calculus (see Cavalier, 2011, Thm. 1.3).

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

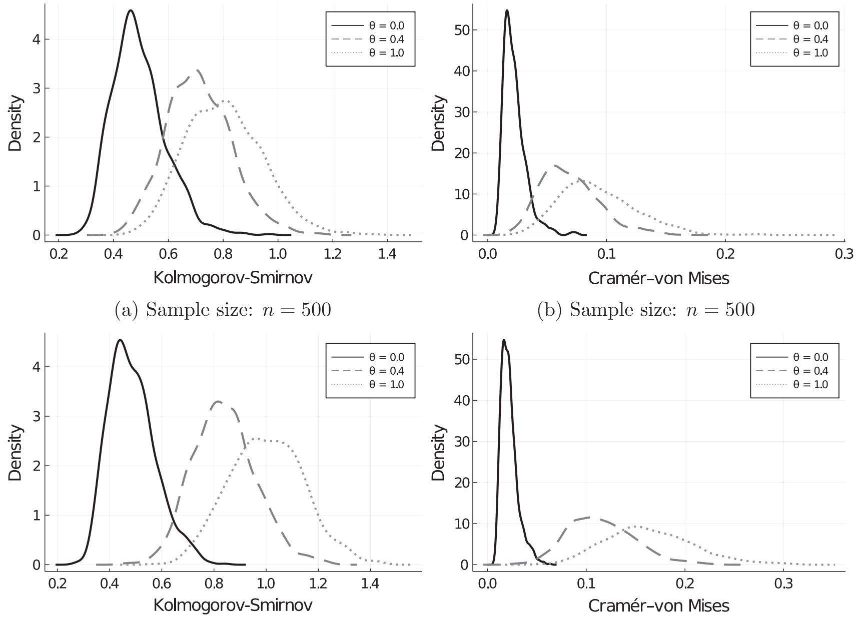
$$\begin{aligned}
II_n(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i\left\langle f_{ZW}(.,W_i),(T^*T)^{-1}f_{UZ}(u,.)\right\rangle + O_P\left(\sqrt{n}h_n^b+h_n^t+\alpha_n^{(\kappa/2a-1)\wedge 1}\right)+o_P(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i[T(T^*T)^{-1}f_{UZ}(u,.)](W_i)+o_P(1).
\end{aligned}$$ □

A.5. 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 A1 shows the distribution of the test statistics under the null hypothesis and the two alternative hypotheses for different sample sizes. The two distributions are

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press



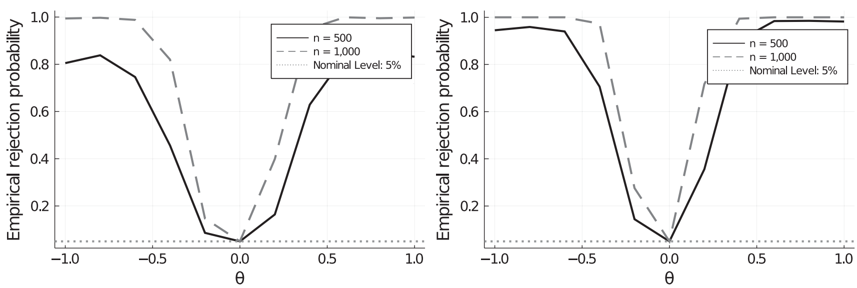
(a) Sample size: $n = 500$

(b) Sample size: $n = 500$

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

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

FIGURE A1. 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 A2. Power curves. The figure shows empirical rejection probabilities as a function of degree of separability θ 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%.

sufficiently distinct once the alternative hypothesis becomes more separated from the null hypothesis.

We plot in Figure A2 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

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

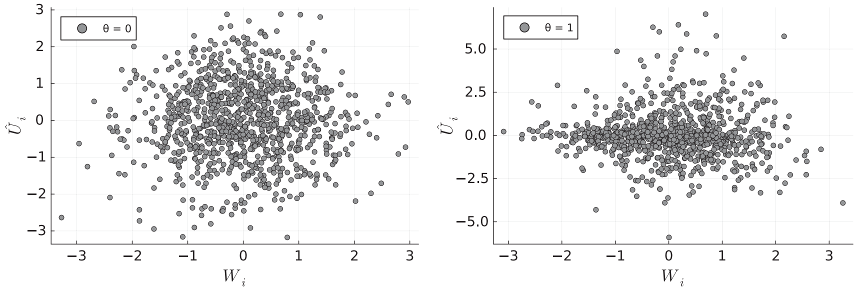


FIGURE A3. Scatter plot of residuals and instruments. The value $\theta = 0$ corresponds to the separable model and $\theta = 1$ to the nonseparable model. Sample size: $n = 1,000$.

hypothesis and when the sample size is larger. The CvM 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.

Lastly, Figure A3 plots the values of the instrumental variable, W_i, and the fitted nonparametric IV residuals, $\hat{U}_i$, under the null hypothesis, H_0 ($\theta = 0$), and under the alternative hypothesis, H_1 ($\theta = 1$). The figure hints that the two are indeed independent under the null and dependent under the alternative.

REFERENCES

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

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

Andrews, D. (1994). Chapter 37: Empirical process methods in econometrics. In *Handbook of econometrics* (vol. 4, pp. 2247–2294). Elsevier.

Babii, A. (2020). Honest confidence sets in nonparametric IV regression and other ill-posed models. *Econometric Theory*, 36(4), 658–706.

Babii, A. (2022). High-dimensional mixed-frequency IV regression. *Journal of Business and Economic Statistics*, 40(3), 1470–1483.

Babii, A., & Florens, J.-P. (2020). *Is completeness necessary? Estimation in nonidentified linear models*. Preprint. arXiv:1709.03473.

Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and Statistics*, 79(4), 527–538.

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

Blundell, R., Horowitz, J., & Parey, M. (2017). Nonparametric estimation of a nonseparable demand function under the Slutsky 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.

Carrasco, M., Florens, J.-P., & Renault, E. (2007). Chapter 77: Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In *Handbook of econometrics* (vol. 6B, pp. 5633–5751). Elsevier.

Carrasco, M., Florens, J.-P., & Renault, E. (2014). Asymptotic normal inference in linear inverse problems. In J. S. Racine, L. Su, & A. Ullah (Eds.), *The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics* (pp. 64–96). Oxford Academic.

Cavalier, L. (2011). Inverse problems in statistics. In P. Alquier, E. Gautier, & G. Stoltz (Eds.), *Inverse problems and high-dimensional estimation* (pp. 3–96). Springer.

Chen, X. (2007). Chapter 76: Large sample sieve estimation of semi-nonparametric models. In *Handbook of econometrics* (vol. 6, pp. 5549–5632). Elsevier.

Chen, X., Chernozhukov, V., Lee, S., & Newey, W. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, 82(2), 785–809.

Chen, X., & Christensen, T. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics*, 9(1), 39–84.

Chen, X., Linton, O., & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5), 1591–1608.

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

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4), 1501–1535.

Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., & Vella, F. (2020). Semiparametric estimation of structural functions in nonseparable triangular models. *Quantitative Economics*, 11(2), 503–533.

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

Chernozhukov, V., & Hansen, C. (2013). Quantile models with endogeneity. *Annual Review of Economics*, 5, 57–81.

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

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

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

Das, M. (2005). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics*, 124(2), 335–361.

Devroye, L. (1986). *Non-uniform random variate generation*. Springer.

Dunker, F., Florens, J.-P., Hohage, T., Johannes, J., & Mammen, E. (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., & Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, 143(1), 88–102.

Engl, H. W., Hanke, M., & Neubauer, A. (2000). *Regularization of inverse problems*. Springer.

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

Evans, L. C. (2010). *Partial differential equations*. Graduate Studies in Mathematics (vol. 19). American Mathematical Society.

Fève, F., Florens, J.-P., & Van Keilegom, I. (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. In M. Dewatripont, L. P. Hansen, & S. J. Turnovsky (Eds.), *Advances in economics and econometrics: Theory and applications, eighth World Congress* (vol. II, pp. 284–311). Cambridge University Press.

Florens, J.-P., Heckman, J., Meghir, C., & Vytlacil, E. (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., Johannes, J., & Van Bellegem, S. (2011). Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(3), 472–496.

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

Gagliardini, P., & Scaillet, O. (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., & Nickl, R. (2015). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press.

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

Heckman, J. J., & Vytlacil, E. (2001). Policy-relevant treatment effects. *American Economic Review*, 91(2), 107–111.

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

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

Imbens, G. (2010). Nonadditive models with endogenous regressors. In *Advances in economics and econometrics: Theory and applications, ninth World Congress* (vol. III, pp. 17–46). Cambridge University Press.

Imbens, G., & Newey, W. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5), 1481–1512.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475.

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

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

Lu, X., & White, H. (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.

Nair, M. (2015). Role of Hilbert scales in regularization theory. In P. G. Romeo, J. C. Meakin, & A. R. Rajan (Eds.), *Semigroups, algebras and operator theory: Kochi, India, February 2014* (pp. 159–176). Springer.

Nakamura, E., Steinsson, J., & Liu, M. (2016). Are Chinese growth and inflation too smooth? Evidence from Engel curves. *American Economic Journal: Macroeconomics*, 8(3), 113–144.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics*, 36(2), 204–228.

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

Newey, W., & Powell, J. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5), 1565–1578.

Nickl, R., & Pötscher, B. (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.

https://doi.org/10.1017/S026646662400001X Published online by Cambridge University Press

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

Su, L., Tu, Y., & Ullah, A. (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.

Torgovitsky, A. (2017). Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics*, 199(1), 35–48.

van der Vaart, A., & Wellner, J. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer.