

WORKING PAPERS

N° 17-802

May 2017

“Distribution of residuals in the nonparametric IV model
with application to separability testing”

Andrii Babii and Jean-Pierre Florens

Distribution of residuals in the nonparametric IV model with application to separability testing*

Andrii Babii[†]

Jean-Pierre Florens[‡]

Toulouse School of Economics

Toulouse School of Economics

May 4, 2017

We develop a uniform asymptotic expansion for the empirical distribution function of residuals in the nonparametric IV regression. Such expansion opens a door for construction of a broad range of residual-based specification tests in nonparametric IV models. Building on obtained result, we develop a test for the separability of unobservables in econometric models with endogeneity. The test is based on verifying the independence condition between residuals of the NPIV estimator and the instrument and can distinguish between the non-separable and the separable specification under endogeneity.

Keywords: separability test, distribution of residuals, nonparametric instrumental regression, Sobolev scales

JEL classification: C12, C14, C26

1 Introduction

This paper is concerned with studying properties of the distribution of residuals in the following nonparametric IV model

$$Y_i = \varphi(Z_i) + U_i, \quad \mathbb{E}[U_i|W_i] = 0, \quad i = 1, \dots, n. \quad (1)$$

*We are grateful to Ingrid van Keilegom and Elia Lapenta for helpful discussions.

[†]Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). Email: babii.andrii@gmail.com.

[‡]Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France).

Several approaches to the estimation of the structural function φ have been developed recently. Seminal papers include [Florens \(2003\)](#), [Newey and Powell \(2003\)](#), [Blundell, Chen, and Kristensen \(2007\)](#), [Darolles, Fan, Florens, and Renault \(2011\)](#), and [Hall and Horowitz \(2005\)](#).

There is also considerable literature on hypothesis testing related to this model. [Horowitz \(2006\)](#) develops a test for the parametric specification, while [Horowitz \(2012\)](#) discusses the test for the existence of the function φ solving the equation generated by the model. [Blundell and Horowitz \(2007\)](#) study the nonparametric exogeneity test using kernel approach. [Breunig \(2015\)](#) develop several tests when the function is estimated with sieves.

Additive separability of unobservables U_i in Eq. (1) is a convenient modeling assumption, however, it can often be hard to justify it from the economic theory. Imposing additive structure rules out heterogeneity of treatment effects across individuals, firms, or other economic entities, [Imbens \(2007\)](#). Aiming to relax separability assumption [Chernozhukov and Hansen \(2005\)](#), [Chernozhukov, Imbens, and Newey \(2007\)](#), [Torgovitsky \(2015\)](#), [D'Haultfœuille and Février \(2015\)](#) study identification in the non-separable model. Strictly speaking, the non-separable model is not more general than the nonparametric IV model, since it imposes a complete stochastic independence assumption, which is stronger than the mean-independence restriction in Eq (1). Moreover, nonparametric estimation of this model leads to the non-linear ill-posed inverse problem, [Horowitz and Lee \(2007\)](#) and [Dunker, Florens, Hohage, Johannes, and Mammen \(2014\)](#). As a result, estimation of non-separable model is considerably more difficult, unless parametric structure is imposed, e.g. see [Torgovitsky \(2017\)](#).

Starting from the nonseparable model, we would like to test whether it actually has a separable representation. To the best of our knowledge no such test is available in the literature. The closest studies are due to [Lu and White \(2014\)](#) and [Lewbel, Lu, and Su \(2015\)](#) who develop a separability test for the different model based on the conditional independence restriction $Y_i = \varphi(Z_i, U_i), U_i \perp\!\!\!\perp Z_i | W_i$.

Our test is based on the idea that under the hypothesis that the model has a separable representation, one can estimate it using the NPIV approach. Estimated residuals asymptotically converge to the unobserved error-terms, which should be stochastically independent from the IV under the maintained hypothesis. In this way, we can test for separability by testing the independence between the error term and residuals. Therefore, the paper also contributes to the residual-based goodness of fit testing literature

As a byproduct we obtain the uniform asymptotic expansion for the series NPIV estimator, which may be of independent interest in the context of other testing problems. To the best of our knowledge, such asymptotic expansions are available in the literature only for estimators in

models *without endogeneity*. Uniform expansion of residuals distribution for the linear regression model is a classical problem in statistics, [Durbin \(1973\)](#) and [Loynes \(1980\)](#), see also [Mammen et al. \(1996\)](#) who studied the cases where the number of regressors is allowed to increase with the sample size.

[Akritas and Van Keilegom \(2001\)](#) study asymptotic properties of the empirical distribution of function of residuals in the nonparametric location-scale model estimated with kernel smoothing. Some examples of residual-based specification tests for the nonparametric location-scale model include: the exogeneity test, [Einmahl and Van Keilegom \(2008\)](#), testing against parametric or semiparametric specifications, or testing the equality of regression curves [Van Keilegom, Manteiga, and Sello \(2008\)](#). Building on our results, similar tests can also be developed for the NPIV model. The uniform asymptotic expansion can also be used to construct prediction intervals, [Akritas and Van Keilegom \(2001\)](#).

It is also worth mentioning that some specification tests exist for the non-separable model, including the monotonicity test, [Hoderlein, Su, White Jr, and Yang \(2014\)](#), the specification tests for quantile IV regression, [Breunig \(2013\)](#) and the endogeneity test, [Fève, Florens, and van Keilegom \(2013\)](#).

Another by-product of our study is investigation of regularization and convergence rates of general Tikhonov-regularized estimator in Sobolev spaces on unbounded domains. Tikhonov-regularization in integer Sobolev spaces on bounded domains was considered previously in [Gagliardini and Scaillet \(2012\)](#). [Florens, Johannes, and Van Belleghem \(2011\)](#) and [Carrasco, Florens, and Renault \(2013\)](#) study more general problem of Tikhonov regularization in Hilbert scales. However, our result is not nested withing any of above studies and provides convergence rates under low-level assumptions when the operator is not known and is estimate from the data.

The paper is organized as follows. In the next section, we describe Tikhonov-regularized estimator with Sobolev penalty. In Section 3, we investigate asymptotic properties of residuals in the nonparametric IV model and establish Donsker-type central limit theorem for the empirical distribution function of residuals. In Section 4 we develop a test for separability. The last section concludes.

2 Regularization in Sobolev scales

Asymptotic developments in this paper rely heavily on the empirical process theory. It turns out that regularization in Sobolev spaces provides a natural link between the empirical process theory and the theory of regularization of ill-posed inverse models. The key argument in the

proof of Theorem 2 relies on the consistency of the NPIV estimator in the Sobolev norm. This result can be obtained considering regularization with Sobolev norm penalty. [Gagliardini and Scaillet \(2012\)](#) studied Tikhonov regularization in integer Sobolev spaces. This idea can be nested within a more general framework of regularization in continuous Hilbert scales, e.g. [Florens et al. \(2011\)](#) and [Carrasco et al. \(2013\)](#). Unfortunately, it is well-known that Sobolev spaces on bounded domains are not Hilbert scales, [Neubauer \(1988\)](#). This motivates using Sobolev spaces on unbounded domains. As a byproduct we obtain convergence rates for the nonparametric IV regression when data have unbounded supports¹.

For the weight function $\langle x \rangle : x \mapsto (1 + \|x\|^2)^{1/2}, x \in \mathbf{R}^p$ and for any $s \in \mathbf{R}$, consider the Sobolev space

$$H^s(\mathbf{R}^p) = \{f \in L_2(\mathbf{R}^p) : \|f\|_s = \|\langle x \rangle^s Ff\| < \infty\}, \quad (2)$$

where $F : \phi \mapsto \frac{1}{(2\pi)^{p/2}} \int_{\mathbf{R}^p} \phi(y) e^{i\langle x, y \rangle} dy$ is a Fourier transform on the space of square-integrable functions $L_2(\mathbf{R}^p)$. The scale of spaces $(H^s(\mathbf{R}^p))_{s \in \mathbf{R}}$ is a Hilbert scale generated by the operator $L : f \mapsto F^{-1}\langle x \rangle Ff$, whose powers are defined as $L^s = F^{-1}\langle x \rangle^s F$. Since $L^2 = I - \Delta$, where $\Delta = \sum_{j=1}^p \frac{\partial^2}{\partial x_j^2}$ is Laplacian, we have yet another interpretation that spaces $H^s(\mathbf{R}^p)$ are generated by the square root of the differential operator $L = \sqrt{I - \Delta}$. For a natural number s , the norm $\|\cdot\|_s$ is equivalent to the usual Sobolev norm defined in terms of weak derivatives, see [Krein and Petunin \(1966\)](#) for more details.

The NPIV model is recognized to be an example of ill-posed inverse problem. The mean-independence restriction in Eq (1) leads to the ill-posed integral equation

$$r(w) := \mathbb{E}[Y|W = w]f_W(w) = \int \varphi(z)f_{ZW}(z, w)dz =: (T\varphi)(w), \quad (3)$$

where $T : L_2(\mathbf{R}^p) \rightarrow L_2(\mathbf{R}^q)$ is integral operator.

Before studying the nonparametric IV regression, we present general result, valid for any ill-posed operator equation $T\varphi = r$ with one-to-one operator $T : L_2(\mathbf{R}^p) \rightarrow L_2(\mathbf{R}^q)$, where r and T can be consistently estimated from the data with \hat{r} and \hat{T} . For $s > 0$, the Tikhonov-regularized estimator $\hat{\varphi}_{\alpha_n}$ solves

$$\min_{\varphi} \left\| \hat{T}\varphi - \hat{r} \right\|^2 + \alpha_n \|\varphi\|_s^2. \quad (4)$$

¹This is yet another contribution of the paper since most of results for the nonparametric IV regression are available for compactly supported data, see e.g. [Darolles et al. \(2011\)](#), [Hall and Horowitz \(2005\)](#), [Chen and Christensen \(2015\)](#) with exception for sieve estimator of [Blundell et al. \(2007\)](#) that allow for the endogenous regressors to have unbounded support.

The estimator has closed-form expression

$$\hat{\varphi} = L^{-s}(\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* \hat{r}, \quad (5)$$

with $\hat{T}_s = \hat{T} L^{-s}$.

The following assumption is natural to describe smoothness in Hilbert spaces. Roughly speaking, it tells us that the operator T acts on the Sobolev scale, increasing smoothness by a , while the structural function φ is assumed to be in the Sobolev space with smoothness b .

Assumption 1. *For some $a, b > 0$, we have*

- (i) *Operator smoothing: $\|T\phi\|_r \sim \|\phi\|_{r-a}$ for all $\phi \in L_2(\mathbf{R}^p)$ and $r > 0$.*
- (ii) *Parameter smoothness: $\varphi \in H^b(\mathbf{R}^p)$.*

The next result describes convergence rates for the $\|\cdot\|_s$ -risk of Tikhonov-regularized estimator in Hilbert scales in terms of the regularization parameter α_n and convergence rates for the operator T and the left-side r .

Theorem 1. *Suppose that Assumption 1 is satisfied. Then*

$$\mathbb{E} \|\hat{\varphi}_{\alpha_n} - \varphi\|_s^2 = O\left(\frac{1}{\alpha_n^2} \mathbb{E} \left\| L^{-s} \hat{T}^* (\hat{r} - \hat{T}\varphi) \right\|^2 + \alpha_n^{\frac{b-s}{a+s}} + \frac{1}{\alpha_n^2} \left(\mathbb{E} \|\hat{T} - T\|^2 + \mathbb{E} \|\hat{T}^* - T^*\|^2 \right) \alpha_n^{\frac{b-s}{a+s}}\right), \quad (6)$$

provided $s \geq (b - a)/2$.

Now we specialize the risk bound in Theorem 1 to the nonparametric IV model. To that end we introduce several additional assumptions, which are standard for non-parametric estimation based on kernel smoothing.

Assumption 2. (i) *there exists some $t > 0$ such that joint density f_{ZW} is in the t -Sobolev ball of radius M ; (ii) kernel functions $K_z : \mathbf{R}^p \rightarrow \mathbf{R}$ and $K_w : \mathbf{R}^q \rightarrow \mathbf{R}$ are such that for $l \in \{w, z\}$, $K_l \in L_2(\mathbf{R})$, $\int K_l(u) du = 1$, $\int \|u\|^s K_l(u) du < \infty$, and $\int u^k K_l(u) du = 0$ for all multindices $|k| = 1, \dots, \lfloor t \rfloor$.*

Let $(\Omega, \Sigma, \Pr) = (\mathcal{X}^{\mathbf{N}_0}, \mathcal{B}^{\mathbf{N}_0}, \mathbf{P}^{\mathbf{N}_0})$ be the probability space and let $X_i = (Y_i, Z_i, W_i), i \in \mathbf{N}_0$ be coordinate projections $X_i : \mathcal{X}^{\mathbf{N}_0} \rightarrow \mathcal{X}$ with law \mathbf{P} . Denote $X = (Y, Z, W) = (Y_0, Z_0, W_0)$. We

estimate T and r in the NPIV model as

$$\begin{aligned}\hat{r}(w) &= \frac{1}{nh_n^q} \sum_{i=1}^n Y_i K_w(h_n^{-1}(W_i - w)) \\ (\hat{T}\phi)(w) &= \int_{\mathbf{R}^p} \phi(z) \hat{f}_{ZW}(z, w) dz, \quad \phi \in L_2(\mathbf{R}^p) \\ \hat{f}_{ZW}(z, w) &= \frac{1}{nh_n^{p+q}} \sum_{i=1}^n K_z(h_n^{-1}(Z_i - z)) K_w(h_n^{-1}(W_i - w)).\end{aligned}\tag{7}$$

The next assumption is a mild regularity restriction in the data and densities corresponding to probability distributions.

Assumption 3. (i) The data $\mathcal{X} = (Y_i, Z_i, W_i)_{i=1}^\infty$ are i.i.d. observations obeying the model in Eq. (1) (ii) the law \mathbf{P} of (Y, Z, W) has Lebesgue density f_{YZW} . Densities $f_{ZW}, f_{UZ}, f_Z, f_W, f_U$ are in L_∞ and uniformly bounded away from zero, and $f_{ZW} \in L_{2+\delta}(\mathbf{R}^{p+q})$ for some $\delta > 0$; (iii) $\mathbb{E}\|Z\|^{2\gamma} < \infty$ and $\mathbb{E}\|W\|^{2\gamma}$ for some $\gamma > 0$; (iv) $\mathbb{E}[U^2|W] \leq \bar{\sigma}^2 < \infty$ a.s.

Proposition 1. Suppose that Assumptions of Theorem 1 are satisfied. Suppose additionally that Assumptions 2 and 3 (i), (ii) are satisfied. Then if $\frac{1}{nh_n^{p+q}} = O(1)$, for any $s > 0$ such that $b \leq a + 2s$

$$\mathbb{E}\|\hat{\varphi}_{\alpha_n} - \varphi\|_s^2 = O\left(\frac{1}{\alpha_n^2} \left(\frac{1}{n} + h_n^{2b}\right) + \alpha_n^{\frac{b-s}{a+s}} + \frac{1}{\alpha_n^2} \left(\frac{1}{nh_n^{p+q}} + h_n^{2t}\right) \alpha_n^{\frac{b-s}{a+s}}\right).\tag{8}$$

This risk bound as well as its proof will be used in the following section to develop the asymptotic distribution of the separability test.

3 CLT for the empirical distribution of residuals

In this section we show that under appropriate assumptions on the dimension parameter α_n , the bias coming from the estimation of the function φ is asymptotically negligible for the residual empirical process. On the other hand, the variance will contribute to the additional term, which would not be present if we knew the error-term. To that end we fix several assumptions.

Assumption 4. The joint density function f_{VZW} is differentiable in the first argument and

$$(i) \quad \|\partial_u f_{UZ}(u, z)\|_\infty < \infty, \quad (ii) \quad \sup_{u, w, z} \left| \int_{-\infty}^w \partial_u f_{UZW}(u, z, \tilde{w}) d\tilde{w} \right| < \infty.\tag{9}$$

Moreover, (iii) $f_{UZ}(u, \cdot) = T^*(T_s T_s^*)^\rho \psi(u, \cdot)$ for some $\rho > \frac{q}{4(a+s)} \vee \frac{1}{2}$ and function $\psi : \mathbf{R}^{p+1} \rightarrow \mathbf{R}$ such that $\|\psi\|_{2, \infty} = \sup_{u \in \mathbf{R}} \|\psi(u, \cdot)\| < \infty$, and (iv) $\int_{\{\tilde{w} \leq w\}} f_{UZW}(u, z, \tilde{w}) d\tilde{w} = T^*(T_s T_s^*)^\rho \tilde{\psi}(u, w, \cdot)$

for the same ρ and some $\tilde{\psi}$ with $\sup_{u \in \mathbf{R}, w \in \mathbf{R}^q} \|\tilde{\psi}(u, w, \cdot)\| < \infty$.

Assumption 4 (iii) is an additional assumption that needed, comparing to the direct regression model, Akritas and Van Keilegom (2001). It tells us that the joint density of the error term and the regressor should be sufficiently smooth in order to overcome ill-posedness of the model. Given that the strength of the instrument is characterized by the smoothness of the joint density of the regressor and the instrument f_{ZW} , this condition can be interpreted as a requirement to rule-out extreme forms of endogeneity.

Next, we introduce conditions on tuning parameters, which allow to obtain a non-degenerate asymptotic distribution for the empirical distribution of residuals as well as the separability test.

Assumption 5. *Suppose that Assumptions 1 and 2 are satisfied with $b > 2$, $b > 2a \geq t > (p+q)/2$ and that $s > p/2$, $s \geq (b-a)/2$. Suppose also that a, b, t, p, q and the sequence of tuning parameters are such that $n\alpha_n^2 h_n^{p+q} \rightarrow \infty$, $n\alpha_n^{b/a} \rightarrow 0$, and $h_n/\alpha_n \rightarrow 0$ as $\alpha_n \rightarrow 0$ and $h_n \rightarrow 0$.*

Consider residuals $\hat{U}_i = Y_i - \hat{\varphi}(Z_i)$. We are interested in the behavior of the residual empirical process

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i \leq u\}} - \mathbf{P}(U \leq u) \right) =: \sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)). \quad (10)$$

The next theorem shows that residual empirical process admits convenient uniform asymptotic expansion.

Theorem 2. *Suppose that Assumptions 1, 2, 3, 4 (i), (iii), and 5, are satisfied. Then under the null hypothesis, the following expansion holds uniformly over $u \in \mathbf{R}$*

$$\sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{1}_{\{U_i \leq u\}} - F_U(u) + U_i [T(T^*T)^{-1} f_{UZ}(u, \cdot)](W_i) \right\} + o_p(1), \quad (11)$$

where the converge is in outer probability \Pr^* .

The next result is a simple consequence of Theorem 2.

Corollary 1. *Under assumptions of Theorem 2,*

$$\sqrt{n}(\hat{F}_{\hat{U}}(u) - F_U(u)) \rightsquigarrow \mathbb{G}(u) \quad \text{in } L_\infty(\mathbf{R}), \quad (12)$$

where $\mathbb{G}(u)$ is a tight centered Gaussian process with uniformly continuous sample path and

covariance structure

$$\begin{aligned}
(u, u') \mapsto & F_U(u \wedge u') - F_U(u)F_U(u') + \mathbb{E} [U_1^2 [T(T^*T)^{-1}f_{UZ}(u, \cdot)] (W_1) [T(T^*T)^{-1}f_{UZ}(u', \cdot)] (W_1)] \\
& + \mathbb{E} [\mathbb{1}_{\{U_1 \leq u\}} U_1 [T(T^*T)^{-1}f_{UZ}(u', \cdot)] (W_1)] + \mathbb{E} [\mathbb{1}_{\{U_1 \leq u'\}} U_1 [T(T^*T)^{-1}f_{UZ}(u, \cdot)] (W_1)]
\end{aligned} \tag{13}$$

Therefore, despite the fact that residuals are coming from the estimation of an ill-posed inverse problem, their empirical distribution function converges to the true distribution at the square-root n speed.

4 Testing separability of unobservables

In this section we discuss the separability test. Assume that the law \mathbf{P} of (Y, Z, W) admits the non-separable representation, i.e. it belongs to the class of probability distributions \mathcal{P} such that there exists a measurable function $\Phi : \mathbf{R}^p \times \mathbf{R} \rightarrow \mathbf{R}$, strictly increasing in the second argument, and a random variable $\varepsilon \perp\!\!\!\perp W$, uniformly distributed on $[0, 1]$ such that $Y = \Phi(Z, \varepsilon)$ a.s.

We would like to test the null hypothesis, that the non-separable model has separable representation

$$H_0 : \mathbf{P} \in \mathcal{P}_0 = \{\mathbf{P} \in \mathcal{P} : \exists \psi : \mathbf{R}^p \rightarrow \mathbf{R}, g : \mathbf{R} \rightarrow \mathbf{R}, \text{ s.t. } Y = \psi(Z) + g(\varepsilon)\}, \tag{14}$$

against the alternative hypothesis

$$H_a : \mathbf{P} \in \mathcal{P} \setminus \mathcal{P}_0. \tag{15}$$

Notice that without loss of generality we can assume that $\mathbb{E}[g(\varepsilon)] = 0$. Then $\mathbb{E}[g(\varepsilon)|W] = 0$ and we can estimate the function φ consistently using the nonparametric IV approach.

Proposition 2. *Suppose that there exists unique φ such that $\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W]$ and let $U := Y - \varphi(Z)$. Then H_0 holds if and only if $U \perp\!\!\!\perp W$.*

Proposition 2 can be used to build the test statistics based on independence test between the error term and the instrument. Under the null hypothesis, the NPIV estimator $\hat{\varphi}$ is consistent for φ and so the difference between the estimated residuals $\hat{U}_i = Y_i - \hat{\varphi}(Z_i)$ and the true error-term U_i should become negligible asymptotically. The test-statistics can be build around the following process

$$\mathbb{G}_n(u, w) = \sqrt{n} \left(\hat{F}_{\hat{U}W}(u, w) - \hat{F}_{\hat{U}}(u) \hat{F}_W(w) \right), \tag{16}$$

which under the null hypothesis by Proposition 2 should have a "small size". On the other

hand, under the alternative hypothesis $F_{UW} \neq F_U F_W$ for some set of positive measure, so that asymptotically the process \mathbb{G}_n will not be zero on this set. The power of the test against the fixed alternative will depend on how different F_{UW} is from $F_U F_W$ for that alternative.

The Donsker central limit theorem can not be applied to the process \mathbb{G}_n , since it is based on pseudo-observations of residuals \hat{U}_i . Using the NPIV estimator, one introduces some bias and additional variance. Our main theorem shows that the bias converge to zero under the appropriate assumptions on the tuning parameter. On the other hand, the variance will affect the covariance structure of the process through additional term.

Theorem 3. *Suppose that Assumptions 1, 2, 3, 4 (ii)-(iv), and 5, are satisfied. Then, under the null hypothesis, the following asymptotic expansion holds uniformly in $(v, w) \in \mathbf{R} \times \mathbf{R}^q$*

$$\mathbb{G}_n(u, w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{1}_{\{U_i \leq u, W_i \leq w\}} - \mathbb{1}_{\{U_i \leq u\}} F_W(w) - \mathbb{1}_{\{W_i \leq w\}} F_U(u) + F_{UW}(u, w) + \delta_i(u, w) \right\} + o_p(1), \quad (17)$$

with

$$\begin{aligned} \delta_i(u, w) &= U_i (T(T^*T)^{-1} g(\cdot, u, w)) (W_i), \\ g(z, u, w) &= \int_{\{\tilde{w} \leq w\}} f_{UZW}(u, z, \tilde{w}) d\tilde{w} - f_{UZ}(u, z) F_W(w). \end{aligned} \quad (18)$$

Corollary 2. *Under assumptions of Theorem 3,*

$$\mathbb{G}_n(u, w) \rightsquigarrow \mathbb{G}(u, w) \quad \text{in } L_\infty(\mathbf{R}), \quad (19)$$

where $\mathbb{G}(u, w)$ is a tight centered Gaussian process with uniformly continuous sample path and contaminated covariance structure

$$\begin{aligned} (u, w, u', w') \mapsto \mathbb{E} \left[\left(\mathbb{1}_{\{U_i \leq u, W_i \leq w\}} - \mathbb{1}_{\{U_i \leq u\}} F_W(w) - \mathbb{1}_{\{W_i \leq w\}} F_U(u) + F_{UW}(u, w) + \delta_i(u, w) \right) \times \right. \\ \left. \times \left(\mathbb{1}_{\{U_i \leq u', W_i \leq w'\}} - \mathbb{1}_{\{U_i \leq u'\}} F_W(w') - \mathbb{1}_{\{W_i \leq w'\}} F_U(u') + F_{UW}(u', w') + \delta_i(u', w') \right) \right]. \end{aligned} \quad (20)$$

5 Monte Carlo experiments

To evaluate the finite-sample behavior of the test, we simulate samples of size $n \in \{1000, 5000\}$ from the model

$$Y_i = \sin(Z_i + \theta V_i) + V_i, \quad (21)$$

where

$$\begin{pmatrix} Z \\ W \\ V \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 & 0.3 \\ 0.9 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix} \right) \quad (22)$$

The degree of separability is governed by the $\theta \in \mathbf{R}$ parameter. The value $\theta = 0$ corresponds to the separable model while any $\theta \neq 0$ to alternative nonseparable model. The number of Monte Carlo replications is 5000.

We look at the Kolmogorov-Smirnov statistics

$$T_\infty = \sup_{u,v} |\mathbb{G}_n(u, v)| \quad (23)$$

and at the Cramer von Mises statistics

$$T_2 = \int \int |\mathbb{G}(u, v)|^2 dudv. \quad (24)$$

Figure 1 shows the distribution of the test statistics under the null hypothesis and under two alternatives for different sample sizes. The two distributions are sufficiently distinct once the alternative hypothesis moves away from the null. Figure 2 shows empirical rejection probabilities for the level fixed at 5%. Again, the power of the test increases once we move away from the null hypothesis. For this particular DGP and the class of alternative hypotheses the Cramer von Mises test seems to have higher power.

6 Conclusion

This paper studies the asymptotic distribution of the empirical distribution of residuals in the nonparametric IV model. Despite the fact that residuals are obtained from the estimation of ill-posed inverse model, their empirical distribution function can converge weakly to the Gaussian process at the root- n speed.

We apply this result to develop a test of separability of unobservables in econometric models with endogeneity. The test can detect whether the non-separable IV model has separable representation. The latter model rules out heterogeneity of treatment effects, but can be estimated significantly easier. Given the plethora of residual-based goodness of fit tests for regression models without endogeneity, similar tests can be developed for IV models using our results.

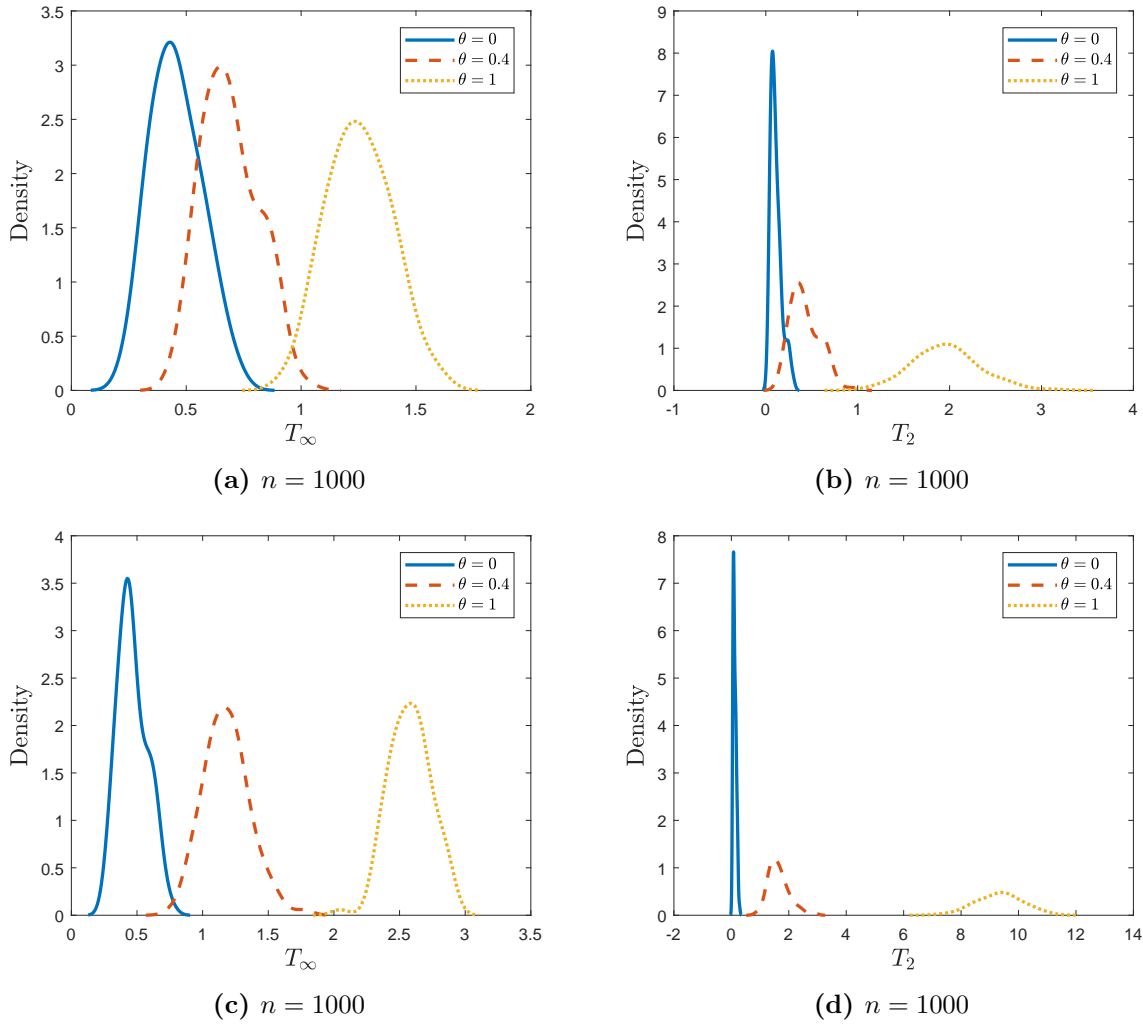


Figure 1: Densities of the finite-sample distribution of the test statistics under the null and two alternatives with different sample sizes.

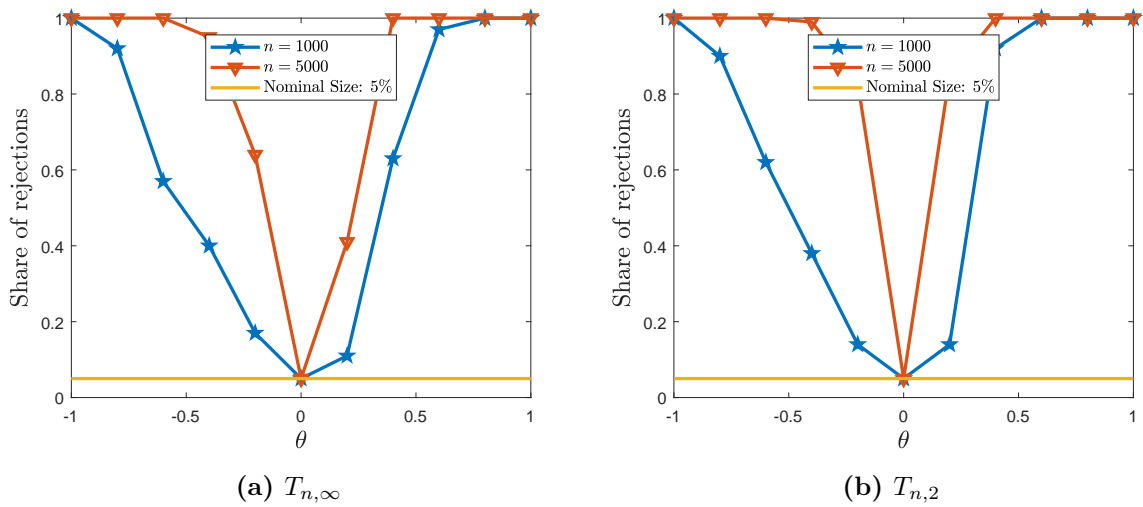


Figure 2: Power of the test.

Appendix: Proofs

Proof of Proposition 2. If the null hypothesis holds, then $\varphi = \psi + \mathbb{E}g$ and $U = g(\varepsilon) - \mathbb{E}g(\varepsilon) \perp\!\!\!\perp W$. On the other side if $U \perp\!\!\!\perp W$, then we can take $\psi = \varphi$ and any $\varepsilon \perp\!\!\!\perp W$, uniformly distributed on $[0, 1]$, so that $g = F_U^{-1}$, where $F^{-1}(u) = \inf \{x \in \mathbf{R} : F(x) \geq u\}$, $u \in (0, 1)$ is the generalized inverse of F_U . \square

Lemma 1. *Suppose that Assumptions 1, 2, 3, 4, 5 are satisfied. Then*

$$\sup_u \left| \hat{F}_{\hat{U}}(u) - \hat{F}_U(u) - \Pr \left(U \leq u + \hat{\Delta}(Z) | \mathcal{X} \right) + F_U(u) \right| = o_p \left(n^{-1/2} \right), \quad (25)$$

where $\hat{\Delta}(z) = \hat{\varphi}(z) - \varphi(z)$, $\mathcal{X} = (Y_i, Z_i, W_i)_{i=1}^{\infty}$ and the convergence is in outer probability \Pr^* .

Proof. The proof is based on the asymptotic equicontinuity argument, inspired by [Akritas and Van Keilegom \(2001\)](#). Let H_M^s be a ball of radius $M < \infty$ in the Sobolev space $H^s(\mathbf{R}^p)$ and consider the following classes of functions

$$\begin{aligned} \mathcal{G} &= \{ \mathbb{1}_{(-\infty, u+\Delta]} - \mathbb{1}_{(-\infty, u]} : (u, \Delta) \in \mathbf{R} \times H_M^s \} \\ &=: \mathcal{G}_1 - \mathcal{G}_2, \end{aligned} \quad (26)$$

where

$$\begin{aligned} \mathcal{G}_1 &= \{ \mathbb{1}_{(-\infty, u+\Delta]} : (u, \Delta) \in \mathbf{R} \times H_M^s \}, \\ \mathcal{G}_2 &= \{ \mathbb{1}_{(-\infty, u]} : u \in \mathbf{R} \}. \end{aligned} \quad (27)$$

\mathcal{G}_2 is a classical Donsker class. If we can show that \mathcal{G}_1 is Donsker, then \mathcal{G} will be Donsker as a sum of two Donsker classes, ([Van Der Vaart and Wellner, 2000](#), Theorem 2.10.6.). To this end, we check that the bracketing entropy condition is satisfied for \mathcal{G}_1 . By ([Nickl and Pötscher, 2007](#), Corollary 4), the ε -bracketing number for the ball H_M^s is bounded by

$$M_\varepsilon := N_{[\cdot]}(\varepsilon, H_M^s, \|\cdot\|_{\mathbf{P}}) \leq \exp(Kh(\varepsilon)), \quad \forall \varepsilon \in (0, 1]. \quad (28)$$

Fix $u \in \mathbf{R}$ and let $[\underline{\Delta}_j, \overline{\Delta}_j]_{j=1}^{M_\varepsilon}$ be a collection of ε -brackets for H_M^s , i.e. for any $\Delta \in H_M^s$, there exists $1 \leq j \leq M_\varepsilon$ such that $\underline{\Delta}_j \leq \Delta \leq \overline{\Delta}_j$ and $\|\overline{\Delta}_j - \underline{\Delta}_j\|_{\mathbf{P}} \leq \varepsilon$ and so

$$\mathbb{1}_{(-\infty, u+\underline{\Delta}_j]} \leq \mathbb{1}_{(-\infty, u+\Delta]} \leq \mathbb{1}_{(-\infty, u+\overline{\Delta}_j]}. \quad (29)$$

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

segment has the following probability measures less or equal to $\varepsilon^2/2$:

$$\begin{aligned} \mathbf{P}(U - \underline{\Delta}_j(Z) \leq \underline{u}_{j,k-1}) - \mathbf{P}(U - \underline{\Delta}_j(Z) \leq \underline{u}_{j,k}) &\leq \varepsilon^2/2, & 2 \leq k \leq \frac{2}{\varepsilon^2} \equiv M_{1\varepsilon}, \\ \mathbf{P}(U - \overline{\Delta}_j(Z) \leq \overline{u}_{j,k-1}) - \mathbf{P}(U - \overline{\Delta}_j(Z) \leq \overline{u}_{j,k}) &\leq \varepsilon^2/2, & 2 \leq k \leq \frac{2}{\varepsilon^2} \equiv M_{2\varepsilon}. \end{aligned} \quad (30)$$

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

$$\left[\mathbf{1}_{(-\infty, \underline{u}_j^* + \underline{\Delta}_j]}, \mathbf{1}_{(-\infty, \overline{u}_j^* + \overline{\Delta}_j]} \right]_{j=1}^{M_\varepsilon}. \quad (31)$$

Under Assumption 4 (ii)

$$\begin{aligned} \left\| \mathbf{1}_{(-\infty, \overline{u}_j^* + \overline{\Delta}_j]} - \mathbf{1}_{(-\infty, \underline{u}_j^* + \underline{\Delta}_j]} \right\|_{\mathbf{P}}^2 &= \mathbf{P}(U \leq \overline{u}_j^* + \overline{\Delta}_j(Z)) - \mathbf{P}(U \leq \underline{u}_j^* + \underline{\Delta}_j(Z)) \\ &\leq \mathbf{P}(U \leq u + \overline{\Delta}_j(Z)) - \mathbf{P}(U \leq u + \underline{\Delta}_j(Z)) + \varepsilon^2 \\ &= \int_{\mathbf{R}^p} \left\{ \int_{-\infty}^{u + \overline{\Delta}_j(z)} f_{UZ}(v, z) dv - \int_{-\infty}^{u + \underline{\Delta}_j(z)} f_{UZ}(v, z) dv \right\} dz + \varepsilon^2 \\ &= \int_{\mathbf{R}^p} \left\{ \int_{u + \underline{\Delta}_j(z)}^{u + \overline{\Delta}_j(z)} f_{U|Z}(u, z) du \right\} f_Z(z) dz + \varepsilon^2 \\ &\leq \|\overline{\Delta}_j - \underline{\Delta}_j\|_{\mathbf{P}} \|f_{U|Z}\|_\infty + \varepsilon^2 = O(\varepsilon^2), \end{aligned} \quad (32)$$

where the last line follows by Cauchy-Schwartz inequality. In this way, we have just constructed brackets of $\|\cdot\|_{\mathbf{P}}$ -size less or equal to ε , covering \mathcal{G}_1 , and we have used at most $O(\varepsilon^{-2} \exp(Kh(\varepsilon)))$ such brackets. Since $s > p/2$, it follows by Jensen's inequality and elementary computations that the bracketing entropy integral is finite²

$$\int_0^1 \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|_{\mathbf{P}})} d\varepsilon < \infty. \quad (33)$$

Therefore, the process

$$\mathbb{G}_n(g) = \sqrt{n}(\mathbf{P}_n f - \mathbf{P} f), \quad f \in \mathcal{G}, \quad (34)$$

with empirical measure $\mathbf{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, is asymptotically equicontinuous, (Van Der Vaart and Wellner, 2000, Theorem 1.5.7), i.e. for any $\varepsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr^* \left(\sup_{f, g \in \mathcal{G}: \rho_{\mathbf{P}}(f-g) < \delta} |\mathbb{G}_n(f-g)| > \varepsilon \right) = 0. \quad (35)$$

²Notice that for $\varepsilon > 1$, $N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|_{\mathbf{P}}) = 1$, since a single bracket $[0, 1]$ contains any $g \in \mathcal{G}$.

Now, we show that the seminorm $\rho_{\mathbf{P}}$ of $\hat{f} = \mathbf{1}_{(-\infty, u + \hat{\Delta}]} - \mathbf{1}_{(-\infty, u]}$ converges to zero in probability

$$\rho_{\mathbf{P}}^2(\hat{f}) = \mathbb{E} \left[\left(\mathbf{1}_{\{U \leq u + \hat{\Delta}(Z)\}} - \mathbf{1}_{\{U \leq u\}} \right)^2 | \mathcal{X} \right] - \left\{ \Pr(U \leq u + \hat{\Delta}(Z) | \mathcal{X}) - \mathbf{P}(U \leq u) \right\}^2 \quad (36)$$

First, we have

$$\begin{aligned} \Pr(U \leq u + \hat{\Delta}(Z) | \mathcal{X}) - \mathbf{P}(U \leq u) &\leq \int_{\mathbf{R}^p} \left| \int_u^{u + \hat{\Delta}(z)} f_{UZ}(v, z) dv \right| dz \\ &\leq \|\hat{\varphi} - \varphi\| \|f_{U|Z}\|_{\infty} = o_p(1), \end{aligned} \quad (37)$$

where the last line follows by Cauchy-Schwartz inequality and Proposition 1 under Assumption 5.

Similarly, we show that

$$\begin{aligned} \mathbb{E} \left[\left(\mathbf{1}_{\{U \leq u + \hat{\Delta}(Z)\}} - \mathbf{1}_{\{U \leq u\}} \right)^2 | \mathcal{X} \right] &= \Pr(U \leq u + \hat{\Delta}(Z) | \mathcal{X}) + \mathbf{P}(U \leq u) - 2 \Pr(U \leq (u + \hat{\Delta}(Z)) \wedge u | \mathcal{X}) \\ &\leq \int_{\mathbf{R}^p} \left| \int_u^{u + \hat{\Delta}(z)} f_{UZ}(v, z) dv \right| dz = o_p(1). \end{aligned} \quad (38)$$

Lastly, under Assumption 5, $\mathbb{E}\|\hat{\Delta}\|_s = o(1)$ by Proposition 1. Therefore, by Markov's inequality $\Pr(\hat{\Delta} \in H_M^s) \rightarrow 1$.

Finally, denote the supremum in Eq (25) by $\|\hat{\nu}_n\|_{\infty}$. Then

$$\Pr^*(\sqrt{n}\|\hat{\nu}_n\|_{\infty} > \varepsilon) \leq \Pr^*(\sqrt{n}\|\hat{\nu}_n\|_{\infty} > \varepsilon, \rho_{\mathbf{P}}(\hat{f}) < \delta, \hat{\Delta} \in H_M^s) + \Pr^*(\rho_{\mathbf{P}}(\hat{f}) \geq \delta) + \Pr^*(\hat{\Delta} \notin H_M^s), \quad (39)$$

where the last two probabilities tend to 0 and so by Eq. (35)

$$\limsup_{n \rightarrow \infty} \Pr^*(\sqrt{n}\|\hat{\nu}_n\|_{\infty} > \varepsilon) = 0, \quad (40)$$

which concludes the proof. \square

Proof of Theorem 2. By Lemma 1,

$$\sqrt{n}(\hat{F}_{\hat{\nu}}(u) - F_U(u)) = \sqrt{n}(\hat{F}_U(u) - F_U(u)) + \sqrt{n} \left\{ \Pr(U \leq u + \hat{\Delta}(Z) | \mathcal{X}) - F_U(u) \right\} + o_p(1), \quad (41)$$

uniformly in $u \in \mathbf{R}$. By Taylor's theorem, there exists some ξ_z between u and $u + \hat{\Delta}(z)$ such that

$$\begin{aligned}
& \sqrt{n} \left(\Pr \left(U \leq u + \hat{\Delta}(Z) \mid \mathcal{X} \right) - \mathbf{P}(U \leq u) \right) \\
&= \sqrt{n} \int_{\mathbf{R}^p} \left\{ \int_{-\infty}^{u + \hat{\Delta}(z)} f_{UZ}(v, z) dv - \int_{-\infty}^u f_{UZ}(v, z) dv \right\} dz \\
&= \sqrt{n} \int_{\mathbf{R}^p} \hat{\Delta}(z) f_{UZ}(u, z) dz + \sqrt{n} \int_{\mathbf{R}^p} \hat{\Delta}^2(z) \frac{1}{2} \partial_u f_{UZ}(\xi_z, z) dz \\
&=: T_{1n}(u) + T_{2n}(u).
\end{aligned} \tag{42}$$

By Assumptions 4

$$\sup_{u \in \mathbf{R}} |T_{2n}(u)| \lesssim \sqrt{n} \|\hat{\varphi} - \varphi\|^2 \|\partial_u f_{UZ}\|_{\infty}, \tag{43}$$

which is $o_p(1)$ under Assumption 5 by Proposition 1. Now, as in the proof of Theorem 1, decompose

$$\hat{\varphi} - \varphi = \frac{1}{n} \sum_{i=1}^n U_i (T^* T)^{-1} f_{ZW}(\cdot, W_i) + \sum_{j=1}^6 R_{j,n}, \tag{44}$$

with

$$\begin{aligned}
R_{1,n} &= L^{-s} [(\alpha_n I + T_s^* T_s)^{-1} - (T_s^* T_s^*)^{-1}] L^{-s} \frac{1}{n} \sum_{i=1}^n U_i f_{ZW}(\cdot, W_i), \\
R_{2,n} &= L^{-s} [(\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} - (\alpha_n I + T_s^* T_s)^{-1}] L^{-s} \frac{1}{n} \sum_{i=1}^n U_i f_{ZW}(\cdot, W_i), \\
R_{3,n} &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} L^{-s} \left(T^* (\hat{r} - \hat{T} \varphi) - \frac{1}{n} \sum_{i=1}^n U_i f_{ZW}(\cdot, W_i) \right), \\
R_{4,n} &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} L^{-s} (\hat{T}^* - T^*) (\hat{r} - \hat{T} \varphi), \\
R_{5,n} &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{T}_s - T_s) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \\
R_{6,n} &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} (\hat{T}_s^* - T_s^*) \alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \\
R_{7,n} &= L^{-s} (\alpha_n I + T_s^* T_s)^{-1} T_s^* T_s L^s \varphi - \varphi.
\end{aligned} \tag{45}$$

Therefore,

$$T_{1n}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i [T (T^* T)^{-1} f_{UZ}(u, \cdot)](W_i) + \sum_{j=1}^7 \sqrt{n} \langle R_{j,n}, f_{UZ}(u, \cdot) \rangle \tag{46}$$

and it remains to show that the second sum is $o_p(1)$.

It follows from the proof of Theorem 1 and Proposition 1 that

$$\begin{aligned}
\mathbb{E}\|R_{3,n}\| &= O\left(\frac{h_n^t + h_n^b}{n^{1/2}} + h_n^b\right), \\
\mathbb{E}\|R_{4,n}\| &= O\left(\left(\frac{1}{n^{1/2}} + h_n^b\right)\left(\frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t\right)\right), \\
\mathbb{E}\|R_{5,n} + R_{6,n} + R_{7,n}\| &= O\left(\alpha_n^{b/2a} + \frac{1}{\alpha_n}\left(\frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t\right)\alpha_n^{b/2a}\right).
\end{aligned} \tag{47}$$

Similarly, by Cauchy-Schwartz and Assumption 4

$$\begin{aligned}
\|\langle R_{2,n}, f_{UZ} \rangle\|_\infty &\leq \left\| (\hat{T}_s^* - T_s^*) \hat{T}_s (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \right\| \left\| \frac{1}{n} \sum_{i=1}^n U_i L^{-s} f_{ZW}(\cdot, W_i) \right\| \left\| (\alpha_n I + T_s^* T_s)^{-1} L^{-s} f_{UZ} \right\|_{2,\infty} \\
&\quad + \left\| (\hat{T}_s - T_s) (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \right\| \left\| \frac{1}{n} \sum_{i=1}^n U_i L^{-s} f_{ZW}(\cdot, W_i) \right\| \left\| T_s (\alpha_n I + T_s^* T_s)^{-1} L^{-s} f_{UZ} \right\|_{2,\infty} \\
&= O_p\left(\frac{1}{\alpha_n n^{1/2}} \left(\frac{1}{\sqrt{nh_n^{p+q}}} + h_n^t\right)\right).
\end{aligned} \tag{48}$$

Lastly, we show that the process

$$\sqrt{n} \langle R_{1,n}, f_{UZ}(u, \cdot) \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i (T_s \alpha_n (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} L^{-s} f_{UZ}(u, \cdot))(W_i), \quad u \in \mathbf{R} \tag{49}$$

indexed by the following class of functions

$$\mathcal{F}_n = \{(u, w) \mapsto u (\alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} L^{-s} f_{UZ}(u, \cdot))(w), u \in \mathbf{R}\} \tag{50}$$

is degenerate. To that tend, notice that under Assumptions 1 (i) and 4 (iii) for any $r \in (q/2, 2\rho(a+s))$ by (Engl, Hanke, and Neubauer, 1996, Corollary 8.22) and isometry of functional calculus

$$\begin{aligned}
\left\| \alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} L^{-s} f_{UZ}(u, \cdot) \right\|_r &\lesssim \alpha_n \left\| (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} T_s^* (T_s T_s^*)^\rho \psi(u, \cdot) \right\|_{-(a+s-r)} \\
&\lesssim \alpha_n \left\| (T_s^* T_s)^{\frac{a+s-r}{2(a+s)}} (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} T_s (T_s T_s^*)^\rho \right\| \\
&= \sup_{[0, \|T_s\|^2]} \left| \lambda^{\rho + \frac{a+s-r}{2(a+s)} - \frac{1}{2}} \right| \\
&= \|T_s\|^{2\rho - \frac{r}{(a+s)}} =: M.
\end{aligned} \tag{51}$$

Therefore, there exists a constant M independent of n such that for a Sobolev ball of radius M ,

denoted H_M^r , we have

$$\mathcal{F}_n \subset \mathcal{H} = \{(v, w) \mapsto vg(w), g \in H_M^r\}. \quad (52)$$

The class \mathcal{H} is Donsker. To see this, let $[l_j, u_j]_{j=1}^{J_\varepsilon}$ be brackets for H_M^r of \mathbf{P} -size less than $\varepsilon/\bar{\sigma}$ for some $\varepsilon > 0$. Denote $v_- = -\min\{v, 0\}$ and $v_+ = \max\{v, 0\}$. Then since $v = v_+ - v_-$ and $v_+, v_- \geq 0$, $[v_+l_j - v_-u_j, v_+u_j - v_-l_j]_{j=1}^{J_\varepsilon}$ are brackets for \mathcal{H} and under Assumption 3 (iv), their size is

$$(\mathbb{E}|V(u_j(W) - l_j(W))|^2)^{1/2} \leq \bar{\sigma} \left(\mathbb{E}|u_j(W) - l_j(W)|^2 \right)^{1/2} < \varepsilon, \quad (53)$$

By (Nickl and Pötscher, 2007, Corollary 4), their number, J_ε is bounded as

$$N_{[]}(\varepsilon, \mathcal{H}, L_2(\mathbf{P})) \lesssim \begin{cases} \varepsilon^{-q/r} & \gamma > r - q/2 \\ \varepsilon^{-q/(\gamma+q/2)} & \gamma < r - q/2 \end{cases}. \quad (54)$$

Moreover, for any $f \in \mathcal{F}_n$, under Assumptions 3 (ii), (iv) and 4 (iii)

$$\begin{aligned} \rho_{\mathbf{P}}(f)^2 &= \mathbb{E} \left[U_i^2 (T_s \alpha_n (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} L^{-s} f_{UZ})^2 (W_i) \right] \\ &\lesssim \bar{\sigma}^2 \left\| \alpha_n T_s^* (\alpha_n I + T_s^* T_s)^{-1} (T_s^* T_s)^{-1} T_s (T_s T_s^*)^\rho \psi \right\|^2 \\ &\lesssim \alpha_n^{2\rho\wedge 2}. \end{aligned} \quad (55)$$

Therefore, there exists a constant C such that $\rho_{\mathbf{P}}^2(f) < C\alpha_n^{2\rho\wedge 2} =: \delta_n$ for all $f \in \mathcal{F}$.

Combining these two observations, for any $\varepsilon > 0$

$$\begin{aligned} \Pr^* \left(\sup_{f \in \mathcal{F}_n} \left| n^{1/2} (\mathbf{P}_n - \mathbf{P})f \right| > \varepsilon \right) &= \Pr^* \left(\sup_{f \in \mathcal{F}_n: \rho_{\mathbf{P}}(f) < \delta_n} \left| n^{1/2} (\mathbf{P}_n - \mathbf{P})f \right| > \varepsilon \right) \\ &\leq \Pr^* \left(\sup_{f \in \mathcal{H}: \rho_{\mathbf{P}}(f) < \delta_n} \left| n^{1/2} (\mathbf{P}_n - \mathbf{P})f \right| > \varepsilon \right), \end{aligned} \quad (56)$$

where the latter probability tend to zero as $n \rightarrow \infty$ by asymptotic equicontinuity due to the fact that \mathcal{H} is Donsker.

Using above estimates, it is easy to verify that, Assumption 5 ensures that

$$\sum_{j=1}^7 \sqrt{n} \langle R_{j,n}, f_{UZ}(u, \cdot) \rangle = o_p(1), \quad (57)$$

uniformly in u , which establishes the result. \square

Proof of Corollary 1. The process given in Theorem 2 is an empirical process indexed by the

following class of functions

$$\mathcal{F} = \{(v, w) \mapsto \mathbb{1}_{\{v \leq u\}} + v (T(T^*T)^{-1} f_{UZ}(u, \cdot))(w), u \in \mathbf{R}\}, \quad (58)$$

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

$$\mathcal{H} = \{(v, w) \mapsto v (T(T^*T)^{-1} f_{UZ})(u, w), u \in \mathbf{R}\}. \quad (59)$$

Hence by (Van Der Vaart and Wellner, 2000, Theorem 2.10.6) it suffices to show that \mathcal{H} is Donsker. Notice that under Assumption 4, the latter admits the following envelop

$$H(v, w) = v \|(T_s^* T_s)^\rho L^{-s} f_{ZW}(\cdot, w)\| \|\psi\|_{2, \infty}, \quad (60)$$

which is square-integrable under Assumption 3 (ii) and (iv). It follows from the proof of Theorem 2 that \mathcal{H} is Donsker, since by (Engl et al., 1996, Corollary 8.22)

$$\|T(T^*T)^{-1} f_{UZ}(u, \cdot)\|_{\rho(a+s)/2} = \|(T_s^* T_s)^\rho \psi(u, \cdot)\|_{\rho(a+s)/2} \lesssim 1. \quad (61)$$

□

Proof of Theorem 1. Decompose

$$\hat{\varphi}_{\alpha_n} - \varphi = I_n + II_n + III_n, \quad (62)$$

with

$$\begin{aligned} I_n &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{\varphi} - \hat{T} \varphi) \\ II_n &= L^{-s} (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* \hat{T}_s L^s \varphi - L^{-s} (\alpha_n I + T_s^* T_s)^{-1} T_s^* T_s L^s \varphi \\ III_n &= L^{-s} (\alpha_n I + T_s^* T_s)^{-1} T_s^* T_s L^s \varphi - \varphi. \end{aligned} \quad (63)$$

Since $L^s \varphi \in H^{b-s}(\mathbf{R}^p)$, by (Engl et al., 1996, Corollary 8.22), there exists a function $\psi \in L_2(\mathbf{R}^p)$ such that

$$L^s \varphi = (T_s^* T_s)^{\frac{b-s}{2(a+s)}} \psi. \quad (64)$$

Then

$$\begin{aligned}
\mathbb{E}\|I_n\|_s^2 &= \mathbb{E} \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{r} - \hat{T} \varphi) \right\|^2 \leq \frac{1}{\alpha_n^2} \mathbb{E} \left\| L^{-s} \hat{T}^* (\hat{r} - \hat{T} \varphi) \right\|^2 \\
\mathbb{E}\|II_n\|_s^2 &= \mathbb{E} \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} (\hat{T}_s^* \hat{T}_s - T_s^* T_s) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|^2 \\
\mathbb{E}\|III_n\|_s^2 &= \mathbb{E} \left\| \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|^2 \leq \sup_{\lambda} \left| \frac{\alpha_n \lambda^{\frac{b-s}{2(a+s)}}}{\alpha_n + \lambda} \right|^2 \|\psi\|^2 = O\left(\alpha_n^{\frac{b-s}{a+s}}\right)
\end{aligned} \tag{65}$$

The second term is decomposed further as

$$\mathbb{E}\|II_n\|_s^2 \leq 2S_{1n} + 2S_{2n} \tag{66}$$

with

$$\begin{aligned}
S_{1n} &= \mathbb{E} \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} \hat{T}_s^* (\hat{T}_s - T_s) \alpha_n (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|^2 \\
&= O\left(\alpha_n^{-1} \mathbb{E} \left\| \hat{T}_s - T_s \right\|^2 \alpha_n^{\frac{b-s}{a+s}}\right) \\
S_{2n} &= \mathbb{E} \left\| (\alpha_n I + \hat{T}_s^* \hat{T}_s)^{-1} (\hat{T}_s^* - T_s^*) \alpha_n T_s (\alpha_n I + T_s^* T_s)^{-1} L^s \varphi \right\|^2 \\
&= O\left(\alpha_n^{-2} \mathbb{E} \left\| \hat{T}_s^* - T_s^* \right\|^2 \alpha_n^{\frac{b+a}{a+s}}\right)
\end{aligned} \tag{67}$$

□

Proof of Proposition 1. In light of Theorem 1, since operator norm can be bounded by the Hilbert-Schmidt norm, both $\|\hat{T}^* - T^*\|$ and $\|\hat{T} - T\|$ are bounded by $\|\hat{f}_{ZW} - f_{ZW}\|$. Under Assumptions 2

$$\mathbb{E} \left\| \hat{f}_{ZW} - f_{ZW} \right\|^2 = O\left(\frac{1}{nh_n^{p+q}} + h_n^{2t}\right). \tag{68}$$

It remains to study the behavior of

$$\mathbb{E} \left\| L^{-s} \hat{T}^* (\hat{r} - \hat{T} \varphi) \right\|^2. \tag{69}$$

To that end we write

$$L^{-s} \hat{T}^* (\hat{r} - \hat{T} \varphi) = L^{-s} T^* (\hat{r} - \hat{T} \varphi) + L^{-s} (\hat{T}^* - T^*) (\hat{r} - \hat{T} \varphi) \tag{70}$$

and

$$\begin{aligned}
L^{-s} T^* (\hat{r} - \hat{T} \varphi) &= \frac{1}{n} \sum_{i=1}^n \{Y_i - [\varphi * K_z](Z_i)\} \{L^{-s} [f_{ZW} * K_w](z, W_i)\} \\
&\equiv I_n + II_n + III_n,
\end{aligned} \tag{71}$$

where

$$\begin{aligned}
I_n &= \frac{1}{n} \sum_{i=1}^n U_i L^{-s} f_{ZW}(z, W_i) \\
II_n &= \frac{1}{n} \sum_{i=1}^n \{\varphi(Z_i) - [\varphi * K_z](Z_i)\} L^{-s} [f_{ZW} * K_z](z, W_i) \\
III_n &= \frac{1}{n} \sum_{i=1}^n U_i L^{-s} \{[f_{ZW} * K_z](z, W_i) - f_{ZW}(z, W_i)\}
\end{aligned} \tag{72}$$

By $\mathbb{E}[U|W] = 0$, I_n and III_n are sums of i.i.d. zero-mean random variables, whence under Assumptions 3 (ii), (iv), 1, and 2 (i).

$$\begin{aligned}
\mathbb{E}\|I_n\|^2 &= \frac{1}{n} \mathbb{E}\|U_i L^{-s} f_{ZW}(\cdot, W_i)\|^2 \\
&\leq \frac{\bar{\sigma}^2}{n} \mathbb{E}\|f_{ZW}(\cdot, W_i)\|_{-s}^2 \\
&= O\left(\frac{1}{n}\right)
\end{aligned} \tag{73}$$

and

$$\begin{aligned}
\mathbb{E}\|III_n\|^2 &= \frac{1}{n} \mathbb{E}\|U_i L^{-s} \{[f_{ZW} * K_z](z, W_i) - f_{ZW}(z, W_i)\}\|^2 \\
&= O\left(\frac{1}{n} \|f_{ZW} - f_{ZW} * K_z\|^2\right) \\
&= O\left(\frac{h_n^{2t}}{n}\right).
\end{aligned} \tag{74}$$

Similarly, by Cauchy-Schwartz inequality

$$\begin{aligned}
\mathbb{E}\|II_n\|^2 &= \frac{1}{n} \mathbb{E}\|(\varphi(Z_i) - [\varphi * K_z](Z_i)) L^{-s} [f_{ZW} * K_z](z, W_i)\|^2 \\
&\quad + \frac{n-1}{n} \|\mathbb{E}(\varphi(Z_i) - [\varphi * K_z](Z_i)) [f_{ZW} * K_z](\cdot, W_i)\|^2 \\
&= \frac{1}{n} \|\varphi - \varphi * K_z\|^2 \|L^{-s} f_{ZW} * K_z\|^2 + \|\varphi - \varphi * K_z\|^2 \|L^{-s} f_{ZW} * K_z\|^2 \\
&= O\left(\frac{h_n^{2b}}{n} + h_n^{2b}\right).
\end{aligned} \tag{75}$$

Combining all estimates, we obtain

$$\mathbb{E}\left\|L^{-s} T^*(\hat{r} - \hat{T}\varphi)\right\|^2 = O\left(\frac{1}{n} + h_n^{2b}\right). \tag{76}$$

Now decompose

$$\begin{aligned}
L^{-s}(\hat{T}^* - T^*)(\hat{r} - \hat{T}\varphi) &= \int (\hat{r} - \hat{T}\varphi) L^{-s} (\hat{f}_{ZW} - \mathbb{E}\hat{f}_{ZW}) + \int (\hat{r} - \hat{T}\varphi) L^{-s} (\mathbb{E}\hat{f}_{ZW} - f_{ZW}) \\
&=: V_{1,n} + V_{2,n} + B_{1,n} + B_{2,n},
\end{aligned} \tag{77}$$

where the four terms are defined below. Using Assumption 3, Young's inequality, and Assumption 2

$$\begin{aligned}
\mathbb{E} \|B_{1,n}\|^2 &= \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n U_i L^{-s} \left[\left(\mathbb{E} \hat{f}_{ZW} - f_{ZW} \right) * K_w \right] (\cdot, W_i) \right\|^2 \\
&= \frac{1}{n} \mathbb{E} \left\| U_i L^{-s} \left[\left(\mathbb{E} \hat{f}_{ZW} - f_{ZW} \right) * K_w \right] (\cdot, W_i) \right\|^2 \\
&= O \left(\frac{\bar{\sigma}^2}{n} \left\| \mathbb{E} \hat{f}_{ZW} - f_{ZW} \right\|^2 \|K_w\|_1^2 \right) \\
&= O \left(\frac{h_n^{2t}}{n} \right)
\end{aligned} \tag{78}$$

and

$$\begin{aligned}
\mathbb{E} \|B_{2,n}\|^2 &= \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n L^{-s} (\varphi(Z_i) - [\varphi * K_z](Z_i)) \left[\left(\mathbb{E} \hat{f}_{ZW} - f_{ZW} \right) * K_w \right] (\cdot, W_i) \right\|^2 \\
&= \frac{1}{n} \mathbb{E} \left\| L^{-s} (\varphi(Z_i) - [\varphi * K_z](Z_i)) \left[\left(\mathbb{E} \hat{f}_{ZW} - f_{ZW} \right) * K_w \right] (\cdot, W_i) \right\|^2 \\
&\quad + \frac{n-1}{n} \left\| \mathbb{E} \left[L^{-s} (\varphi(Z_i) - [\varphi * K_z](Z_i)) \left[\left(\mathbb{E} \hat{f}_{ZW} - f_{ZW} \right) * K_w \right] (\cdot, W_i) \right] \right\|^2 \\
&= O \left(\|\varphi - \varphi * K_w\|^2 \|\mathbb{E} \hat{f}_{ZW} - f_{ZW}\|^2 \right) \\
&= O \left(h_n^{2b+2t} \right).
\end{aligned} \tag{79}$$

Similarly by Young's inequality

$$\begin{aligned}
\mathbb{E} \|V_{2,n}\|^2 &= \mathbb{E} \left\| \int \mathbb{E} [\hat{r} - \hat{T}\varphi] \left(\hat{f}_{ZW} - \mathbb{E} \hat{f}_{ZW} \right) \right\|^2 \\
&\leq \frac{1}{n} \mathbb{E} \left\| \int \mathbb{E} [\hat{r}(w) - (\hat{T}\varphi)(w)] \left(h_n^{-p-q} K_z \left(h_n^{-1} (Z_i - \cdot) \right) K_w \left(h_n^{-1} (W_i - w) \right) \right) dw \right\|^2 \\
&= \frac{1}{nh_n^p} \mathbb{E} \left\| \left[\mathbb{E} [\hat{r}(w) - (\hat{T}\varphi)(w)] * K_w \right] (W_i) \right\|^2 \|K_z\|^2 \\
&\leq \frac{1}{nh_n^p} \left\| \mathbb{E} [\hat{r} - \hat{T}\varphi] \right\|^2 \|K_w\|_1^2 \|K_z\|^2 \\
&= O \left(\frac{1}{h_n^{p+q}} \|\varphi - \varphi * K_z\|^2 \right) \\
&= O \left(\frac{h_n^{2b}}{nh_n^{p+q}} \right).
\end{aligned} \tag{80}$$

For the last term we put

$$\begin{aligned}
V_{1,n} &= \frac{1}{n} \sum_{i=1}^n U_i \left[\left(\hat{f}_{ZW} - \mathbb{E} \hat{f}_{ZW} \right) * K_w \right] (z, W_i) \\
&= \frac{1}{n^2} \sum_{i,j=1}^n U_i [g_j * K_w](z, W_i),
\end{aligned} \tag{81}$$

where g_j denotes

$$g(Z_j, W_j, z, w) = h_n^{-p-q} K_z(h_n^{-1}(Z_j - z)) K_w(h_n^{-1}(W_j - w)) - h_n^{-p-q} \mathbb{E} [K_z(h_n^{-1}(Z_1 - z)) K_w(h_n^{-1}(W_1 - w))]. \quad (82)$$

Decompose further $V_{1,n} = \xi_n + V_n$ with

$$\begin{aligned} \xi_n &= \frac{1}{n^2} \sum_{i=1}^n U_i [g_i * K_w](z, W_i) \\ V_n &= \frac{n-1}{n} \frac{2}{n(n-1)} \sum_{i < j} \frac{1}{2} \{U_i [g_j * K_w](z, W_i) + U_j [\eta_{n,i} * K_w](z, W_j)\}. \end{aligned} \quad (83)$$

Then

$$\mathbb{E} \|\xi_n\|^2 = \frac{1}{n^3} \mathbb{E} \|U_2 [g_2 * K_w](\cdot, W_2)\|^2 + \frac{n-1}{n^3} \|\mathbb{E} [U_2 [g_2 * K_w](\cdot, W_2)]\|^2 = O\left(\frac{1}{n^2 h_n^{p+q}}\right). \quad (84)$$

For V_n by the moment inequality in (Korolyuk and Borovskich, 1994, Theorem 2.1.6.) there exists some constant C such that

$$\mathbb{E} \|V_n\|^2 \leq \frac{C}{n(n-1)} \mathbb{E} \|U_1 [g_2 * K_w](\cdot, W_1) + U_2 [\eta_{n,1} * K_w](\cdot, W_2)\|^2 = O\left(\frac{1}{n^2 h_n^{p+q}}\right). \quad (85)$$

Combining all estimates, we establish

$$\mathbb{E} \left\| L^{-s} (\hat{T}^* - T^*) (\hat{r} - \hat{T}\varphi) \right\|^2 = O\left(\left(\frac{1}{n} + h_n^{2b}\right) \left(\frac{1}{n h_n^{p+q}} + h_n^{2t}\right)\right). \quad (86)$$

□

Lemma 2. *Suppose that 3, 4, 5, 1, and 2 are satisfied. Then*

$$\sup_{u \in \mathbf{R}, w \in \mathbf{R}^q} \left| \hat{F}_{\hat{U}W}(u, w) - \hat{F}_{UW}(u, w) - \Pr\left(U \leq u + \hat{\Delta}(Z), W \leq w \mid \mathcal{X}\right) + F_{UW}(u, w) \right| = o_p\left(n^{-1/2}\right) \quad (87)$$

and

$$\sup_{v \in \mathbf{R}, w \in \mathbf{R}^q} \left| \hat{F}_{\hat{U}}(u) \hat{F}_W(w) - \hat{F}_U(u) \hat{F}_W(w) - \Pr\left(U \leq u + \hat{\Delta}(Z) \mid \mathcal{X}\right) F_W(w) + F_U(u) F_W(w) \right| = o_p\left(n^{-1/2}\right), \quad (88)$$

where $\hat{\Delta}(Z) = \hat{\varphi}(Z) - \varphi(Z)$.

The proof is similar to the proof of Lemma 1 and so we omit it.

Proof of Theorem 3. By Lemma 2, uniformly in (u, w)

$$\mathbb{G}_n(u, w) = T_{1n}(u, w) + T_{2n}(u, w) - T_{3n}(u, w) + o_p(1), \quad (89)$$

where

$$\begin{aligned} T_{1n}(u, w) &= \sqrt{n} \left(\hat{F}_{UW}(u, w) - \hat{F}_U(u) \hat{F}_W(w) \right), \\ T_{2n}(u, w) &= \sqrt{n} \left(\Pr \left(U \leq u + \hat{\Delta}(Z), W \leq w \mid \mathcal{X} \right) - F_{UW}(u, w) \right), \\ T_{3n}(u, w) &= \sqrt{n} \left(\Pr \left(U \leq u + \hat{\Delta}(Z) \mid \mathcal{X} \right) - F_U(u) \right) F_W(w). \end{aligned} \quad (90)$$

Now

$$\begin{aligned} T_{1n}(u, w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{1}_{\{U_i \leq u, W_i \leq w\}} - \mathbb{1}_{\{U_i \leq u\}} F_W(w) - \mathbb{1}_{\{W_i \leq w\}} F_U(u) + F_U(u) F_W(w) \right\} \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{1}_{\{W_i \leq w\}} - F_W(w) \right\} \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{1}_{\{U_i \leq u\}} - F_U(u) \right\}. \end{aligned} \quad (91)$$

By the Donsker CLT, the term in the second line is $O_p(1)O_p(n^{-1/2}) = o_p(1)$ as $n \rightarrow \infty$ uniformly over $(u, w) \in \mathbf{R} \times \mathbf{R}^q$, while the term in the first line will contribute to the asymptotic distribution.

Consider now T_{2n} . Under Assumption 4, using the Taylor expansion, for some ξ_z between u and $u + \hat{\Delta}(z)$

$$\begin{aligned} T_{2n}(u, w) &= \sqrt{n} \int_{\mathbf{R}^p} \int_{\{\tilde{w} \leq w\}} \left\{ \int_{-\infty}^{u + \hat{\Delta}(z)} f_{UZW}(\tilde{u}, z, \tilde{w}) d\tilde{u} - \int_{-\infty}^u f_{UZW}(\tilde{u}, z, \tilde{w}) d\tilde{u} \right\} d\tilde{w} dz \\ &= \sqrt{n} \int_{\mathbf{R}^p} \int_{\{\tilde{w} \leq w\}} \left\{ f_{UZW}(u, z, \tilde{w}) \hat{\Delta}(z) + \frac{1}{2} \partial_u f_{UZW}(\xi_z, z, \tilde{w}) \hat{\Delta}^2(z) \right\} d\tilde{w} dz \\ &=: S_{1n}(u, w) + S_{2n}(u, w). \end{aligned} \quad (92)$$

Under Assumptions 3 (ii) and 4 (ii)

$$\|S_{2n}\| \lesssim \sqrt{n} \|\hat{\varphi} - \varphi\|^2, \quad (93)$$

which is $o_p(1)$ under Assumption 5 by Proposition 1.

In a similar way, we obtain

$$\begin{aligned} T_{3n}(u, w) &= \sqrt{n} \int_{\mathbf{R}^p} \hat{\Delta}(z) f_{UZ}(u, z) dz F_W(w) + \sqrt{n} \int_{\mathbf{R}^p} \hat{\Delta}^2(z) \frac{1}{2} \partial_u f_{UZ}(\xi_z, z) dz F_W(w) \\ &=: S_{3n}(u, w) + o_p(1). \end{aligned} \quad (94)$$

Therefore, uniformly in (u, w)

$$\begin{aligned}
T_{2n}(u, w) - T_{3n}(u, w) &= \sqrt{n} \int_{\mathbf{R}^p} (\hat{\varphi}(z) - \varphi(z)) \left\{ \int_{\{\tilde{w} \leq w\}} f_{UZW}(u, z, \tilde{w}) d\tilde{w} - f_{UZ}(u, z) F_W(w) \right\} dz + o_p(1) \\
&= \sqrt{n} \int_{\mathbf{R}^p} (\hat{\varphi}(z) - \varphi(z)) g(z, u, w) dz + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i (T(T^*T)^{-1} g(\cdot, u, w)) (W_i) + o_p(1),
\end{aligned} \tag{95}$$

where the last line follows by the same argument as in the proof of Theorem 2 under additional Assumption 4 (iv). \square

References

- Michael G Akritas and Ingrid Van Keilegom. Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, pages 549–567, 2001.
- R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.
- Richard Blundell and Joel L Horowitz. A non-parametric test of exogeneity. *The Review of Economic Studies*, 74(4):1035–1058, 2007.
- Christoph Breunig. Specification testing in nonparametric instrumental quantile regression. 2013.
- Christoph Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.
- Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Asymptotic normal inference in linear inverse problems. *J. S. Racine, A. Ullah and L. Su, eds, ‘Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics’*, 2013.
- Xiaohong Chen and Timothy Christensen. Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. 2015.
- Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005.
- Victor Chernozhukov, Guido W Imbens, and Whitney K Newey. Instrumental variable estimation of nonseparable models. *Journal of Econometrics*, 139(1):4–14, 2007.

- S. Darolles, Y. Fan, J.P. Florens, and E. Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.
- Fabian Dunker, Jean-Pierre Florens, Thorsten Hohage, Jan Johannes, and Enno Mammen. Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. *Journal of Econometrics*, 178:444–455, 2014.
- James Durbin. Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, pages 279–290, 1973.
- Xavier D’Haultfoeuille and Philippe Février. Identification of nonseparable triangular models with discrete instruments. *Econometrica*, 83(3):1199–1210, 2015.
- John HJ Einmahl and Ingrid Van Keilegom. Specification tests in nonparametric regression. *Journal of Econometrics*, 143(1):88–102, 2008.
- Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Kluwer Academic Pub, 1996.
- Frédérique Fève, Jean-Pierre Florens, and Ingrid van Keilegom. Estimation of conditional ranks and tests of exogeneity in nonparametric nonseparable models. 2013.
- Jean-Pierre Florens. Inverse problems and structural econometrics. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, page 284. Cambridge University Press, 2003.
- Jean-Pierre Florens, Jan Johannes, and Sébastien Van Belleghem. Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(03):472–496, 2011.
- Patrick Gagliardini and Olivier Scaillet. Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167(1):61–75, 2012.
- Peter Hall and Joel L Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.
- Stefan Hoderlein, Liangjun Su, Halbert L White Jr, and Thomas Tao Yang. Testing for monotonicity in unobservables under unconfoundedness. *Available at SSRN 2448681*, 2014.
- Joel L Horowitz. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74(2):521–538, 2006.

- Joel L Horowitz. Specification testing in nonparametric instrumental variable estimation. *Journal of Econometrics*, 167(2):383–396, 2012.
- Joel L Horowitz and Sokbae Lee. Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75(4):1191–1208, 2007.
- Guido W Imbens. Nonadditive models with endogenous regressors. *Econometric Society Monographs*, 43:17, 2007.
- V. S. Korolyuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273. Springer Science & Business Media, 1994.
- Selim Grigor’evich Krein and Yu I Petunin. Scales of banach spaces. *Russian Mathematical Surveys*, 21(2):85–159, 1966.
- Arthur Lewbel, Xun Lu, and Liangjun Su. Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, 184(1):81–96, 2015.
- RM Loynes. The empirical distribution function of residuals from generalised regression. *The Annals of Statistics*, pages 285–298, 1980.
- Xun Lu and Halbert White. Testing for separability in structural equations. *Journal of Econometrics*, 182(1):14–26, 2014.
- Enno Mammen et al. Empirical process of residuals for high-dimensional linear models. *The Annals of Statistics*, 24(1):307–335, 1996.
- Andreas Neubauer. When do sobolev spaces form a hilbert scale? *Proceedings of the American Mathematical Society*, 103(2):557–562, 1988.
- Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.
- Richard Nickl and Benedikt M Pötscher. 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, 2007.
- Alexander Torgovitsky. Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197, 2015.

Alexander Torgovitsky. Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics*, 2017.

Aad W Van Der Vaart and Jon A Wellner. *Weak Convergence*. Springer, 2000.

Ingrid Van Keilegom, Wenceslao González Manteiga, and César Sánchez Sellero. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, 17(2): 401–415, 2008.