

I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

2012/36

Estimation of conditional ranks and tests of exogeneity in
nonparametric nonseparable models

FÈVE, F., FLORENS, J-P. and I. VAN KEILEGOM

Estimation of conditional ranks and tests of exogeneity in nonparametric nonseparable models ‡

Frédérique FÈVE § Jean-Pierre FLORENS ** Ingrid VAN KEILEGOM §§

December 5, 2012

Abstract

Consider a nonparametric nonseparable regression model $Y = \varphi(Z, U)$, where $\varphi(Z, U)$ is increasing in U and $U \sim U[0, 1]$. We suppose that there exists an instrument W that is independent of U . The observable random variables are Y , Z and W , all one-dimensional. The purpose of this paper is twofold. First, we study the asymptotic properties of a kernel estimator of the distribution of $V = F_{Y|Z}(Y|Z)$, which equals U when Z is exogenous. We show that this estimator converges to the uniform distribution at faster rate than the parametric $n^{-1/2}$ -rate. Next, we construct test statistics for the hypothesis that Z is exogenous. The test statistics are based on the observation that Z is exogenous if and only if V is independent of W , and hence they do not require the estimation of the function φ . The asymptotic properties of the proposed tests are proved, and a bootstrap approximation of the critical values of the tests is shown to work for finite samples via simulations. An empirical example using the U.K. Family Expenditure Survey is also given.

Key Words: Endogeneity; Instrumental variable; Nonseparability; Nonparametric regression.

‡The authors like to thank R. Blundell, J.L. Horowitz and L. Simar for helpful discussions.

§Toulouse School of Economics. Email address: frederique.feve@tse-fr.eu.

**Toulouse School of Economics. Email address: jean-pierre.florens@tse-fr.eu.

§§Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium. E-mail address: ingrid.vankeilegom@uclouvain.be. This research was supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

1 Introduction

Nonseparable models are becoming increasingly popular in econometrics, because of their easy interpretability and their flexible structure. For some of the more recent references on nonparametric nonseparable models with endogeneity we refer to Matzkin (2003), Chernozhukov and Hansen (2005), Hoderlein and Mammen (2007), Horowitz and Lee (2007), Imbens and Newey (2009), Chernozhukov, Scaillet and Gagliardini (2011), Dunker, Hohage, Florens, Johannes and Mammen (2012), and Chen and Pouzo (2012). An important problem in these models is the problem of testing whether the independent variables in these models are exogenous. See Blundell and Horowitz (2007) among others for a detailed discussion on the importance of this kind of tests in general structural models. This problem has so far not been studied in the literature on nonparametric nonseparable models.

An economic variable is exogenous if its generating process does not contain any information on the parameter (possibly functional) of interest (see Engle, Hendry and Richard, 1983, and Florens and Mouchart, 1985). In many models this property is equivalent to an independence condition between this variable and an ‘error’. In a linear regression model $Y = \beta^t Z + U$, exogeneity is equivalent to $E(UZ) = 0$. In a general separable model $Y = m(Z) + U$ exogeneity means that $E(U|Z) = 0$, and m is then the conditional mean function. In a nonseparable model $Y = \varphi(Z, U)$ the exogeneity of Z is equivalent to the full independence between Z and U , and φ is in that case the conditional quantile function if U is normalized by $U \sim U[0, 1]$.

In order to test the assumption of exogeneity, it is necessary to characterize the endogeneity property. Different methods may be followed and we choose the approach based on instrumental variables (IV). Let W be the instruments. The linear model is then characterized by $E(UW) = 0$, the separable model by $E(U|W) = 0$ and the nonseparable model by the independence between U and W . A test of exogeneity may be conducted in several ways. Under the identification condition of the IV model we may estimate β , m or φ under the assumptions $E(UW) = 0$, $E(U|W) = 0$, or U and W independent, respectively, and we may compare this estimator to an ordinary least squares estimator, or a general regression or quantile estimator. An alternative way is to estimate the residuals under the exogeneity assumption and to verify whether the IV condition is satisfied. This second method avoids the IV estimation, which is known to be difficult in a nonparametric framework (see e.g. Johannes, Van Bellegem and Vanhems, 2011, and Florens, Johannes and Van Bellegem, 2011, among many others).

In the linear case exogeneity tests have been introduced by Wu (1973) and Hausman

(1978) a long time ago. More recently, Blundell and Horowitz (2007) have proposed an exogeneity test for nonparametric separable models $Y = \varphi(Z) + U$ with $E(U|W) = 0$ for some instrument W . Their test is based on an estimator $\widehat{V} = Y - \widehat{E}(Y|Z)$ of $V = Y - E(Y|Z)$. They test whether $E(V|W) = 0$ or more precisely whether $E(E(V|W)|Z) = 0$. See also Blundell, Chen and Kristensen (2007) and Sokullu (2011) for related testing procedures. To the best of our knowledge no exogeneity tests exist in the case of nonparametric nonseparable models.

In this paper we fill this gap in the literature by proposing a novel idea to test the independence between U and Z under the nonseparable model $Y = \varphi(Z, U)$. The test statistic is based on an estimator of the conditional distribution $V = F_{Y|Z}(Y|Z)$, say $\widehat{V} = \widehat{F}_{Y|Z}(Y|Z)$. We show that the independence between U and Z is under certain identifiability conditions equivalent to the independence between V and W , and we use Kolmogorov-Smirnov and Crámer-von Mises type statistics to test the independence between V and W , based on the observable quantities \widehat{V} and W .

The problem of estimating the distribution of the ‘error’ V is related to the problem of doing inference for the error distribution in a nonparametric separable regression model. In recent years, this problem has been extensively studied in the statistics and econometrics literature. See e.g. Akritas and Van Keilegom (2001), Lewbel and Linton (2002), Chen, Dahl and Khan (2005), Dette, Neumeyer and Van Keilegom (2007), Einmahl and Van Keilegom (2008), Müller, Schick and Wefelmeyer (2009, 2012), Neumeyer (2009), Neumeyer and Van Keilegom (2010) and the references therein.

The paper is organized as follows. In the next section we give the precise definition of the test statistics, and we also prove that exogeneity is characterized by the independence between V and W , which is a crucial property for our test. In Section 3 we consider a by-product of our test, which is the estimation of the distribution of V and of (V, W) , which leads to peculiar asymptotic properties. The limiting distribution and a bootstrap approximation of the proposed test statistics are studied in Section 4. The behavior of the proposed tests for small samples is studied in Section 5 via simulations and via an empirical example using the U.K. Family Expenditure Survey. In Section 6 we summarize some general conclusions. Finally, Appendix A studies the identification of our model, whereas Appendix B contains the proofs of the main asymptotic results.

2 The test statistics

Consider the nonseparable model

$$Y = \varphi(Z, U), \tag{2.1}$$

where $\varphi(Z, U)$ is increasing in U for each Z , and $U \sim U[0, 1]$. We suppose that there is an instrument W that is independent of U . The observable random variables are Y , Z and W , all one-dimensional. Our goal is to test whether Z is exogenous. More precisely, we may give the following definition.

Definition 2.1. *The variable Z is q -exogenous if $P(U \leq q|Z = z) = q$ for a given $0 < q < 1$, and Z is (completely) exogenous if this property holds true for any q in $[0, 1]$.*

The q -exogeneity of Z is sufficient to identify $\varphi(Z, q)$, but the complete exogeneity is needed to identify φ as a function of Z and U jointly. We analyze in this paper the complete exogeneity, which is equivalent to Z being independent of U . For testing the hypothesis of q -exogeneity we refer to Fu (2010).

We assume throughout this paper that model (2.1) is globally identified. This means that φ is unique in the sense that if $Y = \varphi_1(Z, U_1) = \varphi_2(Z, U_2)$ and U_1 and U_2 are both independent of W and uniform on $[0, 1]$, then necessarily $U_1 = U_2$ a.s. and $\varphi_1 = \varphi_2$. See Appendix A for sufficient conditions for global identification in nonseparable models.

Define

$$V = F_{Y|Z}(Y|Z),$$

where $F_{Y|Z}(y|z) = P(Y \leq y|Z = z)$. It is clear that if Z is exogenous, then $U = V$. The following proposition is crucial for the construction of our test statistic.

Proposition 2.1. *Assume that model (2.1) is globally identified. Then, U and Z are independent if and only if V and W are independent.*

Proof. If U and Z are independent, then $U = V$, and hence V is independent of W . On the other hand, if V and W are independent, then by noting that $Y = \tilde{\varphi}(Z, V)$, with $\tilde{\varphi}(z, v) = F_{Y|Z}^{-1}(v|z)$, we can conclude that $U = V$ a.s. by the global identification of model (2.1). Hence, since V is by construction independent of Z , we also have that U is independent of Z . \square

Our test statistics will be based on an estimator of $P(V \leq v, W \leq w) - P(V \leq v)P(W \leq w)$, which under the conditions of Proposition 2.1 will be close to zero under the null hypothesis

$$H_0 : Z \text{ is exogenous.}$$

Let $\widehat{F}_W(w) = n^{-1} \sum_{i=1}^n I(W_i \leq w)$ be the empirical distribution of $F_W(w) = P(W \leq w)$, and define for any $i = 1, \dots, n$,

$$\widehat{V}_i = \widehat{F}_{Y|Z}(Y_i|Z_i),$$

where

$$\widehat{F}_{Y|Z}(y|z) = \sum_{j=1}^n \frac{k_h(z - Z_j)}{\sum_{k=1}^n k_h(z - Z_k)} L_g(y - Y_j), \quad (2.2)$$

$k_h(\cdot) = k(\cdot/h)/h$, $L_g(u) = \int_{-\infty}^u \ell_g(t) dt$, $\ell_g(\cdot) = \ell(\cdot/g)/g$, k and ℓ are probability density functions, and h and g are appropriate bandwidths tending to zero when n tends to infinity. Then, we estimate the distribution of V by

$$\widehat{F}_{\widehat{V}}(v) = n^{-1} \sum_{i=1}^n I(\widehat{V}_i \leq v),$$

and the bivariate distribution of V and W is estimated by

$$\widehat{F}_{\widehat{V},W}(v, w) = n^{-1} \sum_{i=1}^n I(\widehat{V}_i \leq v, W_i \leq w).$$

The proposed test statistics for the null hypothesis H_0 are now given by the Kolmogorov-Smirnov statistic

$$T_{n,KS} = n^{1/2} \sup_{v \in I, w \in R_W} \left| \widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v) \widehat{F}_W(w) \right|,$$

and the Crámer-von Mises statistic

$$T_{n,CM} = n \int \int_{I \times R_W} \left[\widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v) \widehat{F}_W(w) \right]^2 d\widehat{F}_{\widehat{V},W}(v, w).$$

Here, $I \subset [\delta, 1 - \delta]$ for some small $\delta > 0$ and R_W is the support of W .

3 Asymptotic properties of $\widehat{F}_{\widehat{V}}$ and $\widehat{F}_{\widehat{V},W}$

The estimators $\widehat{F}_{\widehat{V}}$ and $\widehat{F}_{\widehat{V},W}$ are interesting on its own and require separate attention, given that their asymptotic properties are quite peculiar. In fact, as we will see below, the

estimator $\widehat{F}_{\widehat{V}}$ converges faster to the uniform distribution than the empirical distribution based on the true (but unknown) errors V_i , whereas the estimator $\widehat{F}_{\widehat{V},W}$ converges at the usual $n^{-1/2}$ rate. The conditions under which the results below are valid, are collected in Appendix B.

3.1 Uniform rate of convergence

Theorem 3.1. *Assume (A1)-(A5). Then,*

$$\sup_{v \in I} |\widehat{F}_{\widehat{V}}(v) - v| = o_P(n^{-1/2}),$$

and

$$\sup_{v \in I, w \in R_W} |\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w)| = O_P(n^{-1/2}).$$

While this result might seem surprising at first sight (especially the rate of $\widehat{F}_{\widehat{V}}(v) - v$, which is faster than parametric), there is an intuitive interpretation of this result. Consider the case where the variable Z is absent, i.e. $V = F_Y(Y)$ and $\widehat{V} = \widehat{F}_Y(Y)$, where $\widehat{F}_Y(\cdot)$ is the empirical distribution of the Y_i 's. Then, $\widehat{F}_{\widehat{V}}(v) - v$ reduces to $n^{-1} \sum_{i=1}^n I(i/n \leq v) - v$, and this is $O_P(n^{-1})$ uniformly in v . On the other hand, $\widehat{F}_{\widehat{V},W}(v, w)$ is in that case equal to $n^{-1} \sum_{i=1}^n I(i/n \leq v) I(W_i \leq w)$, and the variance of this estimator equals

$$n^{-2} \sum_{i=1}^n I(i/n \leq v) F_W(w) (1 - F_W(w)) = n^{-1} v F_W(w) (1 - F_W(w)) (1 + o(1)).$$

Hence, $\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w) = O_P(n^{-1/2})$. This shows that the result in Theorem 3.1 is in line with what we have in the absence of Z .

3.2 Limiting distribution

For obtaining the limiting distribution we need to slightly adapt the definition of the estimator $\widehat{F}_{\widehat{V}}(v)$. Define for a bandwidth b and a kernel function m ,

$$\widehat{F}_{\widehat{V},b}(v) = n^{-1} \sum_{i=1}^n M_b(v - \widehat{V}_i),$$

where $M_b(u) = \int_{-\infty}^u m_b(t) dt$ and $m_b(\cdot) = m(\cdot/b)/b$. This estimator is a smoothed version of the estimator $\widehat{F}_{\widehat{V}}(v)$, and is close to the latter when b is close to zero. This smoothed

version does not only allow us to use a Taylor expansion, but the smoothing parameter also determines the rate of convergence of the estimator, whereas in standard problems smoothing of a distribution function has no influence on the rate of the estimator. This shows again that the estimation of the distribution of V leads to non-standard and unusual asymptotic rates of convergence.

Theorem 3.2. *Assume (A1)-(A5). Then, under H_0 , for any $0 < v < 1$ and $w \in R_W$,*

$$n(hb)^{1/2}[\widehat{F}_{\widehat{V},b}(v) - v] \xrightarrow{d} N(0, \sigma_V^2(v)),$$

and

$$n^{1/2}[\widehat{F}_{\widehat{V},W}(v, w) - vF_W(w)] \xrightarrow{d} N(0, \sigma_{V,W}^2(v, w)),$$

where

$$\sigma_V^2(v) = v(1-v) \int k^2(t)dt \int m^2(t)dt,$$

and

$$\begin{aligned} \sigma_{V,W}^2(v, w) = & vF_W(w)(1-vF_W(w)) + v(1-v) \int F_{W|V,Z}^2(w|v, z)f_Z(z)dz \\ & -2(1-v) \int F_{W|V,Z}(w|v, z) \int_0^v F_{W|V,Z}(w|u, z)du f_Z(z)dz. \end{aligned}$$

Remark 3.1.

- (a) The proof of the first part of Theorem 3.2 above shows that $\widehat{F}_{\widehat{V},b}(v) - v$ is a U -statistic whose kernel depends on n . Moreover, the U -statistic is degenerate (i.e. its conditional expectations are zero). This explains (in part) the rather unusual rate of convergence of the estimator $\widehat{F}_{\widehat{V},b}(v)$. Also note that the asymptotic variance of $\widehat{F}_{\widehat{V},b}(v)$ contains the factor $v(1-v)$, which is the asymptotic variance of the empirical distribution function.
- (b) Note that the process $n^{1/2}[\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w)]$ ($v \in I, w \in R_W$) is easily seen to be tight (the proof is similar to that of Theorem 4.2 below). Hence, the above asymptotic normality can be extended to the weak convergence of the process. On the other hand, the process $n(hb)^{1/2}[\widehat{F}_{\widehat{V},b}(v) - v]$ ($v \in I$) appears not to be tight. This can be explained by the non-tightness of the process $n^{-1/2} \sum_{i=1}^n m_b(v - V_i)$ ($v \in I$), on which the process $n(hb)^{1/2}[\widehat{F}_{\widehat{V},b}(v) - v]$ ($v \in I$) is based (see the proof in Appendix B for more details).

- (c) The reason for working with an estimator $\widehat{F}_{Y|Z}(y|z)$ that is smooth in y comes from the proof of Theorem 3.1. In all other proofs in this paper the smoothed indicator $L_g(y - Y_i)$ in (2.2) could be replaced by the indicator $I(Y_i \leq y)$, avoiding so the choice of the smoothing parameter g .
- (d) Although we have chosen to work with bandwidths h and b in Theorem 3.2 that lead to asymptotically unbiased estimators, it is easily seen from the proof of Theorem 3.2 that the asymptotic bias is of order $O(h^2) + O(b^2)$ (assuming that $g = 0$, which is possible thanks to the previous remark). However, from assumption (A1) in Appendix B we know that $b = o(h)$ and hence the bias is of order $O(h^2)$. It is interesting to derive the order of h and b for which the asymptotic variance and the squared asymptotic bias are in balance, i.e. for which $n^2 h^5 b$ is constant. Let $h \sim n^{-\alpha}$ and $b \sim n^{-\beta}$. Then, we have that $\alpha = (2 - \beta)/5$. The conditions on h and b in assumption (A1) learn that α and β should then be chosen such that $3/10 < \alpha < 1/3$ and $1/3 < \beta < 1/2$. Moreover, the order of the variance is between $n^{-4/3}$ and $n^{-6/5}$, so smaller than the parametric order n^{-1} .

4 Distribution of test statistics

4.1 Limiting distribution

Theorem 4.1. *Assume (A1)-(A5). Then, under H_0 ,*

$$\begin{aligned} & \widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v)\widehat{F}_W(w) \\ &= n^{-1} \sum_{i=1}^n \left\{ I(V_i \leq v) - v \right\} \left\{ I(W_i \leq w) - F_{W|V,Z}(w|v, Z_i) \right\} + o_P(n^{-1/2}), \end{aligned}$$

uniformly in $(v, w) \in I \times R_W$.

Theorem 4.2. *Assume (A1)-(A5). Then, under H_0 , the process $n^{1/2}(\widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v)\widehat{F}_W(w))$ ($v \in I, w \in R_W$) converges weakly to a Gaussian process $G(v, w)$ with mean zero and covariance given by*

$$\begin{aligned} & \text{Cov}(G(v_1, w_1), G(v_2, w_2)) \\ &= \left\{ v_1 \wedge v_2 - v_1 v_2 \right\} \left\{ F_W(w_1 \wedge w_2) - \int \left[F_{W|V,Z}(w_1|v_1, z) F_{W|Z}(w_2|z) \right. \right. \\ & \quad \left. \left. + F_{W|V,Z}(w_2|v_2, z) F_{W|Z}(w_1|z) - F_{W|V,Z}(w_1|v_1, z) F_{W|V,Z}(w_2|v_2, z) \right] f_Z(z) dz \right\}. \end{aligned}$$

We are now ready to prove the following result concerning the asymptotic limit of our test statistics $T_{n,KS}$ and $T_{n,CM}$.

Corollary 4.1. *Assume (A1)-(A5). Then, under H_0 ,*

$$T_{n,KS} \xrightarrow{d} \sup_{v \in I, w \in R_W} |G(v, w)|$$

and

$$T_{n,CM} \xrightarrow{d} \int \int_{I \times R_W} G^2(v, w) dv dF_W(w).$$

Note that contrary to the results of Theorem 3.2, the choice of the bandwidths h and b is more standard for the above asymptotic results. This is because we now have the usual square root n rate of convergence. Also note that the bandwidth g can be taken equal to zero (cf. Remark 3.1(c)).

4.2 Bootstrap approximation

Although the limiting distribution of $T_{n,KS}$ and $T_{n,CM}$ is explicit and in principle estimable in practice, we prefer to use a bootstrap approximation, since the process $G(v, w)$ has a complicated covariance structure, and since the convergence to the normal limit is rather slow. The bootstrap procedure is defined as follows.

- (1) For $i = 1, \dots, n$, let $Z_i^* = Z_i$ and $W_i^* = W_i$.
- (2) Let V_1^*, \dots, V_n^* be i.i.d. variables drawn randomly from $\widehat{F}_{\widehat{V}}$.
- (3) For $i = 1, \dots, n$, let $Y_i^* = \widehat{F}_{Y|Z}^{-1}(V_i^* | Z_i^*)$.
- (4) For $i = 1, \dots, n$, define $\widehat{V}_i^* = \widehat{F}_{Y^*|Z^*}(Y_i^* | Z_i^*)$, where $\widehat{F}_{Y^*|Z^*}$ is defined as in (2.2), except that we use the resample (Y_j^*, Z_j^*) ($j = 1, \dots, n$) instead of the original sample, and we take $g = 0$.
- (5) Calculate $\widehat{F}_{\widehat{V}^*, W^*}(v, w) - \widehat{F}_{\widehat{V}^*}(v) \widehat{F}_{W^*}(w)$ for all v and w , based on which the bootstrapped test statistics $T_{n,KS}^*$ and $T_{n,CM}^*$ are then obtained.

By repeating this procedure a large number of times (say B times) we get B values of the bootstrapped statistics $T_{n,KS}^*$ and $T_{n,CM}^*$. The order statistics of order $1 - \alpha$ of these B values then give us an approximation of the critical values of the tests.

For the theoretical justification of this bootstrap approximation we believe that an i.i.d. expansion similar to the one in Theorem 4.1 holds true for the bootstrapped process, from which the consistency of the bootstrap will follow. The detailed verification of this claim is however beyond the scope of this paper. On the other hand, in the next section we show by means of simulations that the proposed bootstrap method yields a good approximation of the true unknown distribution of $T_{n,KS}$ and $T_{n,CM}$ for small samples.

5 Finite sample study

5.1 Monte Carlo simulations and bootstrap

Let us consider the following model:

$$\begin{cases} Y = \beta Z^2 + \gamma U \\ Z = aU + bW + c\varepsilon, \end{cases}$$

where $W \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, $U \sim U[0, 1]$ and these three variables are mutually independent. We take $\beta = 0.01$, $\gamma = 1$, $b = 2$ and $c = 2$. The parameter a controls the endogeneity of Z : Z is exogenous when $a = 0$, and the dependence between Z and U is a function of a . We simulate this model for two sample sizes : $n = 200$ and $n = 500$. The number of Monte Carlo replications is $M = 200$.

Figure 1 shows the distribution of the Kolmogorov-Smirnov test statistic $T_{n,KS}$ under the null ($a = 0$) and under two alternatives ($a = 2$ and $a = 5$) for two sample sizes ($n = 200$ at the left and $n = 500$ at the right). The figure shows how well the test statistic is able to discriminate the null from the alternative hypothesis. In particular we see that for $n = 500$, the distribution of $T_{n,KS}$ for $a = 0$ and for $a = 5$ are almost completely separated. Figure 2 gives the power of the test for the two sample sizes. We see that the power increases when the sample size increases and when a becomes larger in absolute value (so when we get further away from the null hypothesis).

In this simulation we also investigate how well the bootstrap procedure proposed in Section 4.2 works for approximating the distribution of the test statistic. For this, we draw $B = 200$ bootstrap samples under the assumption of exogeneity from each of the $M = 200$ generated samples. From these 200 bootstrap replications we may compute the bootstrap P -value for each sample. This bootstrap P -value can then be compared to the true P -value, which is obtained using the distribution of the test statistic previously computed via Monte Carlo simulation.

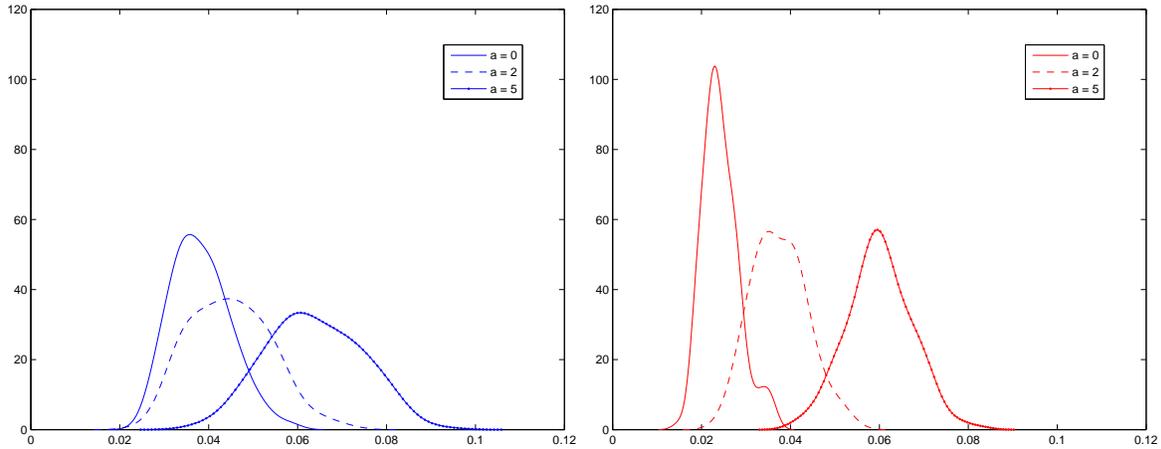


Figure 1: *The distribution of $T_{n,KS}$ under the null and the alternative hypothesis. The left panel shows the results for $n = 200$, the right panel for $n = 500$.*

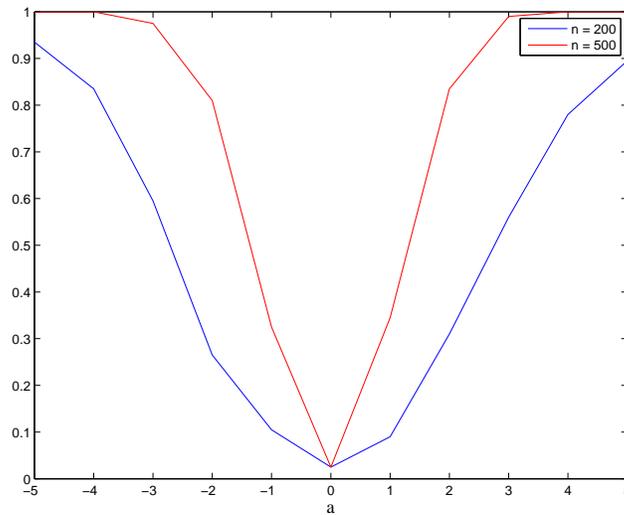


Figure 2: *Power of the test statistic $T_{n,KS}$ for $n = 200$ and $n = 500$.*

Under the null hypothesis ($a = 0$) the true P -values are obviously distributed according to a $U[0, 1]$ distribution. Figure 3 represents the pairs of P -values (‘true’ and ‘bootstrap’) for each sample size ($n = 200$ at the left and $n = 500$ at the right). We see that these P -values are in general close to the bisector in the unit square. This shows that under the null hypothesis the true P -values obtained from Monte Carlo are well approximated by using our bootstrap procedure.

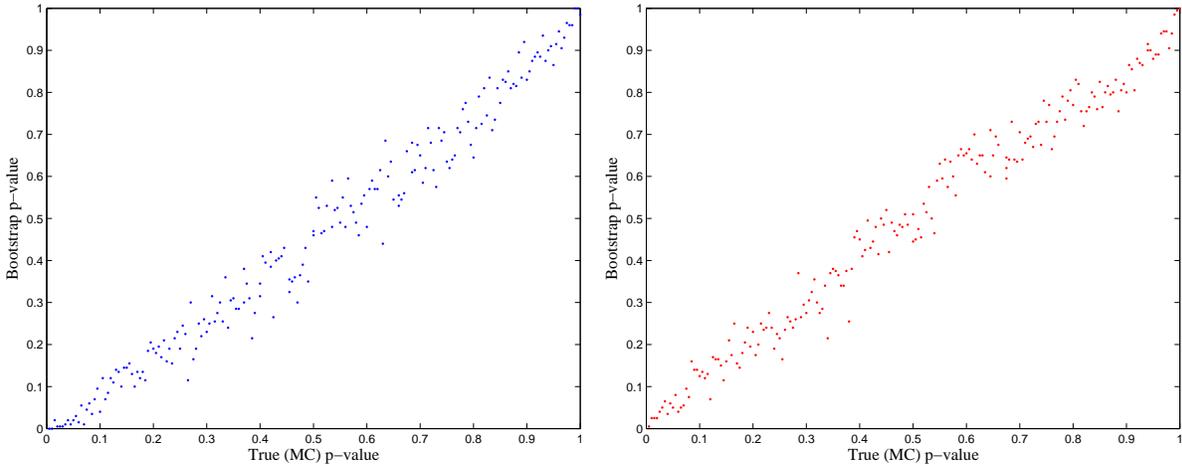


Figure 3: *Monte Carlo and bootstrap P-values under the null hypothesis ($a = 0$) for $n = 200$ (left) and $n = 500$ (right).*

The same exercise is repeated under the alternative for $a = 2$. In order to compare the true and the bootstrap P -values, we calculate the proportion of P -values that are smaller and larger than 0.05 for $n = 200$ and $n = 500$. Table 1 shows that for $n = 500$ the P -values obtained from the Monte Carlo approximation and from the bootstrap approximation are both smaller than 5% in 80.5% of the cases and are both larger than 5% for 14.5% of the generated samples. Moreover, if a test at 5% is performed, the two approximations coincide in 95% of the cases. This shows that also under the alternative, the bootstrap approximation is very accurate.

		Monte Carlo			
		$n = 200$		$n = 500$	
		≤ 0.05	> 0.05	≤ 0.05	> 0.05
Bootstrap	≤ 0.05	0.250	0.070	0.805	0.010
	> 0.05	0	0.680	0.040	0.145

Table 1: *Comparison between Monte Carlo and bootstrap P-values under the alternative hypothesis ($a = 2$) for $n = 200$ and $n = 500$.*

Even if the asymptotic distribution of $\widehat{F}_{\widehat{V},b}(v)$ is not directly used for the test, we use this simulation to compare the asymptotic variance of $\widehat{F}_{\widehat{V},b}(v)$ given in Theorem 3.2 with a

Monte Carlo evaluation of this variance. Figure 4 shows the asymptotic variance (in blue) and the Monte Carlo variance (in red) for different values of v from 0.1 to 0.9. The sample size is $n = 1000$, the bandwidths are respectively $h = 0.2$ and $b = 0.02$ and we use 200 Monte Carlo replications. The simulation is done under the null hypothesis of exogeneity (i.e. $a = 0$ in the above model). The graph shows the relevance of the asymptotic approximation.

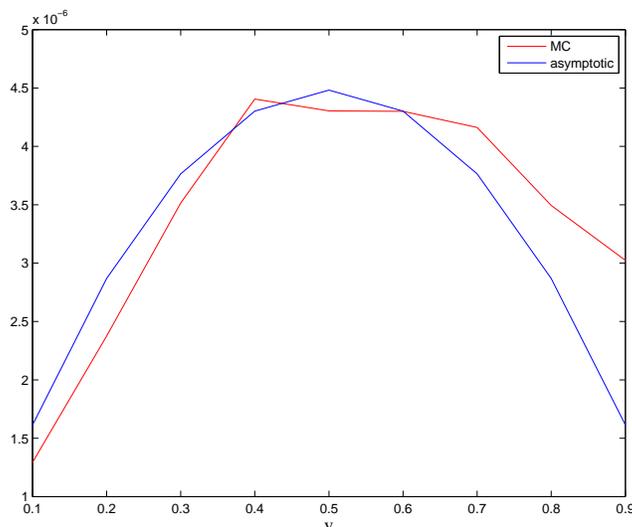


Figure 4: *The asymptotic variance of $\widehat{F}_{\widehat{v},b}(v)$, and the variance obtained from Monte Carlo simulation.*

5.2 Data analysis

We illustrate our approach by an application to Engel curves based on the data of Blundell et al. (2007). We consider three variables: the variable Y is the consumption share of the good, Z is the total expenditure (in log) and the instrument W is the total earning. We consider two goods (food and leisure) and two samples (a sample of $n = 628$ families without children, and a sample of $n = 1027$ families with at least one child). The difference between our test and the previous analyses done by Blundell et al. (2007), Blundell and Horowitz (2007) and Fu (2010) is that we consider a nonseparable model $Y = \varphi(Z, U)$, estimated under the exogeneity assumption and we test the independence between $V = F_{Y|Z}(Y|Z)$ and W . As explained in the paper and illustrated in the simulations, we compute the P -values of the test (based on the Kolmogorov-Smirnov distance) using a bootstrap approximation under the null hypothesis based on 200 resamples. The results are summarized in Table 2.

	No children		At least one child	
	Nonseparable	Separable	Nonseparable	Separable
	model	model	model	model
Food	0.370	0	0	0.045
Leisure	0.095	0.055	0.190	0.100

Table 2: *P-values for the exogeneity tests.*

In Table 2 we also give the P -values for the same null hypothesis (namely the independence between V and W), but where the residuals V are obtained from a separable regression model $Y = m(Z) + V$ with $m(Z) = E(Y|Z)$. The results are similar for the separable and the nonseparable models for three out of the four cases: we reject the exogeneity of the variable ‘Food’ for families with at least one child, and we don’t reject this assumption for the variable ‘Leisure’ for families without children as well as for families with at least one child. On the contrary, we reject the exogeneity hypothesis in a separable model for ‘Food’ for families without children, but we accept this property in a nonseparable model. This result may be viewed as a rejection of the validity of a separable regression model for the estimation of this Engel curve.

6 Conclusions

The first contribution of this paper is the characterization of the asymptotic limit of the empirical distribution of the conditional ranks, which are the residuals of a nonseparable model with uniform noise. The speed of convergence of this empirical distribution is non standard (between \sqrt{n} and n), but the limit is Gaussian. The result is applied to the construction of an exogeneity test based on the independence property between the conditional ranks and some instrumental variable. This test only uses the estimation under the null and not under the alternative. Moreover, it does not require the resolution of a nonlinear integral equation. We illustrate the power of the test by some Monte Carlo simulations and we present a bootstrap strategy to implement the test in practice. The validity of the bootstrap procedure is analyzed via simulations. An application to Engel curves is given as an illustration of our approach. A possible next step could be a theoretical study of the bootstrap validity and a comparison with other possible tests based for example on the estimation of the nonseparable model under the null hypothesis (so assuming exogeneity) or using instruments. The general

methodology presented here may also be applied to other relevant problems: for example, an extension to a test of separability in case of endogeneity is on our research agenda.

7 Appendix A: Identification

For reasons of completeness and self-sufficiency of the paper, we summarize here some results on the identification of the nonparametric nonseparable model (2.1).

Let $(Y, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$ be a random vector and let $f(y|z, w)$ and $F(y|z, w)$ be the conditional density and distribution functions of Y given $Z = z$ and $W = w$. The function $\varphi(z, u)$ (also denoted by $\varphi_z(u)$ in this Appendix) is supposed to be a differentiable strictly increasing function of u for any z . The conditional (residual) distribution and density of $U = \varphi_Z^{-1}(Y)$ are then

$$\begin{aligned} G(u|z, w) &= F(\varphi_z(u)|z, w) \\ g(u|z, w) &= \varphi'_z(u)f(\varphi_z(u)|z, w). \end{aligned}$$

Let us define the perturbed distribution of the residuals in the direction $\Delta_z(u)$ of size $\delta \in [0, 1]$:

$$\begin{aligned} \tilde{G}_{\delta, \Delta}(u|z, w) &= F(\varphi_z(u) + \delta\Delta_z(u)|z, w) \\ \tilde{g}_{\delta, \Delta}(u|z, w) &= [\varphi'_z(u) + \delta\Delta'_z(u)]f(\varphi_z(u) + \delta\Delta_z(u)|z, w). \end{aligned}$$

Note that if $\delta = 0$, $\tilde{g}_{\delta, \Delta}$ reduces to g . Finally, let us construct

$$\tilde{g}_{\delta, \Delta}(u, z|w) = f(z|w)\tilde{g}_{\delta, \Delta}(u|z, w),$$

which defines a family of perturbations of the joint density of Z and U given $W = w$.

We can now define the following concept.

Definition 7.1. *Z is strongly complete by W given U (in the L^2 sense) if for all $a_{\delta, \Delta} \in L^2_{z, u}$ (a family of functions indexed by δ and Δ) and for all Δ_z we have :*

$$\int \tilde{E}_{\delta, \Delta}[a_{\delta, \Delta}(Z, u)|u, w]\tilde{g}_{\delta, \Delta}(u|w) d\delta \equiv 0 \Rightarrow a_{\delta, \Delta}(z, u) \equiv 0,$$

where $\tilde{E}_{\delta, \Delta}[a_{\delta, \Delta}(Z, u)|u, w] = \int a_{\delta, \Delta}(z, u)\tilde{g}_{\delta, \Delta}(z|u, w) dz$ and $L^2_{z, u} = \{(z, u) \rightarrow a(z, u) : \int a^2(z, u)dF_{Z, U}(z, u) < \infty\}$.

If δ is fixed at zero, this concept reduces to the completeness of Z by W for given U (or Z is simply identified by W given U).

Theorem 7.1 (Chernozhukov and Hansen, 2005). *If Z is strongly complete by W given U , the nonseparable instrumental variable model (2.1) is globally identified within the class of functions $\{(z, u) \rightarrow \varphi_z(u) : \varphi_z(u)$ is differentiable and strictly increasing in u for any $z\}$.*

Proof. Let φ be the true function, which is a solution of

$$\int F(\varphi_z(u)|z, w)f(z|w)dz = u,$$

where F is identifiable. Let ψ be another solution of this equation. We have :

$$\int \left[F(\psi_z(u)|z, w) - F(\varphi_z(u)|z, w) \right] f(z|w)dz = 0.$$

Let $\Delta_z = \psi_z - \varphi_z$ for any z . This equality is equivalent to

$$\int f(z|w) \int_0^1 f(\varphi_z(u) + \delta\Delta_z(u)|z, w)\Delta_z(u) d\delta dz = 0.$$

By assumption $\varphi'_z > 0$ and $\psi'_z > 0$. for any z . Then, $\varphi'_z + \delta\Delta'_z > 0$ for $0 \leq \delta \leq 1$, and the latter equality is equivalent to

$$\int_0^1 \int \frac{\Delta_z(u)}{\varphi'_z(u) + \delta\Delta'_z(u)} \tilde{g}_{\delta, \Delta}(u, z|w) dz d\delta = 0$$

or

$$\int_0^1 \tilde{g}_{\delta, \Delta}(u|w) \int \frac{\Delta_z(u)}{\varphi'_z(u) + \delta\Delta'_z(u)} \tilde{g}_{\delta, \Delta}(z|u, w) dz d\delta = 0,$$

for any u and w , which implies that $\Delta_z(u)/[\varphi'_z(u) + \delta\Delta'_z(u)] \equiv 0$ and hence $\Delta(z, u) \equiv 0$. \square

8 Appendix B: Proofs

We start this Appendix by giving the conditions under which the main results in this paper are valid.

(A1) *[On the bandwidths]* The bandwidths h , g and b satisfy $n^2bh(h^4 + g^4) \rightarrow 0$, $nb^2 \rightarrow \infty$, $nh^{3+2\gamma}(\log n)^{-1} \rightarrow \infty$ for some $0 < \gamma < 1$, and $(\log n)^2(b/h) \rightarrow 0$.

(A2) *[On the kernels]* The functions k , ℓ and m are symmetric, twice continuously differentiable probability density functions with compact support.

(A3) [On the support] The support R_Z and R_W (of Z and W) are compact subsets of \mathbb{R} .

(A4) [On $F_{Y|Z}$] The distribution $F_{Y|Z}(y|z)$ is twice continuously differentiable with respect to y and z , the derivatives are continuous in (y, z) , and

$$\sup_{y,z} \left| \frac{\partial^k}{\partial z^{k_1} \partial y^{k_2}} F_{Y|Z}(y|z) \right| < \infty,$$

where $k = k_1 + k_2$, $k = 1, 2$ and $0 \leq k_1, k_2 \leq 2$. Moreover, $\inf_{y,z} f_{Y|Z}(y|z) > 0$.

(A5) [On F_Z , $F_{Y,W|Z}$ and $F_{W|V,Z}$] The distribution $F_Z(z)$ is three times continuously differentiable, and $\inf_z f_Z(z) > 0$. The distribution $F_{Y,W|Z}(y, w|z)$ is twice continuously differentiable with respect to y and

$$\sup_{y,z,w} \left| \frac{\partial^k}{\partial y^k} F_{Y,W|Z}(y, w|z) \right| < \infty$$

for $k = 1, 2$. Finally, the distribution $F_{W|V,Z}(w|v, z)$ is continuously differentiable with respect to z and $\sup_{w,v,z} \left| \frac{\partial}{\partial z} F_{W|V,Z}(w|v, z) \right| < \infty$.

Before proving the asymptotic results of Sections 3 and 4, we start with two technical lemmas.

Lemma 8.1. *Assume (A1)-(A5). Then,*

$$\sup_{v \in I} \left| n^{-1} \sum_{i=1}^n \left\{ I(\widehat{V}_i \leq v) - I(V_i \leq v) - P(\widehat{V} \leq v) + P(V \leq v) \right\} \right| = o_P(n^{-1/2}), \quad (8.1)$$

where the probability $P(\widehat{V} \leq v) = P(\widehat{F}_{Y|Z}(Y|Z) \leq v)$ is calculated with respect to the law of (Y, Z) , conditional on $\widehat{F}_{Y|Z}$ (and where (Y, Z) is independent of the data (Y_i, Z_i) , $i = 1, \dots, n$).

Proof. The main idea of the proof is as follows : the expression inside the absolute values in (8.1) will be embedded into an empirical process indexed by v and by a class of functions to which the function $(y, z) \rightarrow \widehat{F}_{Y|Z}(y|z)$ belongs with probability tending to one. We show that this empirical process is Donsker, from which it will follow that the expression on the left hand side in (8.1) is $O_P(n^{-1/2})$. Finally, the required order $o_P(n^{-1/2})$ will follow by showing that this empirical process is degenerate.

Let's now look at each of these steps in more detail. Define the following class :

$$\mathcal{F} = \left\{ (y, z) \rightarrow I(F(y|z) \leq v) : v \in I, F(\cdot|z) \text{ is monotone onto } [0, 1] \text{ for all } z \in R_Z, \right. \\ \left. F^{-1}(v|\cdot) \in C_M^{1+\gamma}(R_Z) \text{ for all } v \in I \right\}.$$

Here, $C_M^{1+\gamma}(R_Z)$ is the class of all differentiable functions h defined on the domain R_Z of Z such that $\|h\|_{1+\gamma} \leq M < \infty$, where

$$\|h\|_{1+\gamma} = \max \left\{ \sup_z |h(z)|, \sup_z |h'(z)| \right\} + \sup_{z_1, z_2} \frac{|h'(z_1) - h'(z_2)|}{|z_1 - z_2|^\gamma},$$

and $0 < \gamma < 1$ is defined in assumption (A1). Note that $\frac{\partial}{\partial z} \widehat{F}_{Y|Z}^{-1}(v|z)$ exists for any v and z , since $\widehat{F}_{Y|Z}(y|z)$ is smooth in y and z . It can then be shown that

$$\sup_{v \in I, z \in R_Z} \left| \widehat{F}_{Y|Z}^{-1}(v|z) - F_{Y|Z}^{-1}(v|z) \right| = o_P(1), \quad (8.2)$$

$$\sup_{v \in I, z \in R_Z} \left| \frac{\partial}{\partial z} \widehat{F}_{Y|Z}^{-1}(v|z) - \frac{\partial}{\partial z} F_{Y|Z}^{-1}(v|z) \right| = o_P(1), \quad (8.3)$$

$$\sup_{v \in I, z_1, z_2 \in R_Z} \frac{\left| \frac{\partial}{\partial z} \widehat{F}_{Y|Z}^{-1}(v|z_1) - \frac{\partial}{\partial z} F_{Y|Z}^{-1}(v|z_1) - \frac{\partial}{\partial z} \widehat{F}_{Y|Z}^{-1}(v|z_2) + \frac{\partial}{\partial z} F_{Y|Z}^{-1}(v|z_2) \right|}{|z_1 - z_2|^\gamma} = o_P(1). \quad (8.4)$$

In fact, among equations (8.2)–(8.4) the most difficult result to prove is the last one. Using Propositions 4.1 and 4.2 in Akritas and Van Keilegom (2001), (8.4) is easily seen to be $O_P((nh^{3+2\gamma})^{-1/2}(\log n)^{1/2})$ and this is $o_P(1)$ if $nh^{3+2\gamma}(\log n)^{-1} \rightarrow \infty$. Hence, since $F_{Y|Z}^{-1}(v|\cdot) \in C_M^{1+\gamma}(R_Z)$ for all $v \in I$, the same is true for the estimator $\widehat{F}_{Y|Z}^{-1}(v|\cdot)$ with probability tending to one.

We will now show that the class \mathcal{F} is Donsker. For the class of functions of the form $z \rightarrow F^{-1}(v|z)$ that belong to $C_M^{1+\gamma}(R_Z)$ and with $v \in I$, there exist ϵ^2 -brackets $b_j^L \leq b_j^U$, $j = 1, \dots, M_\epsilon = O(\exp(\epsilon^{-2/(1+\gamma)}))$, such that for a given v and F there exists a $1 \leq j \leq M_\epsilon$ satisfying

$$b_j^L(\cdot) \leq F^{-1}(v|\cdot) \leq b_j^U(\cdot)$$

(see Corollary 2.7.2 in Van der Vaart and Wellner, 1996), and hence we also have that

$$I(b_j^L(\cdot) \geq y) \leq I(F(y|\cdot) \leq v) \leq I(b_j^U(\cdot) \geq y).$$

Moreover,

$$\begin{aligned} & \int [I(b_j^U(z) \geq y) - I(b_j^L(z) \geq y)]^2 dF_{Y,Z}(y, z) \\ &= \int [F_{Y|Z}(b_j^U(z)|z) - F_{Y|Z}(b_j^L(z)|z)] dF_Z(z) \\ &\leq K \|b_j^U - b_j^L\|_{L_1^P} \leq K \|b_j^U - b_j^L\|_{L_2^P} \leq K \epsilon^2, \end{aligned}$$

provided $\sup_{y,z} f_{Y|Z}(y|z) < \infty$. It now follows that

$$\int_0^{2M} (\log N_{[]}(\mathcal{F}, \epsilon, L_2(P)))^{1/2} d\epsilon \leq \int_0^{2M} \epsilon^{-1/(1+\gamma)} d\epsilon = K[(2M)^{\gamma/(1+\gamma)}] < \infty,$$

where P is the joint probability measure of (Y, Z) . Hence, the class \mathcal{F} is Donsker.

Next, in order to show (8.1), we will use the following result, which is valid for any Donsker class \mathcal{F} (see Corollary 2.3.12 in Van der Vaart and Wellner, 1996) :

$$\lim_{\alpha \rightarrow 0} \lim_{\epsilon \rightarrow 0} P \left(\sup_{\substack{f, g \in \mathcal{F} \\ \rho_P(f-g) < \alpha}} \left| n^{-1/2} \sum_{i=1}^n \left\{ f(Y_i, Z_i) - g(Y_i, Z_i) - Ef(Y, Z) + Eg(Y, Z) \right\} \right| > \epsilon \right) = 0,$$

where $\rho_P^2(f) = \text{Var}(f(Y, Z))$. Therefore, we calculate

$$\begin{aligned} & \text{Var} \left(I(\widehat{F}_{Y|Z}(Y|Z) \leq v) - I(F_{Y|Z}(Y|Z) \leq v) \right) \\ & \leq E \left(I(\widehat{F}_{Y|Z}(Y|Z) \leq v) - I(F_{Y|Z}(Y|Z) \leq v) \right)^2 \\ & = \int \left[F_{Y|Z}(\widehat{F}_{Y|Z}^{-1}(v|z)|z) - F_{Y|Z}(\widehat{F}_{Y|Z}^{-1}(v|z) \wedge F_{Y|Z}^{-1}(v|z)|z) \right] dF_Z(z) \\ & \quad + \int \left[F_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) - F_{Y|Z}(\widehat{F}_{Y|Z}^{-1}(v|z) \wedge F_{Y|Z}^{-1}(v|z)|z) \right] dF_Z(z) \\ & \leq \sup_{v \in I, z \in R_Z} |\widehat{F}_{Y|Z}^{-1}(v|z) - F_{Y|Z}^{-1}(v|z)| = o(1) \quad a.s. \end{aligned}$$

Hence, the result follows. □

Lemma 8.2. *Assume (A1)-(A5). Then,*

$$\begin{aligned} & \sup_{v \in I, w \in R_W} \left| n^{-1} \sum_{i=1}^n \left\{ I(\widehat{V}_i \leq v, W_i \leq w) - I(V_i \leq v, W_i \leq w) \right. \right. \\ & \quad \left. \left. - P(\widehat{V} \leq v, W \leq w) + P(V \leq v, W \leq w) \right\} \right| = o_P(n^{-1/2}). \end{aligned}$$

The proof of this result is similar to the proof of Lemma 8.1 and is therefore omitted.

Proof of Theorem 3.1. We start with the first statement. Lemma 8.1 implies that

$$\begin{aligned} \widehat{F}_{\widehat{V}}(v) - v &= \left[n^{-1} \sum_{i=1}^n I(V_i \leq v) - v \right] + \left[P(\widehat{V} \leq v) - v \right] + o_P(n^{-1/2}) \\ &= T_1(v) + T_2(v) + o_P(n^{-1/2}), \end{aligned}$$

uniformly in $v \in I$. Consider $T_2(v)$:

$$\begin{aligned}
T_2(v) &= \int \left[F_{Y|Z}(\widehat{F}_{Y|Z}^{-1}(v|z)|z) - F_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) \right] f_Z(z) dz \\
&= \int f_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) \left[\widehat{F}_{Y|Z}^{-1}(v|z) - F_{Y|Z}^{-1}(v|z) \right] f_Z(z) dz + o_P(n^{-1/2}) \\
&= - \int \left[\widehat{F}_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) - v \right] f_Z(z) dz + o_P(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n \int k_h(Z_i - z) \left[L_g(F_{Y|Z}^{-1}(v|z) - Y_i) - v \right] dz + o_P(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n \left[L_g(F_{Y|Z}^{-1}(v|Z_i) - Y_i) - v \right] \\
&\quad - n^{-1} \sum_{i=1}^n \int k(t) \left[L_g(F_{Y|Z}^{-1}(v|Z_i - th) - Y_i) - L_g(F_{Y|Z}^{-1}(v|Z_i) - Y_i) \right] dt + o_P(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n I(V_i \leq v) + v + o_P(n^{-1/2}),
\end{aligned}$$

since $nh^4 = o(1)$ and $ng^4 = o(1)$. This shows that $T_2(v) = -T_1(v) + o_P(n^{-1/2})$ uniformly in v , and hence the first statement follows.

For the second one, we use a similar decomposition into two terms, this time by using Lemma 8.2 :

$$\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w) = S_1(v, w) + S_2(v, w) + o_P(n^{-1/2}), \quad (8.5)$$

uniformly in v and w , where $S_1(v, w) = n^{-1} \sum_{i=1}^n I(V_i \leq v, W_i \leq w) - F_{V,W}(v, w)$ and $S_2(v, w) = P(\widehat{V} \leq v, W \leq w) - F_{V,W}(v, w)$. Using similar arguments as above, we can show that

$$S_2(v, w) = - \int \frac{\frac{\partial}{\partial y} F_{Y,W|Z}(y, w|z)|_{y=F_{Y|Z}^{-1}(v|z)}}{f_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z)} \left[\widehat{F}_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) - v \right] f_Z(z) dz + o_P(n^{-1/2}).$$

Note that since $F_{Y,W|Z}(y, w|z) = \int_{-\infty}^y F_{W|Y,Z}(w|t, z) f_{Y|Z}(t|z) dt$, it follows that

$$\frac{\partial}{\partial y} F_{Y,W|Z}(y, w|z) = F_{W|Y,Z}(w|y, z) f_{Y|Z}(y|z),$$

and hence

$$\begin{aligned}
S_2(v, w) &= - \int F_{W|Y,Z}(w|F_{Y|Z}^{-1}(v|z), z) \left[\widehat{F}_{Y|Z}(F_{Y|Z}^{-1}(v|z)|z) - v \right] f_Z(z) dz + o_P(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n F_{W|V,Z}(w|v, Z_i) \left[I(V_i \leq v) - v \right] + o_P(n^{-1/2}), \quad (8.6)
\end{aligned}$$

where the latter equality can be proved similarly as in the first part of the proof. This shows that $S_1(v, w) + S_2(v, w) = O_P(n^{-1/2})$ uniformly in v and w (but contrarily to the first part of the proof, we do not have $o_P(n^{-1/2})$, since $S_1(v, w)$ and $S_2(v, w)$ do not compensate each other). \square

Proof of Theorem 3.2. Using the fact that $\max_i |\widehat{V}_i - V_i| = O_P((nh)^{-1/2}(\log n)^{1/2})$, we can write

$$\begin{aligned}
& \widehat{F}_{\widehat{V}, b}(v) - v \\
&= n^{-1} \sum_{i=1}^n M_b(v - V_i) - v + n^{-1} \sum_{i=1}^n m_b(v - V_i)(V_i - \widehat{V}_i) + O_P((nh)^{-1} \log n) \\
&= n^{-2} \sum_{i=1}^n \sum_{j \neq i} \widehat{f}_Z^{-1}(Z_i) k_h(Z_i - Z_j) \left\{ M_b(v - V_i) - v + m_b(v - V_i)[V_i - L_g(Y_i - Y_j)] \right\} \\
&\quad + T_n(v) + o_P((n^2bh)^{-1/2}), \tag{8.7}
\end{aligned}$$

provided $bh^{-1}(\log n)^2 = o(1)$, where

$$T_n(v) = n^{-2} \sum_{i=1}^n \widehat{f}_Z^{-1}(Z_i) k_h(0) \left\{ M_b(v - V_i) - v + m_b(v - V_i)[V_i - 0.5] \right\},$$

and where $\widehat{f}_Z(z) = n^{-1} \sum_{i=1}^n k_h(z - Z_i)$ is a kernel estimator of the density $f_Z(z)$. It is readily seen that $T_n(v) = O_P((nh)^{-1}) = o_P((n^2hb)^{-1/2})$ since $b = o(h)$. Let $Q_{nij}(v)$ be the expression between curled brackets in (8.7). Then, the main term in (8.7) can be decomposed as follows :

$$\begin{aligned}
& n^{-2} \sum_{i \neq j} \widehat{f}_Z^{-1}(Z_i) k_h(Z_i - Z_j) Q_{nij}(v) \\
& - n^{-2} \sum_{i \neq j} \widehat{f}_Z^{-2}(Z_i) (\widehat{f}_Z(Z_i) - f_Z(Z_i)) k_h(Z_i - Z_j) Q_{nij}(v) \\
& + n^{-2} \sum_{i \neq j} \widehat{f}_Z^{-1}(Z_i) \widehat{f}_Z^{-2}(Z_i) (\widehat{f}_Z(Z_i) - f_Z(Z_i))^2 k_h(Z_i - Z_j) Q_{nij}(v) \\
& = R_1(v) + R_2(v) + R_3(v).
\end{aligned}$$

The term $R_3(v)$ is $O_P((nh)^{-1} \log n) = o_P((n^2bh)^{-1/2})$. We focus in what follows on the first term $R_1(v)$, which is a U -statistic of order two. The derivation for the term $R_2(v)$ (which is

a U -statistic of order three), is similar and leads to $R_2(v) = o_P((n^2hb)^{-1/2})$. Write

$$\begin{aligned} R_1(v) &= \binom{n}{2}^{-1} \sum_{i < j} \frac{1}{2} \left[f_Z^{-1}(Z_i) k_h(Z_i - Z_j) Q_{nij}(v) + f_Z^{-1}(Z_j) k_h(Z_i - Z_j) Q_{nji}(v) \right] \\ &:= [n(n-1)]^{-1} \sum_{i < j} \left[h_n(T_i, T_j) - E(h_n(T_i, T_j) | T_i) \right] + [n(n-1)]^{-1} \sum_{i < j} E(h_n(T_i, T_j) | T_i) \\ &= R_{11}(v) + R_{12}(v), \end{aligned}$$

where $T_i = (Z_i, V_i, Y_i)$. We start with $R_{12}(v)$:

$$\begin{aligned} &R_{12}(v) \\ &= n^{-1} \sum_{i=1}^n f_Z^{-1}(Z_i) E \left[k_h(Z_i - Z) \left\{ M_b(v - V_i) - v + m_b(v - V_i) [V_i - E(L_g(Y_i - Y) | Z, Y_i)] \right\} \middle| T_i \right] \\ &\quad + n^{-1} \sum_{i=1}^n E \left[f_Z^{-1}(Z) k_h(Z_i - Z) \left\{ v - E[m_b(v - V) L_g(\varphi(Z, V) - Y_i) | Z, Y_i] \right\} \middle| T_i \right] + O(b^4) \\ &= n^{-1} \sum_{i=1}^n \left[M_b(v - V_i) - v + m_b(v - V_i) V_i - f_Z^{-1}(Z_i) m_b(v - V_i) E \left\{ k_h(Z_i - Z) F_{Y|Z}(Y_i | Z) \middle| T_i \right\} \right] \\ &\quad + n^{-1} \sum_{i=1}^n \left[v - E \left\{ f_Z^{-1}(Z) k_h(Z_i - Z) M_b(v - \varphi_Z^{-1}(Y_i)) \middle| T_i \right\} \right] + O(b^2) + O(h^2) + O(g^2), \end{aligned}$$

where $\varphi_z^{-1}(\cdot)$ is the inverse of $\varphi(z, \cdot)$ for fixed z . The latter can be easily seen to be equal to $O(b^2) + O(h^2) + O(g^2) = o((n^2bh)^{-1/2})$ using the assumptions on b, h and g .

Next we consider $R_{11}(v)$. This is a degenerate U -statistic with kernel depending on n . We use Theorem 1 in Hall (1984) to obtain its limiting distribution. For this we need to show that

$$\frac{E(G_n^2(T_1, T_2)) + n^{-1} E(H_n^4(T_1, T_2))}{\{E(H_n^2(T_1, T_2))\}^2} \rightarrow 0, \quad (8.8)$$

where $G_n(t_1, t_2) = E(H_n(T, t_1)H_n(T, t_2))$ and $H_n(t_1, t_2) = [n(n-1)]^{-1}h_n(t_1, t_2)$. It is easy to show that $E(H_n^2(T_1, T_2)) = O(n^{-4}h^{-1}b^{-1})$, $E(H_n^4(T_1, T_2)) = O(n^{-8}h^{-3}b^{-3})$ and $E(G_n^2(T_1, T_2)) = O(n^{-8}h^{-1}b^{-2})$. Hence, the left hand side of (8.8) is $O(h + (nhb)^{-1}) = o(1)$ if $nhb \rightarrow \infty$. Hence, Theorem 1 in Hall (1984) implies that $n(hb)^{1/2}R_{11}(v)$ is asymptotically normally distributed with zero mean and variance given by

$$\frac{1}{2} \lim_{n \rightarrow \infty} \left[n^4 h b \text{Var} \{ H_n(T_1, T_2) \} \right] = \frac{1}{2} \lim_{n \rightarrow \infty} \left[h b \text{Var} \{ h_n(T_1, T_2) \} \right].$$

Note that $\text{Var} \{ h_n(T_1, T_2) \}$ is asymptotically equivalent to $E(\{h_n(T_1, T_2)\}^2)$, and by the law of iterated expectation, this is asymptotically equal to

$$2E \left[f_Z^{-2}(Z_1) m_b^2(v - V_1) E \left(\left\{ k_h(Z_1 - Z_2) [V_1 - L_g(Y_1 - Y_2)] \middle| T_1 \right\}^2 \right) \right], \quad (8.9)$$

since the covariance between the two terms of $h_n(T_1, T_2)$ is asymptotically negligible with respect to the variance of each of these terms. The interior expectation is asymptotically equivalent to

$$\begin{aligned} & \int k_h^2(Z_1 - z) \{V_1 - I(v \leq F_{Y|Z}^{-1}(Y_1|z))\}^2 f_Z(z) dz dv \\ &= h^{-1} V_1 (1 - V_1) f_Z(Z_1) \int k^2(t) dt (1 + o(1)), \end{aligned}$$

and hence (8.9) equals

$$\begin{aligned} & 2h^{-1} E[f_Z^{-1}(Z) m_b^2(v - V) V(1 - V)] \int k^2(t) dt (1 + o(1)) \\ &= 2(hb)^{-1} v(1 - v) \int m^2(t) dt \int k^2(t) dt (1 + o(1)). \end{aligned}$$

This shows that

$$\frac{1}{2} \lim_{n \rightarrow \infty} [hb \text{Var}\{h_n(T_1, T_2)\}] = v(1 - v) \int m^2(t) dt \int k^2(t) dt.$$

This shows the first statement of the theorem. Note that in the proof we could have chosen the bandwidth g equal to zero without any complication. This agrees with Remark 3.1(c), which states that the bandwidth g is only necessary for the proof of Theorem 3.1.

For the second statement, note that it follows from (8.5) and (8.6) that

$$\begin{aligned} & \widehat{F}_{\widehat{V}, W}(v, w) \tag{8.10} \\ &= n^{-1} \sum_{i=1}^n \left\{ I(V_i \leq v, W_i \leq w) - F_{W|V, Z}(w|v, Z_i) [I(V_i \leq v) - v] \right\} + o_P(n^{-1/2}). \end{aligned}$$

Hence, $n^{1/2}(\widehat{F}_{\widehat{V}, W}(v, w) - F_{V, W}(v, w))$ is asymptotically normal with mean zero and variance given by

$$\begin{aligned} & E\left(\left\{ I(V \leq v, W \leq w) - vF_W(w) - F_{W|V, Z}(w|v, Z) [I(V \leq v) - v] \right\}^2\right) \\ &= vF_W(w)(1 - vF_W(w)) + v(1 - v) \int F_{W|V, Z}^2(w|v, z) f_Z(z) dz \\ &\quad - 2(1 - v) \int F_{W|V, Z}(w|v, z) \int_0^v F_{W|V, Z}(w|u, z) du f_Z(z) dz. \end{aligned}$$

This shows the second statement. □

Proof of Theorem 4.1. Using the first part of Theorem 3.1 and equation (8.10), we can write

$$\begin{aligned}
& \widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v)\widehat{F}_W(w) \\
&= n^{-1} \sum_{i=1}^n I(V_i \leq v, W_i \leq w) - n^{-1} \sum_{i=1}^n F_{W|V,Z}(w|v, Z_i) [I(V_i \leq v) - v] \\
&\quad - v n^{-1} \sum_{i=1}^n I(W_i \leq w) + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n [I(V_i \leq v) - v] [I(W_i \leq w) - F_{W|V,Z}(w|v, Z_i)] + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in v and w . □

Proof of Theorem 4.2. The class of functions $\{z \rightarrow F_{W|V,Z}(w|v, z) : v \in I, w \in R_W\}$ is Donsker, since $F_{W|V,Z}(w|v, z)$ is continuously differentiable in z and $\frac{\partial}{\partial z} F_{W|V,Z}(w|v, z)$ is bounded uniformly in v, w and z (see Corollary 2.7.2 in Van der Vaart and Wellner, 1996). Hence, the class

$$\mathcal{F} = \left\{ (V, W, Z) \rightarrow [I(V \leq v) - v] [I(W \leq w) - F_{W|V,Z}(w|v, Z)] : v \in I, w \in R_W \right\}$$

is also Donsker. This, combined with Theorem 4.1, shows the weak convergence of the process $n^{1/2}(\widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v)\widehat{F}_W(w))$ indexed by $(v, w) \in I \times R_W$ (see Theorem 2.5.6 in Van der Vaart and Wellner, 1996). The covariance function of the limiting process is given by

$$\begin{aligned}
& \text{Cov}(G(v_1, w_1), G(v_2, w_2)) \\
&= \left\{ v_1 \wedge v_2 - v_1 v_2 \right\} \left\{ F_W(w_1 \wedge w_2) - \int [F_{W|V,Z}(w_1|v_1, z) F_{W|Z}(w_2|z) \right. \\
&\quad \left. + F_{W|V,Z}(w_2|v_2, z) F_{W|Z}(w_1|z) - F_{W|V,Z}(w_1|v_1, z) F_{W|V,Z}(w_2|v_2, z)] f_Z(z) dz \right\}.
\end{aligned}$$

This finishes the proof. □

Proof of Corollary 4.1. The convergence of $T_{n,KS}$ follows readily from the continuous mapping theorem. For $T_{n,CM}$ it suffices to show that $d\widehat{F}_{\widehat{V},W}(v, w)$ can be replaced by $dF_{V,W}(v, w)$. Using the weak convergence of the processes $\widehat{G}(v, w) := n^{1/2}(\widehat{F}_{\widehat{V},W}(v, w) - \widehat{F}_{\widehat{V}}(v)\widehat{F}_W(w))$ and $n^{1/2}(\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w))$, and the Skorohod construction (see e.g. Serfling, 1980) we can write

$$\sup_{v,w} |\widehat{G}(v, w) - G(v, w)| \rightarrow_{a.s.} 0 \quad \text{and} \quad \sup_{v,w} |\widehat{F}_{\widehat{V},W}(v, w) - F_{V,W}(v, w)| \rightarrow_{a.s.} 0 \quad (8.11)$$

(for simplicity we keep the same notation as for the original processes). Here, $G(v, w)$ is the limiting process given in Theorem 4.2. Now write

$$\begin{aligned} & \left| \int \widehat{G}^2(v, w) d\widehat{F}_{\widehat{V}, W}(v, w) - \int G^2(v, w) dF_{V, W}(v, w) \right| \\ & \leq \left| \int (\widehat{G}^2(v, w) - G^2(v, w)) d\widehat{F}_{\widehat{V}, W}(v, w) \right| + \left| \int G^2(v, w) d(\widehat{F}_{\widehat{V}, W}(v, w) - F_{V, W}(v, w)) \right|. \end{aligned}$$

The first expression in (8.11) implies that the first term above is $o(1)$ a.s. For the second term, note that the trajectories of the limit process $G(v, w)$ are bounded and continuous almost surely. Therefore, taking into account the second expression in (8.11), we can apply the Helly-Bray Theorem (see e.g. Rao, 1965, p. 97) to each of these trajectories and conclude that $|\int G^2(v, w) d(\widehat{F}_{\widehat{V}, W}(v, w) - F_{V, W}(v, w))| \rightarrow_{a.s.} 0$. \square

References

- Akritis, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **28**, 549-568.
- Blundell, R., Chen, X. and Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, **75**, 1613-1669.
- Blundell, R. and Horowitz, J.L. (2007). A non-parametric test of exogeneity. *Review of Economic Studies*, **74**, 1035-1058.
- Chen, S., Dahl, G.B. and Khan, S. (2005). Nonparametric identification and estimation of a location-scale censored regression model. *Journal of the American Statistical Association*, **100**, 212-221.
- Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, **80**, 277-321.
- Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73**, 245-261.
- Chernozhukov, V., Scaillet, O. and Gagliardini, P. (2011). Nonparametric instrumental variable estimators of structural quantile effects. Technical report.
- Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society - Series B*, **69**, 903-917.
- Dunker, F., Florens, J.-P., Hohage, T., Johannes, J. and Mammen, E. (2012). Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. Technical report.

- Einmahl, J. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, **143**, 88-102.
- Engle, R.F., Hendry, D.F. and Richard, J.-P. (1983). Exogeneity. *Econometrica*, **51**, 277-304.
- Florens, J.-P., Johannes, J. and Van Bellegem, S. (2011). Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, **27**, 472-496.
- Florens, J.-P. and Mouchart, M. (1985). Conditioning in dynamic models. *Journal of Time Series Analysis*, **6**, 15-34.
- Fu, J.-Y. M. (2010). Testing exogeneity in nonparametric instrumental variables models identified by conditional quantile restrictions. Technical report.
- Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, **14**, 1-16.
- Hausman, J.A. (1978). Specification tests in econometrics. *Econometrica*, **46**, 1251-1271.
- Hoderlein, S. and Mammen, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, **75**, 1513-1518.
- Horowitz, J.L. and Lee, S. (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, **75**, 1191-1208.
- Imbens, G. and Newey, W. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77**, 1481-1512.
- Johannes, J., Van Bellegem, S. and Vanhems, A. (2011). Convergence rates for ill-posed inverse problems with an unknown operator. *Econometric Theory*, **27**, 522-545.
- Lewbel, A. and Linton, O. (2002). Nonparametric censored and truncated regression. *Econometrica*, **70**, 765-779.
- Matzkin, R.L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, **71**, 1339-1375.
- Müller, U.U., Schick, A. and Wefelmeyer, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statistics and Probability Letters*, **79**, 957-964.
- Müller, U.U., Schick, A. and Wefelmeyer, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference*, **142**, 552-566.
- Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scandinavian Journal of Statistics*, **36**, 204-228.

- Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101**, 1067-1078.
- Rao, C.R. (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.
- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Sokullu, S. (2011). Nonparametric analysis of two-sided markets. Technical report.
- Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.
- Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, **41**, 733-750.