

CONSISTENT TESTS OF CONDITIONAL MOMENT RESTRICTIONS

Miguel A. Delgado
Universidad Carlos III de Madrid, Spain

Manuel A. Domínguez
ITAM, Mexico

Pascal Lavergne
INRA-ESR Toulouse, France

Abstract

We address the issue of building consistent specification tests in econometric models defined through multiple conditional moment restrictions. In this aim, we extend the two methodologies developed for testing the parametric specification of a regression function to testing general conditional moment restrictions. Two classes of tests are proposed, which can be both interpreted as M-tests based on integrated conditional moment restrictions. The first class depends upon nonparametric functions that are estimated by kernel smoothers. The second type of test is built as a functional of a marked empirical process. For both tests, a simulation procedure for obtaining critical values is shown to be asymptotically valid. Comparison of finite sample performances of the tests are investigated by means of several Monte-Carlo experiments.

JEL Classification: Primary C52; Secondary C14.

Key words: Hypothesis testing, Nonlinear models, Nonparametric methods, Empirical processes.

CONSISTENT TESTS OF CONDITIONAL MOMENT RESTRICTIONS

1 Introduction

Econometric models are frequently defined through conditional moment restrictions. This is the case for instance for models parameterizing different conditional moments (e.g. conditional mean and conditional variance) without specific distributional assumptions, transformation models, models defined by means of instrumental variables, and nonlinear-in-variables simultaneous equation models. Efficient estimation under conditional moment restrictions is considered by Chamberlain (1987), Newey (1990) and Robinson (1991). The so-called M-tests, as proposed by Newey (1985a, 1985b), Tauchen (1985) and Wooldridge (1990) among others, take as the null hypothesis a finite number of arbitrary unconditional moment restrictions implied by the conditional moment restrictions. These tests are therefore unable to detect all departures from the null hypothesis. Tests that are consistent against nonparametric alternatives can be obtained by considering some unconditional moment restrictions that uniquely characterize the conditional moment restrictions. Two competing methodologies have been developed in the recent literature for specification testing of univariate regression models. The first approach compares the parametric estimated function with a semiparametric or nonparametric function is estimated using smoothers. This leads to asymptotically pivotal tests statistics, see Eubank and Spiegelman (1990), Kozek (1991), Härdle and Mammen (1993), Horowitz and Härdle (1993), Hong and White (1995), Fan and Li (1996), Zheng (1996) among others. Hart's monograph (1997) surveys part of the statistical literature on the topic. The second approach compares integral transforms of the competing regression curves rather than the curves themselves. Indeed, a function can be uniquely characterized by an integral transform, see Apostol (1957, Chap 11). For instance, a probability density function is uniquely characterized by two integral transforms: the distribution function and the characteristic function. The integral regression function generalizes the distribution function concept to the regression case, see Prakasa Rao (1983 pp. 256-258), and is used for testing purposes by Buckley (1991), Hong-Zhy and Bin (1991), Delgado (1993), Stute (1997), Koul and Stute (1999) and Whang (2000), to mention just a few. Bierens and coauthors use the generalization of the characteristic function to the regression case and propose a specification test, see Bierens (1982, 1990), de Jong and Bierens

(1994) and Bierens and Ploberger (1997). The methodology based on integral transforms yields test statistics that have case dependent limiting distributions, but do not depend upon the choice of a smoothing parameter.

Surprisingly, work on consistent tests in econometric models mainly focuses on regression models, with few exceptions.¹ Stichcombe and White (1998) and Whang (2001) propose tests for an univariate conditional moment restriction that extends Bierens' work. Using the same methodology, Chen and Fan (1999) provide consistent procedures for testing parametric restrictions in semiparametric models and nonparametric restrictions in nonparametric models, e.g., testing for omitted variables. However, they do not allow the null hypothesis to depend upon unknown parameters. Using a smoothing approach, Li (1999) considers testing portfolio conditional mean-variance efficiency, an hypothesis also studied by Chen and Fan (1999).

The aim of this paper is to propose tests for multiple conditional moment restrictions with unknown parameters and thus to provide procedures that can prove useful in a variety of econometric models. The characteristic feature of our study with respect to previous work is to simultaneously follow and generalize the two methodologies developed for specification testing of regression models. We restrict ourselves to an iid context. Extension to a time-series context should follow along the lines of de Jong (1996), who considers Bierens' (1990) test under data dependence, and Li (1999), who generalizes Fan and Li's (1996) results. Here, we will concentrate on the particular features arising when the conditional moment restrictions are multidimensional and possibly nonlinear in the endogenous variables. This allows us to point out the inherent problems of the generalization and to raise some open questions. From a practical viewpoint, we look throughout our paper at some examples of applications and we show how to implement each type of tests in practice. Finally, we compare the behavior of the two types of tests by means of several Monte-Carlo simulations, as there exists few such studies in the econometric literature.

The paper is organized as follows. In the next section, we detail our general testing framework and we discuss examples of applications. In Section 3, we study the asymptotics of some tests statistics based upon either nonparametric kernel estimation or integral-transform regression estimation. In Section 4, we propose to approximate critical values of each type of test by simulation procedures. Section 5 reports the results of the Monte-Carlo study. Section 6 gives some directions for further research. Technical proofs are confined to the last section.

¹See Bierens and Ginther (1999) and Zheng (1998) for quantile regression models.

2 Testing Framework

Let $\mathcal{Z}_n = \{Z_i, i = 1, \dots, n\}$ be a sample of independent observations, identically distributed as the random variable Z with values in $\mathbb{R}^s$, and let X be a subvector of Z with dimension q . As done by Newey (1985), Chamberlain (1987), Wooldridge (1990) and White (1994), we consider a particular parametric model indexed by $\theta \in \Theta$, a compact subset of $\mathbb{R}^p$, and defined through conditional moment restrictions of the form

$$H_0 : \exists \theta_0 \in \Theta, \quad E[\psi(Z, \theta_0) | X] = \mathbf{0} \quad \text{a.s.}, \quad (1)$$

where $\psi(\cdot, \cdot) : \mathbb{R}^s \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a vector of known functions and $\mathbf{0}$ is the null vector of $\mathbb{R}^m$. We call $\psi(\cdot, \cdot)$ a generalized residual vector, as Wooldridge (1990) does by analogy with regression models, because the null hypothesis specifies that its conditional expectation given X is zero. Our general framework allows to deal with a wide range of parametric models considered in the econometric literature.

Example 1 *Our framework includes specification testing of models that jointly parameterize the conditional mean and the conditional variance of a dependent variable. Such models are defined as*

$$Y = \mu(X, \theta_0) + U, \quad E[U|X] = 0, \quad \text{a.s.} \quad E[U^2|X] = \omega^2(X, \theta_0) > 0 \quad \text{a.s.},$$

where $\mu(\cdot, \cdot)$ and $\omega^2(\cdot, \cdot)$ are known functions up to the value θ_0 , see e.g. Wooldridge (1990). The parametric model is completely defined through restrictions (1), where $Z = (Y', X')'$ and

$$\psi(Z, \theta) = \begin{pmatrix} Y - \mu(X, \theta) \\ [Y - \mu(X, \theta)]^2 - \omega^2(X, \theta) \end{pmatrix}.$$

Testing H_0 allows to check the full specification of the model. We may also be interested in testing only a subset of these restrictions. If we consider only restrictions relative to conditional mean, we deal with specification testing of a standard regression model, as studied by many authors. If we consider only the second set of restrictions, e.g. if we are sure about the functional form of the conditional mean, we entertain a test about the functional form of the conditional variance, as studied by Hong (1993). Finally, a particular application of our framework allows to test the null hypothesis of homoskedasticity by considering the specific restriction

$$E \left[(Y - \mu(X, \theta_0))^2 - E(Y - \mu(X, \theta_0))^2 | X \right] = 0 \quad \text{a.s.}$$

Example 2 Consider the model

$$\tau(Y, \lambda) = \mu(X, \beta) + U, \quad E[U|X] = 0 \quad \text{a.s.},$$

where $\mu(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ are known functions. When λ is unknown and $\tau(\cdot, \cdot)$ is a nonlinear transform, this model is not a regression model. Choices for $\tau(\cdot, \cdot)$ include the popular Box-Cox transformation, and the family of transformations defined by

$$\tau(Y, \lambda) = \kappa(\lambda Y)/\lambda, \quad \kappa(0) = 0, \quad \dot{\kappa}(0) = 1,$$

where $\dot{\kappa}(\cdot)$ denotes the first derivative of $\kappa(\cdot)$, see MacKinnon and Magee (1990). Simple instances of such transformations are $\kappa(y) = y^2 + y$ and $\kappa(y) = \text{arcsinh}(y)$, which have several advantages over the Box-Cox transform, see Burbidge, Magee and Robb (1988). When the residual's distribution is unknown, the parametric model is simply defined through

$$E[\tau(Y, \lambda_0) - \mu(X, \beta_0)|X] = 0 \quad \text{a.s.},$$

that is, through restrictions of the type (1). This allows to use any function of X as an instrumental variable (IV) for estimation purposes. Rejecting the validity of the conditional moment restrictions indicates that the functional form of the model is inadequate and would then invalidate the IV estimation method.

Example 3 Our framework further includes models defined through conditional moment restrictions given a set of instrumental variables, as considered by Newey (1990). An important example of such a model is one where $\psi(\cdot, \cdot)$ is a vector (or subvector) of residuals from a (possibly nonlinear) simultaneous equations system. As a benchmark, consider the simple equilibrium model

$$\begin{aligned} Q &= a_0 P + b_0 I + U, & a_0 < 0 & \quad (\text{Demand}) \\ Q &= \alpha_0 P + \beta_0 W + V, & \alpha_0 > 0 & \quad (\text{Supply}), \end{aligned}$$

where Q and P respectively denote quantity and price, I and W are equation-specific exogenous variables, and the error terms U and V are uncorrelated. Here $Z = (Q, P, I, W)'$ is the vector of all variables entering the model and $X = (I, W)'$ is the vector of exogenous variables. The model assumes that the error terms are unpredictable given the exogenous variables, i.e.

$$E[Q - a_0 P - b_0 I | X] = 0 \quad \text{a.s.} \quad \text{and} \quad E[Q - \alpha_0 P - \beta_0 W | X] = 0 \quad \text{a.s.}$$

Besides its economic meaning, this restriction ensures identification of the coefficients and is the basis for IV estimation methods either in a parametric context, see e.g. Davidson and MacKinnon

(1993), or in a nonparametric setup, see Darolles, Florens and Renault (1999). Our framework allows to test jointly the restrictions related to the demand and supply equations. Under the assumption that X is exogenous, rejecting this conditional moment restriction means that the postulated functional forms of the demand and supply curves are invalid.

As seen from the above examples, our framework goes far beyond testing the parametric specification of an univariate regression function and applies to testing econometric models defined by several conditional moment restrictions, which can be tested simultaneously or separately. Other examples can be considered, such as testing that the conditional score function is identically zero conditional in a fully parametric model, or that the information matrix equality holds conditionally upon the independent variables.

We now introduce some conditions upon the considered econometric model. To keep a great level of applicability, we formulate general assumptions that can accommodate various models and estimation methods. For the sake of simplicity, we focus on the case of cross-section data.

2.1 $\mathcal{Z}_n$ is an i.i.d. random sample from a random variable Z on $\mathbb{R}^s$. The subvector X admits a probability distribution function $F(\cdot)$ and a density function $f(\cdot)$.

2.2 Θ is compact in $\mathbb{R}^p$. There is an estimator θ_n that admits the first order asymptotic expansion $\theta_n = \theta^* + O_p(n^{-1/2})$, for some interior point θ^* of Θ . Under H_0 , $\theta^* = \theta_0$.

2.3 (i) $E \|\psi(Z, \theta^*)\|^2 < \infty$. For each component $\psi^{(k)}(\cdot, \cdot)$ of $\psi(\cdot, \cdot)$, $k = 1, \dots, m$, $\dot{\psi}^{(k)}(\cdot, \cdot) = \partial \psi^{(k)}(\cdot, \cdot) / \partial \theta$ and $\ddot{\psi}^{(k)}(\cdot, \cdot) = \partial^2 \psi^{(k)}(\cdot, \cdot) / \partial \theta \partial \theta'$ exist almost surely in an open neighborhood $\mathcal{N}(\theta^*)$ of θ^* , $E \|\dot{\psi}^{(k)}(Z, \theta^*)\| < \infty$ and $\sup_{\theta \in \mathcal{N}(\theta^*)} \|\ddot{\psi}^{(k)}(\cdot, \theta^*)\| < S(\cdot)$, with $E[S(Z)] < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. (ii) For each $k = 1, \dots, m$, $\ddot{\psi}^{(k)}(\cdot, \cdot)$ are continuous in θ for $\theta \in \mathcal{N}(\theta^*)$ and uniformly in Z .

Assumption 2.1 restricts our analysis to an iid context where the exogenous variables are continuous. Allowing some discrete components in X is not difficult, but would involve cumbersome notations.² Assumption 2.2 says that we have at hand an estimator θ_n that is $\sqrt{n}$ -consistent for a pseudo-true value θ^* of the parameter, which equals θ_0 under H_0 . This allows for several estimation methods, such as nonlinear least-squares, instrumental variables, generalized method of moments, or pseudo-maximum likelihood, see e.g. White (1994) and Gourieroux and Monfort (1995) for a general theory. For instance, given a suitable $(p \times m)$ matrix of instruments M , the pseudo-true value can be defined as the solution to the equation system

$$E[M(X)\psi(Z, \theta^*)] = 0, \quad (2)$$

²See Zheng (1996) or Su and Wei (1991) for details.

and θ_n is the solution to the sample analog of (2), that is, the solution to

$$\frac{1}{n} \sum_{i=1}^n M(X) \psi(Z, \theta) = 0. \quad (3)$$

Assumption 2.3 imposes some regularity on the functions entering the restrictions, see e.g. Newey (1985) or Robinson (1991) for similar conditions.

3 The two methodologies: rationale and asymptotics

3.1 Smooth tests

To test for the null hypothesis, we formulate the conditional moment restrictions as a completely equivalent unconditional moment restriction. Specifically, H_0 is equivalent to

$$H_0 : E [\psi' (Z, \theta_0) E (\psi (Z, \theta_0) | X) w_S (X)] = 0 \quad \text{a.s.}, \quad (4)$$

for any weight function $w_S(\cdot)$ is strictly positive almost surely onto the support of $f(X)$. This states the orthogonality between the generalized residuals $\psi(\cdot, \theta_0)$ and their conditional expectation $E(\psi(Z, \theta_0) | X = \cdot)$. Alternatively, one could weight each of the generalized residuals differently. A sample analog of the expectation in (4) is given by the infeasible estimator

$$\frac{1}{n} \sum_{i=1}^n \psi'(Z_i, \theta_0) E(\psi(Z_i, \theta_0) | X_i) w_S(X_i),$$

A feasible estimator thus requires on the one hand a consistent estimator of θ_0 , as given by θ_n , and on the other hand a nonparametric estimator of $E(\psi(Z, \theta_0) | X)$, such as a kernel estimator. A possible and natural choice for $w_S(\cdot)$ is the density $f(\cdot)$ itself. This is technically convenient in problems involving sums of kernel estimators, see e.g. Powell, Stock and Stoker (1989), Fan and Li (1996), Zheng (1996) and Lavergne and Vuong (2000) for an analogous device. Estimating $E(\psi(Z, \theta_0) | X = \cdot) f(\cdot)$ by kernel estimators leads to consider the statistic

$$T_n = \frac{1}{n(n-1)h^q} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \psi'(Z_i, \theta_n) \psi(Z_j, \theta_n) K_{ij}, \quad (5)$$

where $K_{ij} = K((X_i - X_j)/h)$, $K(\cdot) : \mathbb{R}^q \rightarrow \mathbb{R}$ is a kernel function and $h = h(n)$ is a positive bandwidth number. This statistic is easily computed from the estimated generalized residuals $\psi(Z_i, \theta_n)$. It resembles the statistic proposed in Fan and Li (1996) and Zheng (1996) for testing the specification of regression functions and it is constructed similarly, with the generalized vector residuals $\psi(Z_i, \theta_n)$ in place of standard regression residuals.

The asymptotic behavior of T_n depends on whether H_0 holds. Under misspecification, T_n is asymptotically consistent and efficient for the unconditional moment in (4) with the usual $\sqrt{n}$ -rate of convergence. But, under the null hypothesis, the $\sqrt{n}$ -asymptotic distribution of T_n is degenerate. This leads us to consider higher-order terms in its asymptotic expansion under H_0 . To state our formal result, we introduce some standard assumptions on the kernel function and the bandwidth parameter. We also impose the supplementary smoothness Assumption 2.6 stated in the proofs' section.

2.4 $K(\cdot)$ is even and bounded, integrates to one and $\lim_{\|u\| \rightarrow \infty} \|u\|^q |K(u)| = 0$.

2.5 $\lim_{n \rightarrow \infty} \left\{ h + (nh^q)^{-1} \right\} = 0$.

Theorem 1 Under H_0 and Assumptions 2.1 to 2.6,

$$nh^{q/2}T_n \xrightarrow{d} N(0, V), \text{ where } V = 2 \sum_{k=1}^m \sum_{l=1}^m E[\sigma_{kl}^2(X) f(X)] \int_{\mathbb{R}^q} K^2(u) du,$$

with $\sigma_{kl}(X) = E[\psi^{(k)}(Z, \theta^*) \psi^{(l)}(Z, \theta^*) | X]$.

To build a test statistic in practice, we need an estimator of the asymptotic variance of T_n under the null hypothesis. For instance, it can be estimated by

$$V_n = \frac{2}{n(n-1)} \frac{1}{h^q} \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1, j \neq i}^n \psi^{(k)}(Z_i, \theta_n) \psi^{(l)}(Z_i, \theta_n) \psi^{(k)}(Z_j, \theta_n) \psi^{(l)}(Z_j, \theta_n) K_{ij}^2.$$

The test statistic is defined as

$$t_n = nh^{q/2}T_n / V_n^{1/2}.$$

Corollary 2 Under Assumptions 2.1 to 2.6, $t_n \rightarrow_d N(0, 1)$ under H_0 and otherwise $(nh^{q/2})^{-1} t_n \rightarrow_p E[\psi'(Z, \theta^*) E[\psi(Z, \theta^*) | X] f(X)] / V^{1/2} > 0$.

The resulting test is a *one-sided* normal test and is consistent in the direction of any nonparametric alternative to H_0 . In particular, negative values of the test statistic can occur under the null hypothesis only.

Example 1 (continued) Given a consistent estimator θ_n of θ^* , e.g. a generalized nonlinear least-squares estimator, the generalized residuals are

$$\psi(Z_i, \theta_n) = \begin{pmatrix} Y_i - \mu(X_i, \theta_n) \\ [Y_i - \mu(X_i, \theta_n)]^2 - \omega^2(X_i, \theta_n) \end{pmatrix} = \begin{pmatrix} U_{ni} \\ U_{ni}^2 - \omega^2(X_i, \theta_n) \end{pmatrix}.$$

The statistic T_n writes

$$\begin{aligned} T_n &= T_{1n} + T_{2n} \\ &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \{U_{ni}U_{nj} + [U_{ni}^2 - \omega^2(X_i, \theta_n)] [U_{nj}^2 - \omega^2(X_j, \theta_n)]\} K_{ij}. \end{aligned}$$

The asymptotic variance estimator is computed as

$$\begin{aligned} V_n &= \frac{2}{n(n-1)h^q} \sum_{i \neq j} \{U_{ni}^2 U_{nj}^2 + [U_{ni}^2 - \omega^2(X_i, \theta_n)]^2 [U_{nj}^2 - \omega^2(X_j, \theta_n)]^2 \\ &\quad + 2U_{ni} [U_{ni} - \omega^2(X_i, \theta_n)] U_{nj} [U_{nj} - \omega^2(X_j, \theta_n)]\} K_{ij}^2. \end{aligned}$$

Testing only the specification of the conditional mean is based upon T_{1n} , as proposed by Zheng (1996) and Li and Wang (1998). Testing only the specification of the conditional variance can be based upon T_{2n} . Our joint specification test of conditional mean and variance is simply the addition of the two latter statistics.

Among the problems related to the practical implementation of our test is the choice of the bandwidth parameter. First, though our general theory is developed for a generic bandwidth, our test can be readily extended to vanishing individual bandwidths h_j for each conditioning variable $X^{(j)}$. The rate of convergence of T_n is then $n \prod_{j=1}^q h_j^{1/2}$. Second, a data-dependent bandwidth is often used in practice. This allows to adapt smoothing to the variability of $X^{(j)}$, by considering for instance $h_j = h \times s_j$ where s_j^2 is the empirical variance of $X^{(j)}$. The following theorem shows that this can be done without affecting the properties of our asymptotic test.

Theorem 3 *If $K(\cdot)$ is differentiable, with bounded partial derivatives on its support, Corollary 1 extends to the case of random $\hat{h}$, if there exists a deterministic h that fulfills the assumptions of Corollary 1 such that $(\hat{h} - h)/h = o_p(1)$.*

3.2 Integral-transform tests

Another way to construct a test for H_0 is to replace the conditional moment restrictions by an equivalent infinite set of unconditional moment restrictions that are easier to deal with than the original ones. Indeed, the null hypothesis of interest is equivalent to

$$H_0 : E[\psi(Z, \theta_0) w_P(X, x)] = \mathbf{0}, \quad \forall x \in \mathbb{R}^q, \quad (6)$$

where $w_P(\cdot, \cdot)$ is a properly chosen weight function. Stinchcombe and White (1998) show that there exists a wide class of functions that generate consistent tests, including among others the logistic cumulative distribution function or the exponential weight function employed by Bierens

(1982). Here, we use the computationally convenient step function $w_P(X, x) = \mathbf{1}(X \leq x) = \prod_{j=1}^q \mathbf{1}(X^{(j)} \leq x^{(j)})$, where $\mathbf{1}(\cdot)$ is the indicator function. Previous work where a similar choice is made includes Andrews (1997), Stute (1997), Koul and Stute (1999) and Whang (2000).

It is convenient to consider the expectations in (6) because they are easily estimated by sample analogs. In the case where $w_P(X, x) = \mathbf{1}(X \leq x)$, the estimator is

$$R_n(x) = \frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta_n) \mathbf{1}(X_i \leq x), \quad (7)$$

which is an empirical process of dimension m marked by the generalized residuals $\psi(Z_i, \theta_n)$. A test statistic for (6) can then be any well-chosen continuous functional of $R_n(\cdot)$. For instance, one may consider a statistic of the form

$$n \int_{\mathbb{R}^q} R'_n(x) R_n(x) d\nu(x).$$

where $\nu(\cdot)$ is an arbitrary probability measure, as done by Bierens and Ploberger (1997). In our formal analysis, we specifically choose a simple and natural measure as the empirical distribution function of the X_i 's. Hence, we consider the Cramer-von-Mises type statistic

$$c_n = n \int_{\mathbb{R}^q} R'_n(x) R_n(x) dF_n(x) = \sum_{i=1}^n R_n(X_i)' R_n(X_i),$$

where $F_n(\cdot)$ is the empirical distribution function of X . A Kolmogorov-Smirnov type test can also be constructed. We focus here on the Cramer-von-Mises test, as Whang (2000) reports better performances for this type of test in a regression context.

We first provide a functional central limit theorem for $R_n(\cdot)$ under the following assumption.

2.7 $\theta_n = \theta^* + \frac{1}{n} \sum_{i=1}^n \ell(Z_i, \theta^*) + o_p(n^{-1/2})$, where the function $\ell(\cdot, \cdot)$ is such that $E[\ell(Z, \theta^*)] = 0$ and $E[\ell(Z, \theta^*) \ell'(Z, \theta^*)]$ exists.

When θ_n is defined as the solution to (3), then $\ell(Z, \theta^*) = \left(E \left[M(X) \dot{\psi}(Z, \theta^*) \right] \right)^{-1} M(X) \psi(Z, \theta^*)$, where $\dot{\psi} = \left(\dot{\psi}^{(1)'} , \dots, \dot{\psi}^{(m)'} \right)'$. Define

$$r_i(x, \theta) = \psi(Z_i, \theta) \mathbf{1}(X_i \leq x) + E \left[\dot{\psi}(Z, \theta) \mathbf{1}(X \leq x) \right] \ell(Z_i, \theta). \quad (8)$$

Theorem 4 Under H_0 and Assumptions 2.1, 2.3(i) and 2.7,

$$n^{1/2} R_n \Rightarrow R_\infty,$$

where $\Rightarrow$ denotes weak convergence in the Skorokhod space $\times_{k=1}^m D[-\infty, \infty]^q$ and R_∞ is a Gaussian process centered at zero with covariance structure $\Omega(x, s) = E[r_1(x, \theta_0) r_1'(s, \theta_0)]$, $\forall x, s \in \mathbb{R}^q$.

The following corollary gives the asymptotic behavior of our statistic c_n .

Corollary 5 *Under Assumptions 2.1, 2.3(i) and 2.7,*

$$c_n \xrightarrow{d} c_\infty = \int_{\mathbb{R}^q} R_\infty(x) R'_\infty(x) dF(x) \text{ under } H_0,$$

$$n^{-1}c_n \rightarrow_p \int_{\mathbb{R}^q} E[\psi(Z, \theta_0) \mathbf{1}(X \leq x)]' E[\psi(Z, \theta_0) \mathbf{1}(X \leq x)] dF(x) > 0, \quad \text{otherwise.}$$

Therefore, an asymptotic test can be based on the statistic c_n , as soon as one can compute critical values. Because the asymptotic distribution of c_n is known only in special cases, see e.g. Delgado (1993), we will propose in the next section a simulation method to approximate it. The resulting test is then a *one-sided* normal test and is consistent against any nonparametric alternative to H_0 .

Example 1 (continued) *Recall that $U_{ni} = Y_i - \mu(X_i, \theta_n)$. The statistic c_n is then equal to*

$$\begin{aligned} c_n &= c_{1n} + c_{2n} \\ &= \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1}^n U_{nj} \mathbf{1}(X_j \leq X_i) \right)^2 + \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1}^n [U_{nj}^2 - \omega^2(X_j, \theta_n)] \mathbf{1}(X_j \leq X_i) \right)^2. \end{aligned}$$

Testing only the specification of the conditional mean leads to consider c_{1n} , which is exactly the statistic proposed by Stute (1997) for univariate regression models and extended by Whang (2000) to a multivariate context. Testing only the specification of the conditional variance is based upon c_{2n} . Our test statistic for testing the specification of both conditional mean and variance is the sum of the two latter statistics.

3.3 Variations on two themes

It should be noted that a number of different valid test statistics can be built. First, many alternative equivalent formulations of the null hypothesis could be considered by replacing each of the generalized residuals $\psi^{(k)}(\cdot, \cdot)$ by $a_k \psi^{(k)}(\cdot, \cdot)$, for given constants $a_k, k = 1, \dots, m$. The null hypothesis (4) for smooth tests would then write

$$\sum_{k=1}^m a_k^2 E \left[\psi^{(k)}(Z, \theta_0) E \left(\psi^{(k)}(Z, \theta_0) \mid X \right) w_S(X) \right] = 0 \quad \text{a.s.}$$

Thus, the sum of squares generalized residuals $\psi'(\cdot, \theta_n) \psi(\cdot, \theta_n) = \sum_{k=1}^m [\psi^{(k)}(\cdot, \theta_n)]^2$ would be replaced in T_n by a weighted sum of squares $\sum_{k=1}^m [a_k \psi^{(k)}(\cdot, \theta_n)]^2$, and similarly, each of the $\psi^{(k)}(\cdot, \theta_n)$

would be changed into $a_k \psi^{(k)}(\cdot, \theta_n)$ in the expressions of V and V_n . The null hypothesis (6) for integral-transform statistics would write

$$a_k E \left[\psi^{(k)}(Z, \theta_0) w_P(X, x) \right] = \mathbf{0}, \quad \forall x \in \mathbb{R}^q, \quad \forall k = 1, \dots, m$$

and the statistic c_n would be changed accordingly. More generally, one could adopt different weighting functions $w_S(X)$ or $w_P(X, x)$ in the formulation of the null hypothesis. Whether there exists some optimal weighting is an open question and has not been investigated even for testing regression models.

Second, as easily seen from our Example 1, one could consider testing (4) by smooth tests by standardizing first each of the components of T_n , T_{1n} and T_{2n} in our example, and then determining the asymptotic distribution of the sum of the standardized individual statistics. Yet another approach could be to show that the joint distribution of (T_{1n}, T_{2n}) is asymptotically normal and to construct a normalized statistic with an asymptotic χ^2 distribution. This approach is adopted in Li (1999) when testing portfolio mean-variance efficiency. However, such a test ignores the one-sided nature of the testing problem, and thus has undesirable properties. Indeed, it rejects the null hypothesis whenever T_{1n} and T_{2n} take large negative values, which can occur under H_0 only. As a consequence, it is also expected to have inferior power properties.

While these variants lead to statistics with analogous properties than the ones that we have studied, they should be compared on their respective power. This is a quite intricate question, because one would like to consider nonparametric alternatives of various forms, and the power of each form of test can depend on the specific form of the alternatives considered. Moreover, smooth tests and integral-transform tests have their own strengths and drawbacks, depending upon on the theoretical approach for comparing them, see Fan and Li (2000) and Guerre and Lavergne (2002). Hence, we leave this issue for future research.

4 Simulation-based critical values

On the one hand, the practical implementation of the asymptotic test from Section 3.1 involves some difficulties, as the asymptotic approximation of the null distribution can be slow, depending upon the chosen bandwidth and the number of exogenous variables in the model, see e.g. Härdle and Mammen (1993). On the other hand, a procedure for computing critical values of the c_n statistic is needed, because its asymptotic null distribution is case-dependent. In the following, we first briefly review the methods proposed for computing critical values for specification tests of regression functions, and explain the difficulties that arise for extending these methods to testing general conditional moment restrictions. We then propose a simple simulation approach that yields

asymptotically valid critical values for our tests.

4.1 Methods for approximating critical values

The integral-transform statistic raises the fundamental problem that its limiting distribution under the null hypothesis depends on the unknown data-generating process, and therefore cannot not be tabulated in general. One way of solving this problem is to derive case-independent upperbounds of the asymptotic critical values, as proposed by Bierens and Ploberger (1997). Another possibility consists of transforming the marked empirical process $R_n(\cdot)$, as suggested by Stute, Thies and Zhu (1998) and Koul and Stute (1999). Indeed, in a regression context with $q = 1$, if the errors are independent of X , it is possible to perform a transformation S_n of R_n depending on the data, such that

$$S_n \circ R_n \Rightarrow B,$$

where B denotes the standard Brownian motion. As Stute, Thies and Zhu (1998) show, this transformation is also feasible when the errors are conditionally heteroskedastic, using nonparametric estimators of the conditional variance. However, when $q > 1$, the proposed transformation is no longer asymptotically pivotal, even when the parameters are known. Hence, the implementation of such transformations in our context would require too restrictive assumptions and would be rather involved for practical purposes.

For smooth specification tests of regression models, wild bootstrap procedures have been proposed to compute accurate small sample critical values. When $\psi(Z, \theta) = Y - m(X, \theta)$, with a scalar Y and a known function $m(\cdot, \theta)$, Härdle and Mammen (1993) propose to generate external bootstrap resamples from a model that fulfills the null hypothesis, which is necessary to obtain valid critical values. A typical resample $\mathcal{Z}_n^*$ is thus obtained as $\{(Y_i^*, X_i), i = 1, \dots, n\}$, where $Y_i^* = m(X_i, \theta_n) + U_{ni}^*$, with $U_{ni}^* = U_{ni}\zeta_i$ and

B1 The ζ_i 's are iids, independent of $\mathcal{Z}_n$, with zero mean and unit variance.

Bootstrap critical values at level α are obtained by (i) generating several resamples $\mathcal{Z}_n^*$ and computing the corresponding test statistic t_n^* for each resample, (ii) computing the empirical quantile of order $(1 - \alpha)$ of the statistics t_n^* . This gives a critical value to be compared to the original test statistic t_n . For a given α , the estimation accuracy of the critical value depends upon the number of simulated statistics.

Such a wild bootstrap procedure is also applicable to statistics based on integral transforms of a regression, see Stute, González-Manteiga and Presedo (1997) and Whang (2000). However, it can be extremely difficult to apply in our context for practical and theoretical reasons. In practice,

the number of generalized residuals may be different than the number of endogenous variables. For instance, in Example 1, when one wants to jointly test for the parametric specification of the conditional mean and the conditional variance, there are two generalized residuals for only one endogenous variable. On the other hand, if one wants to consider a single equation coming from a simultaneous equation system, as in Example 3, there is less generalized residuals than endogenous variables. Moreover, the model can be nonlinear in the endogenous variables, as in a the transformation model of Example 2 or a nonlinear simultaneous equation model, so that a simple reduced form for the endogenous variables may not be available. From a theoretical viewpoint, non-linearity of the model in endogenous variables complicates a lot the technical analysis. Furthermore, it may not be easy in that case to find an external distribution which mimics what would happen under the null hypothesis. Similar difficulties are also encountered when bootstrapping parametric tests based on Generalized Method of Moments estimators, see Hall and Horowitz (1996).

Because it seems difficult to propose a generalization of the wild bootstrap that would be valid in any case, we adapt a simple simulation technique proposed by Hansen (1996) and also used by de Jong (1996) and Chen and Fan (1999) for integral-transform type statistics. This method is simple and always applicable, but is likely not to be as good as one tailored for a specific model.

4.2 Smooth tests

Because the statistic T_n is a function of $\mathcal{Y}_n = \{(\psi(Z_i, \theta_n), X_i), i = 1, \dots, n\}$ only, we directly simulate resamples as $\mathcal{Y}_n^* = \{(\psi_{ni}^*, X_i), i = 1, \dots, n\}$, where $\psi_{ni}^* = \psi(Z_i, \theta_n) \zeta_i$ and the ζ_i 's satisfy Assumption B1. This gives us statistics of the form

$$T_n^* = \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_{ni}^* \psi_{nj}^* K_{ij} = \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi'(Z_i, \theta_n) \psi(Z_j, \theta_n) \zeta_i \zeta_j K_{ij}.$$

The simulated version of the test statistic is then

$$t_n^* = \frac{nh^{q/2} T_n^*}{V_n^{1/2}}.$$

Critical values are computed as the corresponding empirical quantiles for the set of simulated statistics. This gives a critical value to be compared to the original test statistic t_n . The next theorem justifies the validity of this procedure.

Theorem 6 *Under the assumptions of Theorem 1 and Assumption B1,*

$$\sup_t |\Pr(t_n^* \leq t | \mathcal{Z}_n) - \Phi(t)| = o_p(1),$$

where $\mathcal{Z}_n = \{Z_i, i = 1, \dots, n\}$ and $\Phi(\cdot)$ is the standard normal distribution.

Whether much accuracy is gained with respect to the asymptotic approximation depends upon several features. For the linear regression model, i.e. $\psi(Z_i, \theta_n) = Y_i - X_i' \theta$, and when θ_n is the ordinary least squares (OLS) estimator, this simulation method is equivalent to a wild bootstrap procedure, and then has the same properties. In that case, Li and Wang (1998) show that moments up to order four are accurately matched by the wild bootstrap under the supplementary assumption $E(\zeta_1^3) = 1$.³ Fan and Linton (1999) provide Edgeworth's expansions for the distribution of a similar test statistic in a regression context, assuming symmetric errors. In general, however, the performances of our simulation technique may depend upon the form of the moment restrictions, the estimation technique and the distribution of the ζ_i 's.

4.3 Integral-transform tests

The procedure used for the smooth test cannot be directly applied to the statistic c_n . This is because a version of $R_n(\cdot)$ based upon $\mathcal{Y}_n^*$ would not mimic its covariance structure under the null hypothesis. However, Theorem 3 shows that $R_n(\cdot)$ asymptotically depends only upon the $r_i(x, \theta)$. This suggests to consider the statistic

$$c_n^* = \sum_{i=1}^n \bar{R}_n^*(X_i)' \bar{R}_n^*(X_i), \quad \text{where} \quad \bar{R}_n^*(x) = \frac{1}{n} \sum_{i=1}^n r_{ni}^*(x) \quad \text{and} \quad r_{ni}^*(x) = r_i(x, \theta_n) \zeta_i.$$

It is easy to see that $n^{1/2} \bar{R}_n^*(\cdot)$ is, conditional on $\mathcal{Z}_n$, centered at zero and with covariance structure $\Omega_n(x, s) = \frac{1}{n} \sum_{i=1}^n r_i(x, \theta_n) r_i'(s, \theta_n)$, which converges to the covariance structure of $R_\infty(\cdot)$ under H_0 .

In practice, computation of $r_i(x, \theta_n)$ requires to evaluate $\ell(Z_i, \theta_n)$, which is unknown in general, but can be adequately approximated for usual estimators. For instance, if θ_n is the solution to (3) then $\ell(Z_i, \theta_n)$ can be replaced by $L_n^{-1} M(X_i) \psi(Z_i, \theta_n)$, where $L_n = n^{-1} \sum_{i=1}^n M(X_i) \dot{\psi}(Z_i, \theta_n)$. Critical values for testing H_0 at level α are obtained as for the smooth test. Note that, when $\psi(Z, \theta) = Y - X' \theta$ and θ_n is the OLS estimator, c_n^* can be simply computed from a resample $\mathcal{Z}_n^* = \{(Y_i^*, X_i), i = 1, \dots, n\}$, where $Y_i^* = X_i' \theta_n + \psi_{ni}^*$, which is the method proposed by Stute et al. (1998) and Whang (2000).

Theorem 7 *Under the assumptions of Theorem 3 and 2.8 (see Proof Section), if $|\zeta_1| \leq c < \infty$ for some $c > 0$,*

$$\sup_t |\Pr(c_n^* \leq t | \mathcal{Z}_n) - \Pr(c_\infty \leq t)| = o_p(1),$$

where c_∞ is as in Corollary 2 under H_0 , and otherwise similarly defined but depending upon θ^* in place of θ_0 .

³This is to match the skewness terms in the Edgeworth's expansions of the statistic distribution and the conditional bootstrap statistic distribution.

5 Monte Carlo Results

We investigate the small sample behavior of the tests considered in this paper within the setup of our Examples 1 and 2. Note that, even in the case of testing a regression function, there is little evidence in the econometric literature on the comparative performances of the different approaches. A notable exception is Whang (2000), who compares Kolmogorov-Smirnov and Cramer-von-Mises tests to the tests of Härdle and Mammen (1993) and Bierens and Ploberger's (1997).

Testing for linearity of a regression function We first consider testing for the linear specification of a regression function. In this case, the smooth test is identical to the test proposed by Zheng (1996) and Li and Wang (1996), while the integral-transform test is the one studied by Stute (1997). The null hypothesis of interest is

$$H_0 : \exists (\alpha_0, \beta_0) : E[Y - \alpha_0 - \beta_0 X \mid X] = 0 \quad \text{a.s.}$$

The data generating process is chosen as

$$Y_i = 1 + 2X_i + \sin(\delta\pi X_i) + u_i, \quad i = 1, \dots, n,$$

where the u_i 's are iids $N(0, 1)$ and independently distributed of the X_i 's which are iids $N(0, 1)$. The null hypothesis corresponds to $\delta = 0$ and is denoted by DGP_0 . We investigate three alternatives denoted as DGP_δ for $\delta = 1, 2$ and 3 . By increasing δ , we obtain higher frequency alternatives that are more difficult to distinguish from pure noise. This allows us to observe large variations in the behavior of our tests.

In each case, two different sample sizes, $n = 50$ and 100 , are considered. For each experiment, i.e. each data generating process and sample size, we run 2000 simulations. For each simulation, the critical value is estimated using 500 bootstrap replications. The parameter vector (α_0, β_0) is estimated by ordinary least-squares. For the test based on smoothers, we choose the bandwidth following the rule-of-thumb $h = dn^{-1/5}$, with d varying in the grid $\{0.025, 0.05, 0.1, 0.5, 1, 1.5, 2\}$. The results of each experiment are reported on a graph that show the empirical rejection probabilities for the three tests at nominal level 5% with respect to d . The solid line corresponds to the rejection probability for the test based upon c_n (which does not depend upon d), the grey solid line is the rejection probability for the smooth test based upon t_n using bootstrap critical values, and the dashed line is the rejection probability for the smooth test using asymptotic critical values.

Figures 1 and 2 report the result of DGP_0 for the two sample sizes $n = 50$ and 100 respectively. In both cases, the test based on nonparametric estimation is too conservative. This is because the test statistic is negatively biased in small samples, as already noted by Li and Wang (1998).

However, the empirical size of the test becomes closer to its nominal size when the sample size increases. It is also seen that use of bootstrap critical values does improve upon the asymptotic critical values, but generally not as much as required. The empirical level for Stute's test is very close to its nominal level even for a small sample size of 50.

Results for alternative hypotheses DGP_1 to DGP_3 are reported in Figures 3 to 8. For the low frequency alternative DGP_1 , Stute's test has high power even for a sample size of 50 observations, and its power is one when $n=100$. When the frequency of the alternative increases, the rejection probability for Stute's test rapidly decreases and is less than 20% for DGP_3 for a sample size of 100. As expected, power of the smooth test varies with the frequency of the alternative and the bandwidth. The first phenomenon has already been noted in some other contexts, see e.g. Lavergne and Vuong (2000). A small bandwidth always corresponds a low rejection frequency, while a large one may yield low or high power depending upon the alternatives. Moreover, the smooth test has high power for a range of bandwidths that narrows as δ increases. However, a bandwidth close to 0.5 gives high power for each experiment in our study. As is the case under the null hypothesis, using bootstrap critical values does lead to some but limited improvement in most cases. Compared to Stute's test, the smooth test has higher or equivalent power for a large range of bandwidths.

Testing jointly for linearity of the regression function and for homoscedasticity We now consider a similar regression model, where one aims to test jointly for the specification of the regression function and for homoskedasticity. The null hypothesis writes

$$H_0 : \exists (\alpha_0, \beta_0, \sigma_0^2) : \begin{pmatrix} E[Y - \alpha_0 - \beta_0 X | X] \\ E[(Y - \alpha_0 - \beta_0 X)^2 - \sigma_0^2 | X] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{a.s.}$$

We consider the design

$$Y_i = \alpha_0 + \beta_0 X_i + \delta u_i(1 + X_i^2)/\sqrt{5}, \quad i = 1, \dots, n,$$

where the u_i 's and the X_i 's are generated as before. The null hypothesis corresponds to $\delta = 0$ and is denoted DGP'_0 . The alternative DGP'_1 corresponds to $\delta = 1$. Two sample sizes are considered, $n = 100$ and 250. Other details are otherwise similar than in the previous Monte-Carlo study. The results are reported in Figures 9 to 12. Under the null hypothesis, the tests essentially exhibits the same features as when testing for a linear regression only. That is, the smooth test is undersized and it makes little difference to use bootstrap critical values, while the test based upon c_n has an empirical size close to the nominal one. Under the alternative hypothesis, there is much more difference between the two tests. The smooth test is more powerful than the integral-transform test for all bandwidths, except a very small one ($d = 0.025$). However, power is larger than 70% for the integral-transform based test when the sample size equals 250.

Testing the transformation model We now consider a well-studied model with a nonlinear transformation in the endogenous variable, namely the arcsinh transformation. The hypothesis of interest here writes

$$H_0 : \exists (\lambda_0, \alpha_0, \beta_0) : E \left[\frac{\text{arcsinh}(\lambda_0 Y)}{\lambda_0} - \alpha_0 - \beta_0 X \mid X \right] = 0 \quad \text{a.s.}$$

We consider the design

$$\frac{\text{arcsinh}(2Y_i)}{2} = 1 + 2X_i + \sin(\delta\pi X_i) + u_i, \quad i = 1, \dots, n,$$

where the u_i 's and the X_i 's are generated as before, except that the variance of the error term is chosen as 0.5. The parameters are estimated by a *GMM* estimator, with vector of instruments $M(X) = (1, X, X^2)'$, where the identity matrix has been chosen to weight the discrepancy conditions in the estimating objective function. Two sample sizes are considered, $n = 100$ and 250 . The notation DGP_δ'' denotes the model with $\delta = 0, 1, 2, 3$. Other details are otherwise similar than in the previous studies. The results are reported in Figures 13 to 20.

Under the null hypothesis (Figures 13 and 14), the only novel feature compared to the previous cases is that the integral-transform test is now undersized for the smaller sample size. Under the alternative hypothesis DGP_1'' (Figures 15 and 16), the power of the smooth test is essentially one over the whole range of considered bandwidths, while the power of its competitor is around 20% for $n = 100$ and increases to 50% when $n = 250$. Under the alternative hypothesis DGP_2'' (Figures 17 and 18), the smooth test behaves quite similarly, while the power of its competitor increases up to 95% for the larger sample size. For the higher frequency alternative DGP_3'' (Figures 19 and 20), the test based on c_n has essentially no power for $n = 100$, but its rejection probability reaches 70% when $n = 250$. The smooth test's power is not very sensitive to the bandwidth choice as soon as it is not too large, i.e. for d less than one.

Our limited set of simulations sheds light on their comparative behavior of the tests. For testing the specification of an univariate regression function, the integral-transform test has good size. Its power is quite high against an alternative of low frequency even for a small sample size, but it deteriorates against an alternative of high frequency. The smooth test is undersized and can be outperformed by its competitor, but seems well adapted to alternatives of varying frequencies. In other testing situations, the differences can be more marked. While the smooth test remains undersized, it seems that it can deal more easily with nonlinearity in the endogenous variable.

6 Conclusion

We have shown in this paper how the two methodologies used for testing the specification of regression function can be extended to testing a general set of conditional moment restrictions,

which can prove useful for many econometric models. Several problems warrant further research. First, it seems important to determine whether there is some optimal and feasible way to combine the conditional moment restrictions, as discussed in Section 3.3. Second, for the smooth test, it would be helpful to have some data-driven methods for bandwidth's choice, as investigated by Horowitz and Spokoiny (2001) and Guerre and Lavergne (2002) in the regression context. Third, it is clear that procedures for computing critical values are crucial for practical implementation. We have explained the difficulties that arise in generalizing bootstrap methods proposed in a regression context. Some unreported simulations results suggest that more sophisticated methods can substantially improve upon the simple simulation technique we use here. However, a general formal theory is still required in this respect.

7 Proofs

Definition 1 G^α , $\alpha > 0$, is the class of functions $g(\cdot) : \mathbb{R}^q \rightarrow \mathbb{R}$ such that $\exists \rho > 0$ such that for all $z \in \mathbb{R}^q$, $\sup_{\|y-z\| \leq \rho} |g(y) - g(z)| / \|y - z\| \leq G(z)$ and $g(\cdot)$ and $G(\cdot)$ have finite α -th moments (or are bounded if $\alpha = +\infty$).

Let $\alpha_k(x) = E \left[\left| \psi^{(k)}(Z, \theta^*) \right| \mid X = x \right]$, $\gamma_k(x) = E \left[\dot{\psi}^{(k)}(Z, \theta^*) \mid X = x \right]$, $\sigma_k^4(x) = E \left[\left(\psi^{(k)}(Z, \theta^*) \right)^4 \mid X = x \right]$, $\sigma_{kl}(x) = E \left[\psi^{(k)}(Z, \theta^*) \psi^{(l)}(Z, \theta^*) \mid X = x \right]$ for $k, l = 1, \dots, m$, and let $\lambda(x) = E[S(Z) \mid X = x]$. The next assumption summarizes the smoothness conditions on the different nonparametric functions.

2.6 $f(\cdot) \in G^\infty$. For all $k, l = 1, \dots, m$, $\sigma_{kl}(\cdot) \in G^4$, each element of $\gamma_k(\cdot)$ belong to $G^{8/3}$, $\sigma_k^4(\cdot)$, $\alpha_k(\cdot)$ and $\lambda(\cdot)$ belong to G^2 and $E \left[\left\| \dot{\psi}^{(k)}(Z, \theta^*) \right\|^{8/3} \right] < \infty$.

Proof of Theorem 1 Henceforth, $\psi_i^{(k)} = \psi^{(k)}(Z_i, \theta_0)$, $\hat{\psi}_i^{(k)} = \psi^{(k)}(Z_i, \theta_n)$, $\dot{\psi}_i^{(k)} = \dot{\psi}^{(k)}(Z_i, \theta_0)$, $\ddot{\psi}_i^{(k)}(\theta) = \ddot{\psi}^{(k)}(Z_i, \theta)$, for $i = 1, 2, \dots, n$ and $k = 1, \dots, m$, and $\sum_{i \neq j}$ stand for $\sum_{i=1}^n \sum_{j=1, j \neq i}^n$. We have

$$T_n = \frac{1}{n(n-1)} \sum_{i \neq j} D_n(Z_i, Z_j) + \sum_{k=1}^m \left(2T_{1n}^{(k)} + T_{2n}^{(k)} \right), \quad (9)$$

with $D_n(Z_i, Z_j) = \sum_{k=1}^m \psi_i^{(k)} \psi_j^{(k)} h^{-q} K_{ij}$,

$$\begin{aligned} T_{1n}^{(k)} &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{(k)} \left(\hat{\psi}_j^{(k)} - \psi_j^{(k)} \right) K_{ij}, \\ T_{2n}^{(k)} &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \left(\hat{\psi}_i^{(k)} - \psi_i^{(k)} \right) \left(\hat{\psi}_j^{(k)} - \psi_j^{(k)} \right) K_{ij}. \end{aligned}$$

We prove in a first step that $T_{1n}^{(k)} = O_p(n^{-1})$ and $T_{2n}^{(k)} = O_p(n^{-1})$ for all $k = 1, \dots, m$. Using a mean value theorem argument,

$$\begin{aligned} T_{1n}^{(k)} &= (\theta_n - \theta)' S_{1n} + (\theta_n - \theta_0)' S_{2n} (\theta_n - \theta_0), \\ T_{2n}^{(k)} &= (\theta_n - \theta_0)' S_{3n} (\theta_n - \theta_0), \end{aligned}$$

$$\begin{aligned} \text{where } S_{1n} &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{(k)} \dot{\psi}_j^{(k)} K_{ij}, \\ S_{2n} &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{(k)} \ddot{\psi}_j^{(k)} (\bar{\theta}_n) K_{ij}, \\ S_{3n} &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \left[\dot{\psi}_i^{(k)} + \ddot{\psi}_i^{(k)} (\bar{\theta}_n) (\theta_n - \theta_0) \right] \left[\dot{\psi}_j^{(k)} + \ddot{\psi}_j^{(k)} (\bar{\theta}_n) (\theta_n - \theta_0) \right]' K_{ij}, \\ &\quad \text{with } \|\bar{\theta}_n - \theta_0\| \leq \|\theta_n - \theta_0\| \text{ and } \|\tilde{\theta}_n - \theta_0\| \leq \|\theta_n - \theta_0\|. \end{aligned}$$

To study S_{1n} , we use the following lemma.

Lemma 8 (Powell, Stock and Stoker, 1989) *Let $U_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n H_n(Z_i, Z_j)$ be a U -statistic with symmetric kernel $H_n(Z_i, Z_j)$ and let the Z_i 's be iid. Let $q_n(Z_i) = E[H_n(Z_i, Z_j) | Z_i]$, $\bar{q}_n = E(q_n(Z_i))$. If $E(\|H_n(Z_1, Z_2)\|^2) = o(n)$, then $U_n = \bar{q}_n + (2/n) \sum_{i=1}^n [q_n(Z_i) - \bar{q}_n] + o_p(n^{-1/2})$.*

The quantity S_{1n} is a U -statistic with kernel $H_n(Z_1, Z_2) = h^{-q} (\psi_1^{(k)} \dot{\psi}_2^{(k)} + \psi_2^{(k)} \dot{\psi}_1^{(k)}) K_{12}/2$ and

$$\begin{aligned} &E(\|H_n(Z_1, Z_2)\|^2) \\ &= \frac{1}{h^{2q}} E\left[\left(\dot{\psi}_1^{(k)}\right)^2 \sigma_{kk}(X_2) K_{12}^2\right] = \frac{1}{h^q} E\left[\left(\dot{\psi}_1^{(k)}\right)^2 \int \sigma_{kk}(X_1 + hu) f(X_1 + hu) K^2(u) du\right] \\ &= \frac{1}{h^q} E\left[\left(\dot{\psi}_1^{(k)}\right)^2 \sigma_{kk}(X_1) f(X_1)\right] \int K^2(u) du + o\left(\frac{1}{h^q}\right) = O\left(\frac{1}{h^q}\right) = o(n) \end{aligned}$$

by Assumptions 2.4 to 2.6 together with Hölder's inequality. As $E[H_n(Z_i, Z_j)] = 0$, Lemma 8 implies that $S_{1n} = (2/n) \sum_{i=1}^n q_n(Z_i) + o_p(n^{-1/2})$, where $q_n(Z_i) = E[H_n(Z_i, Z_j) | Z_i]$. Moreover

$$E[q_n^2(Z)] = E[\sigma_{kk}(X) \gamma_k^2(X) f(X)] \left[\int K(u) du \right]^2 + o(1) = O(1),$$

so that $S_{1n} = O_p(n^{-1/2})$. For S_{2n} , we have

$$E|S_{2n}| \leq \frac{1}{h^q} E\{S(Z_2) E[\alpha_k(X_1) | K_{12} | Z_2]\} = E\{S(Z_2) \alpha_k(X_2) f(X_2)\} \int |K(u)| du + o(1) = O(1).$$

Hence, $S_{2n} = O_p(1)$. Similarly, one can show that $S_{3n} = O_p(1)$. These results imply that $T_{1n}^{(k)}$ and $T_{2n}^{(k)}$ are both $O_p(n^{-1})$, as $\theta_n - \theta^* = O_p(n^{-1/2})$ by Assumption 2.2.

We now determine the asymptotic distribution of the first term in (9). We use of the following result for degenerate U -statistics.

Lemma 9 (Hall, 1984) Let U_n be as in Lemma 8, with $E[H_n(Z_i, Z_j) | Z_i] = 0$ a.s. Let $G_n(Z_1, Z_2) = E[H_n(Z_3, Z_1)H_n(Z_3, Z_2) | Z_1, Z_2]$.

$$\text{If } \lim_{n \rightarrow \infty} \frac{EG_n^2(Z_1, Z_2) + n^{-1}EH_n^4(Z_1, Z_2)}{E^2H_n^2(Z_1, Z_2)} = 0, \quad \text{then } \frac{nU_n}{2E^{1/2}H_n^2(Z_1, Z_2)} \xrightarrow{d} N(0, 1).$$

The first term of (9) is a degenerate U -statistic with kernel $D_n(\cdot, \cdot)$, and the corresponding $G_n(\cdot, \cdot)$ is such that $E[G_n(Z_1, Z_2)^2] = \sum_{k=1}^m \sum_{l=1}^m \sum_{k'=1}^m \sum_{l'=1}^m \lambda_{kk' ll'}$, where

$$\begin{aligned} \lambda_{kk' ll'} &= h^{-4q} E \{ \sigma_{kk'}(X_1) \sigma_{ll'}(X_2) E[K_{13}K_{23}\sigma_{kl}(X_3) | X_1, X_2] E[K_{13}K_{23}\sigma_{k'l'}(X_3) | X_1, X_2] \} \\ &= \frac{1}{h^{2q}} E \left\{ \sigma_{kk'}(X_1) \sigma_{ll'}(X_2) \left[\int K(u) K\left(u + \frac{X_2 - X_1}{h}\right) \sigma_{kl}(X_1 + hu) f(X_1 + hu) du \right] \right. \\ &\quad \left. \left[\int K(u') K\left(u' + \frac{X_2 - X_1}{h}\right) \sigma_{k'l'}(X_1 + hu') f(X_1 + hu') du' \right] \right\} \\ &= \frac{1}{h^q} E \left\{ \sigma_{kk'}(X_1) \sigma_{ll'}(X_1 + hv) \left[\int K(u) K(u + v) \sigma_{kl}(X_1 + hu) f(X_1 + hu) du \right] \right. \\ &\quad \left. \left[\int K(u') K(u' + v) \sigma_{k'l'}(X_1 + hu') f(X_1 + hu') du' \right] f(X_1 + hv) dv \right\} \\ &= \frac{1}{h^q} E \{ \sigma_{kk'}(X_1) \sigma_{ll'}(X_1) \sigma_{kl}(X_1) \sigma_{k'l'}(X_1) f^3(X_1) \} \\ &\quad \left[\int \int \int K(u) K(u + v) K(u') K(u' + v) du du' dv \right] + o(h^{-q}) = O\left(\frac{1}{h^q}\right), \end{aligned}$$

by Assumptions 2.4-2.6. Moreover,

$$\begin{aligned} E[H_n(Z_1, Z_2)^2] &= \sum_{k=1}^m \sum_{l=1}^m \frac{1}{h^{2q}} E \left(K^2\left(\frac{X_1 - X_2}{h}\right) \sigma_{kl}(X_1) \sigma_{kl}(X_2) \right) \\ &= \sum_{k=1}^m \sum_{l=1}^m \frac{1}{h^q} E \left(\sigma_{kl}(X_1) \int K^2(u) \sigma_{kl}(X_1 + hu) f(X_1 + hu) du \right) \\ &= \sum_{k=1}^m \sum_{l=1}^m \frac{1}{h^q} E [\sigma_{kl}^2(X_1) f(X_1)] \int K^2(u) du + o(h^{-q}) = O\left(\frac{1}{h^q}\right), \end{aligned}$$

$$\begin{aligned} E[H_n(Z_1, Z_2)^4] &= \sum_{k=1}^m \sum_{l=1}^m \sum_{k'=1}^m \sum_{l'=1}^m \frac{1}{h^{4q}} E \left[K_{12}^4 \psi_1^{(k)} \psi_2^{(k)} \psi_1^{(k')} \psi_2^{(k')} \psi_1^{(l)} \psi_2^{(l)} \psi_1^{(l')} \psi_2^{(l')} \right] \\ &\leq \sum_{k=1}^m \sum_{l=1}^m \sum_{k'=1}^m \sum_{l'=1}^m \frac{1}{h^{3q}} \prod_{p \in \{k, l, k', l'\}} \{ E[h^{-q} K_{12}^{16} \sigma_p^4(X_1) \sigma_p^4(X_2)] \}^{1/4} \\ &\leq \sum_{k=1}^m \sum_{l=1}^m \sum_{k'=1}^m \sum_{l'=1}^m \frac{O(1)}{h^{3q}} \prod_{p \in \{k, l, k', l'\}} \{ E[\sigma_p^4(X_1) \sigma_p^4(X_1) f(X_1)] \}^{1/4} + o(h^{-3q}) \\ &\leq O\left(\frac{1}{h^{3q}}\right). \end{aligned}$$

Assumption 2.5 ensures that the conditions of Lemma 9 are fulfilled, and Theorem 1 follows.

Proof of Corollary 2 Let us first consider the properties of T_n when H_0 does not hold. Notice that (9) holds with θ^* in place of θ_0 . By a weak law of large numbers, it is straightforward to check that the corresponding S_{1n} , S_{2n} and S_{3n} are all $O_p(1)$, so that $T_{1n}^{(k)} = O_p(n^{-1/2})$ and $T_{2n}^{(k)} = O_p(n^{-1})$, for all $k = 1, \dots, m$, using Assumption 2.2. Similarly, we get

$$\frac{1}{n(n-1)} \sum_{i \neq j} D_n(Z_i, Z_j) \rightarrow_p E[D(Z_1, Z_2)] = \sum_{k=1}^m E \left[E^2 \left[\psi^{(k)}(Z, \theta^*) | X \right] f(X) \right] + o(1),$$

so that T_n converges to a strictly positive limit when H_0 does not hold. By a similar reasoning, it is easily shown that $V_n \rightarrow_p V$, whether H_0 holds or not.

Proof of Theorem 3 Under the null hypothesis, we shall show that $n\hat{h}^{q/2}T_n(\hat{h}) - nh^{q/2}T_n(h) = o_p(1)$, where the dependence of T_n on the bandwidth is made explicit. For this equality to hold, we need to show tightness of the process $n(\nu h)^{q/2}T_n(\nu h)$ for $\nu \in [B_1, B_2]$, with $0 < B_1 < 1 < B_2 < \infty$. It can be seen that the second and third term in (9) are both $O_p(n^{-1})$ uniformly for $\nu \in [B_1, B_2]$. Let $\tilde{T}_n(h)$ be the first term in (9). It is asymptotically normal at a fixed point and converges to the same limit for any ν . Moreover, for $\nu_1, \nu_2 \in [B_1, B_2]$,

$$\begin{aligned} & E \left[n(\nu_1 h)^{q/2} \tilde{T}_n(\nu_1 h) - n(\nu_2 h)^{q/2} \tilde{T}_n(\nu_2 h) \right]^2 \\ &= \frac{4n}{n-1} \sum_{k=1}^m \sum_{l=1}^m E \left\{ \sigma_{kl}(X_1) \sigma_{kl}(X_2) \left[(\nu_1 h)^{-q/2} K \left(\frac{X_1 - X_2}{\nu_1 h} \right) - (\nu_2 h)^{-q/2} K \left(\frac{X_1 - X_2}{\nu_2 h} \right) \right]^2 \right\} \end{aligned}$$

is $O[(\nu_1 - \nu_2)^2]$ by a Taylor expansion of $\nu_2^{-q/2} K[x/\nu_2]$ around ν_1 , using Assumption 2.6 and Hölder's inequality. Hence, $n(\nu h)^{p/2} \tilde{T}_n$ is tight for $\nu \in [B_1, B_2]$, see Billingsley (1968). Under the alternative hypothesis, it is sufficient to show that $T_n(\nu h)$ is tight for $\nu_1, \nu_2 \in [B_1, B_2]$, which is shown similarly. An analogous result for V_n then implies the desired result.

Proof of Theorem 6 Henceforth, $\psi_i^{*(k)} = \psi^{(k)}(Z_i, \theta^*) \zeta_i$ and $\hat{\psi}_i^{*(k)} = \psi^{(k)}(Z_i, \theta_n) \zeta_i$, $\dot{\psi}_i^{*(k)} = \dot{\psi}^{(k)}(Z_i, \theta^*) \zeta_i$ and $E^*[\cdot] \equiv E[\cdot | \mathcal{Z}_n]$. We have a decomposition similar to (9), that is,

$$T_n^* = \binom{n}{2}^{-1} \sum_{i \neq j} D_n^{*(k)}(Z_i, Z_j) + \sum_{k=1}^m 2T_{1n}^{*(k)} + T_{2n}^{*(k)}, \quad (10)$$

where $D_n^{*(k)}$, $T_{1n}^{*(k)}$ and $T_{2n}^{*(k)}$ are defined similarly to $D_n^{(k)}$, $T_{1n}^{(k)}$ and $T_{2n}^{(k)}$ in (9), with $\psi_i^{*(k)}$ and $\hat{\psi}_i^{*(k)}$ in place of $\psi_i^{(k)}$ and $\hat{\psi}_i^{(k)}$. We now show that $T_{1n}^{*(k)}$ and $T_{2n}^{*(k)}$ are both $o_p(n^{-1}h^{-q/2})$ for all $k = 1, \dots, m$. Using a mean value theorem argument,

$$T_{1n}^{*(k)} = (\theta_n - \theta^*)' S_{1n}^* + (\theta_n - \theta^*)' S_{2n}^* (\theta_n - \theta^*),$$

$$\begin{aligned} \text{where} \quad S_{1n}^* &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{*(k)} \dot{\psi}_j^{*(k)} K_{ij}, \\ S_{2n}^* &= \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{*(k)} \ddot{\psi}_j^{*(k)}(\bar{\theta}_n) K_{ij}, \quad \|\bar{\theta}_n - \theta^*\| \leq \|\theta_n - \theta^*\| \end{aligned}$$

and $\ddot{\psi}_j^{*(k)}(\theta) = \ddot{\psi}^{(k)}(Z_j, \theta) \zeta_i$. Since $E^*(\zeta_i \zeta_j) = 0$ for $j \neq i$, S_{1n}^* and S_{2n}^* are degenerate U -statistics. Hence, $E^*[S_{1n}^*] = 0$ and as $E^*(\zeta_i)^2 = 1$ for all i ,

$$E^*[S_{1n}^{*2}] = \left[\frac{1}{n(n-1)h^q} \right]^2 \sum_{i=1}^n \sum_{j \neq i}^n \left(\psi_i^{(k)} \right)^2 \left(\psi_j^{(k)} \right)^2 K_{ij}^2.$$

Using similar arguments as in the proof of Theorem 1,

$$E[E^*[S_{1n}^{*2}]] = \left[\frac{1}{n(n-1)h^q} \right]^2 n(n-1)O(h^q) = O(n^{-2}h^{-q}).$$

Similarly, we can prove that $E[E^*[S_{2n}^{*2}]] = O(1)$. As $(\theta_n - \theta^*) = O_p(n^{-1/2})$, we get that $nh^{q/2}T_{1n}^{*(k)} = o_p(1)$ and $nT_{2n}^{*(k)} = o_p(1)$. These terms are then negligible conditional upon the initial sample.

Let us now determine the asymptotic distribution of the first term in (10). For the sake of simplicity, we treat the case where $m = 1$. We then consider

$$\tilde{T}_n^* = \frac{1}{n(n-1)} \sum_{i \neq j} D_n^*(Z_i, Z_j) = \frac{1}{n(n-1)} \sum_{i \neq j} h_j^{-q} \psi_i^* \psi_j^* K_{ij},$$

where $E^*[D_n^*(Z_i, Z_j) | \xi_i] = 0$, for all i . By Proposition 3.2 in De Jong (1987), $\sigma_n^{-1} \tilde{T}_n^* \rightarrow_{d^*} N(0, 1)$ in probability if G_1, G_2 and G_3 are of lower order in probability than $(\sigma_n^2)^2$, where

$$\begin{aligned} \sigma_n^2 &\equiv E^*[\tilde{T}_n^{*2}] = \left[\frac{1}{n(n-1)} \right]^2 \sum_{i \neq j} \sum_{k=1}^{n-1} \sum_{l=k+1}^n E^*[D_n^*(Z_i, Z_j) D_n^*(Z_k, Z_l)] = \frac{1}{n(n-1)h^q} V_n, \\ G_1 &\equiv \left[\frac{1}{n(n-1)} \right]^4 \sum_{i \neq j} E^*[D_n^{*4}(Z_i, Z_j)] = E^2(V_1^4) \left[\frac{1}{n(n-1)} \right]^2 \sum_{i \neq j} h_j^{-4q} \psi_i^4 \psi_j^4 K_{ij}^4, \\ G_2 &\equiv \binom{n}{2}^{-4} \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n E^*[D_n^{*2}(Z_i, Z_j) D_n^{*2}(Z_i, Z_k)] \\ &= E(V_1^4) \binom{n}{2}^{-4} \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n h_j^{-4q} \psi_i^4 \psi_j^2 \psi_k^2 K_{ij}^2 K_{ik}^2, \\ G_3 &\equiv \binom{n}{2}^{-4} \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^n E^*[D_n^*(Z_i, Z_j) D_n^*(Z_i, Z_k) D_n^*(Z_l, Z_j) D_n^*(Z_l, Z_k)] \\ &= \binom{n}{2}^{-4} \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^n h_j^{-4q} \psi_i^2 \psi_j^2 \psi_k^2 \psi_l^2 K_{ij} K_{ik} K_{lj} K_{lk}. \end{aligned}$$

Now, G_1, G_2 and $(\sigma_n^2)^2$ are positive and it is easily checked, as in the proof of Theorem 1, that

$$E[(\sigma_n^2)^2] = E \left[\binom{n}{2}^{-2} \sum_{i=1}^n \sum_{j \neq i}^n h_j^{-2q} \psi_i^2 \psi_j^2 K_{ij}^2 \right]^2 = n^{-8} O[n^2 h^{-3q} + n^3 h^{-2q} + n^4 h^{-2q}] = O(n^{-4} h^{-2q}),$$

and that similarly $E[G_1] = O(n^{-6} h^{-3q})$, $E[G_2] = O(n^{-5} h^{-2q})$ and $E[G_3] = O(n^{-4} h^{-q})$. The convergence of the distribution function is then uniform by Polyá's theorem.

Proof of Theorem 4 Write, uniformly in x

$$\begin{aligned} n^{1/2} R_n(x) &= \frac{1}{n^{1/2}} \sum_{i=1}^n \psi(Z_i, \theta^*) \mathbf{1}(X_i \leq x) + G(x, \theta^*) \frac{1}{n^{1/2}} \sum_{i=1}^n l(Z_i, \theta^*) + o_p(1) \\ &= R_n^0(x) + R_n^1(x) + o_p(1), \end{aligned}$$

where $G(x, \theta) = E[\psi(Z, \theta) \mathbf{1}(X \leq x)]$. The limit process is identified by the convergence of the finite dimensional distributions. Choose $(x_1, \dots, x_p) \in \mathbb{R}^q$ and normalized vectors $(a_1, \dots, a_p) \in \mathbb{R}^m$. Then apply a Central Limit Theorem to obtain that

$$n^{1/2} \sum_{j=1}^p a'_j [R_n^0(x_j) + R_n^1(x_j)] \rightarrow_d \sum_{j=1}^p a'_j R_\infty(x_j).$$

We now show tightness of the process. Note that the index parameter in R_n^1 is included in a deterministic continuous bounded function. Therefore, R_n^1 is tight. For R_n^0 , tightness will be proved when the marginal distributions of each component of X are uniform in $[0, 1]$. The general case is dealt with applying the usual quantile transformation coordinate by coordinate. When the marginal distributions of X are uniform in the interval $[0, 1]$, R_n^0 takes its values in $\times_{k=1}^m D[0, 1]^q$. Let $R_n^0(x) = (R_{n1}^0(x), \dots, R_{nm}^0(x))'$. Since we endow $\times_{k=1}^m D[0, 1]^q$ with the product topology (which is generated by the metric $d^m(f, g) = \max\{d(f_k, g_k) : k = 1, \dots, m\}$, where f_k, g_k are the k -th coordinate of f, g respectively and d is the metric in the Skorokhod Space $D[0, 1]^q$), tightness follows if each coordinate is tight. The increment of the process R_{nk}^0 around $B = (s, t] = \times_{j=1}^m (s_j, t_j]$ is defined in Bickel and Wichura (1971) as

$$R_{nk}^0(B) = \sum_{e_1=0}^1 \dots \sum_{e_q=0}^1 (-1)^{q-\sum_p e_p} R_{nk}^0(s_1 + e_1(t_1 - s_1), \dots, s_q + e_q(t_q - s_q)).$$

Then, it suffices to check the tightness condition in Bickel and Wichura (1971). That is, we have to show that, for any two neighbor intervals B and $B' = (s', t']$, i.e. they abut and for some $j \in \{1, \dots, q\}$, they have the same j -th face $\times_{k \neq j} (s_k, t_k) = \times_{k \neq j} (s'_k, t'_k)$,

$$E \left(R_{nk}^0(B)^2 R_{nk}^0(B')^2 \right) \leq 3\mu(B) \mu(B'),$$

for all $k = 1, \dots, m$, where μ is an arbitrary measure. Notice that, applying Stute's (1997) Lemma 1,

$$E \left(R_{nk}^0(B)^2 R_{nk}^0(B')^2 \right) \leq n E \left(\alpha_1^2 \beta_1^2 \right) + 3n(n-1) E \left(\alpha_1^2 \right) E \left(\beta_1^2 \right),$$

with $\alpha_i = n^{-1/2} \psi_i^{(k)} \mathbf{1}(X_i \in B)$ and $\beta_i = n^{-1/2} \psi_i^{(k)} \mathbf{1}(X_i \in B')$ with $\psi_i^{(k)} = \psi^{(k)}(Z_i, \theta^*)$. Since $\alpha_i \beta_i = 0$,

$$E \left(R_{nk}^0(B)^2 R_{nk}^0(B')^2 \right) \leq 3E \left[\left(\psi_1^{(k)} \right)^2 \mathbf{1}(X_1 \in B) \right] E \left[\left(\psi_1^{(k)} \right)^2 \mathbf{1}(X_1 \in B') \right].$$

Thus, tightness in $\times_{k=1}^m D[0, 1]^q$ is proved when the marginal distributions of each component of X are uniform in $[0, 1]$. The general case is proved applying the usual quantile transformation coordinate by coordinate.

Proof of Corollary 5 By Lemma in Kieffer (1959, p. 424), $\int_{\mathbb{R}^q} R_n(x)' R_n(x) [dF_n(x) - dF(x)] = o_p(1)$. The result, is an immediate consequence of the continuous mapping theorem.

Proof of Theorem 7

2.8 The function $\ell(\cdot, \cdot)$ is such that $E[\ell(Z, \theta^*)] = 0$ and $E[\ell(Z, \theta^*) \ell(Z, \theta^*)']$. (ii) $\dot{\ell}(\cdot, \cdot) = \partial \ell(\cdot, \cdot) / \partial \theta$ exists almost surely in an open neighborhood $\mathcal{N}(\theta^*)$ of θ^* and $\sup_{\theta \in \mathcal{N}(\theta^*)} \|\dot{\ell}(\cdot, \theta)\| < L(\cdot)$, with $E[L(Z)] < \infty$.

It is immediate to show that $\sup_{x,s} \|\Omega_n(x, s) - \Omega(x, s)\| = o_p(1)$. It is also straightforward to obtain that $\sup_x \|\bar{R}_n^*(x) - \bar{R}_n^{0*}(x)\| = o_p(n^{-1/2})$, with $\bar{R}_n^{0*}(x) = n^{-1} \sum_{i=1}^n r_i(x, \theta^*) \zeta_i$, where $r_i(x, \theta)$ is defined in (8). Fix some $(x_1, \dots, x_p) \in \mathbb{R}^q$ and normalized vectors $(a_1, \dots, a_p) \in \mathbb{R}^m$, and define $\psi_i^* = \psi_i \zeta_i$ with $\psi_i = \psi(Z_i, \theta^*)$. By the Cràmer-Wold device, the convergence of the finite distributions follows from the convergence of

$$n^{1/2} \sum_{j=1}^p a_j' \bar{R}_n^{0*}(x_j) = \frac{1}{n^{1/2}} \sum_{i=1}^n \left[\sum_{j=1}^p a_j' \psi_i (1(X_i \leq x_j) + G(x_j, \theta^*) l(Z_i, \theta^*)) \right] \zeta_i = \frac{1}{n^{1/2}} \sum_{i=1}^n W_i \zeta_i.$$

Asymptotic normality is proved by showing the Lindeberg-Levi's condition. That is, for each $\delta > 0$,

$$L_n(\delta) = \frac{1}{n} \sum_{i=1}^n W_i^2 E \left[\zeta_i^2 1 \left(|W_i \zeta_i| > n^{1/2} \delta \right) \middle| \mathcal{Z}_n \right] = o_p(1).$$

Since $|\zeta_i| \leq c$, for all i and some $c > 0$,

$$L_n(\delta) \leq \frac{c^2}{n} \sum_{i=1}^n W_i^2 1 \left(|W_i| \geq \frac{n^{1/2} \delta}{c} \right) = o_p(1),$$

using the fact that W_i^2 are iid with finite first moment. Then, Theorem 7 will follow from the tightness of $n^{1/2} \bar{R}_n^{0*}$. As in the proof of Theorem 3, assume without loss of generality that each coordinate in X is uniform in the interval $[0, 1]$. We have

$$n^{1/2} \bar{R}_n^{0*}(x) = \frac{1}{n^{1/2}} \sum_{i=1}^n \psi_i 1(X_i \leq x) \zeta_i + G(x, \theta^*) \frac{1}{n^{1/2}} \sum_{i=1}^n l(Z_i, \theta^*) \zeta_i = R_n^{00*}(x) + R_n^{01*}(x).$$

The tightness of R_n^{01*} follows from the continuity of G and applying a Central Limit Theorem to the random sum, which does not depend on x . Consider the increment $R_{nk}^{00*}(B)$ around the interval B , as in Theorem 2. Bickel and Wichura's (1971) tightness condition is satisfied, since for two neighbor intervals B and B' ,

$$E \left[R_{nk}^{00*}(B)^2 R_{nk}^{00*}(B')^2 \right] \leq 3 \mu_n(B) \mu_n(B'),$$

where $\mu_n(B) = n^{-1} \sum_{i=1}^n \psi_i^2 1(X_i \in B)$. Applying, the uniform LLN, we have that,

$$\sup_B |\mu_n(B) - \mu(B)| = o_p(1),$$

where $\mu(B) = E[\psi_1^2 1(X \in B)]$ is a continuous measure.

REFERENCES

- Andrews, D. W. K. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1128.
- Apostol, T. M. (1957): *Mathematical Analysis*. Reading: Addison-Wesley.
- Bickel, P.J. and M.J. Wichura (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," *Annals of Mathematical Statistics*, 42, 1656–1670.
- Bierens, H. (1982): "Consistent Model Specification Test," *Journal of Econometrics*, 20, 105–134.
- Bierens, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.
- Bierens, H. and D. Ginther (2001): "Integrated Conditional Moment Testing of Quantile Regression Models," *Empirical Economics*, 26, 307–324.
- Bierens, H. and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.
- Buckley, M. J. (1991): "Detecting a Smooth Signal: Optimality of Cusum Based Procedures," *Biometrika*, 78, 253–262..
- Burbidge, J.B., L. Magee and A.L. Robb (1988): "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 83, 123–127.
- Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.
- Chen, X. and Y. Fan (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373–401.
- Darolles, S., J-P. Florens and E. Renault (1999): "Nonparametric Instrumental Regression," Université Toulouse 1.
- Davidson, R. and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics*. New-York: Oxford University Press.
- De Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.
- De Jong, R.M. (1996): "The Bierens Test Under Data Dependence," *Journal of Econometrics*, 72, 1–32.
- De Jong, R. M. and H. J. Bierens (1994): "On Limit Behavior of a Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 9, 70–90.
- Delgado, M (1993): "Testing the Equality of Nonparametric Regression Curves," *Statistics and Probability Letters*, 17, 199–204.
- Eubank, R. L. and Spiegelman, C.H. (1990): "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387–392.

- Fan, Y. and Q. Li (1996): “Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms,” *Econometrica*, 64, 865–890.
- Fan Y. and Q. Li (2000): “Consistent Model Specification Tests: Nonparametric Versus Bierens’ Tests,” *Econometric Theory*, 16(6), 1016–1041.
- Fan, Y. and O. Linton (1999): “Some Higher Order Theory For a Consistent Nonparametric Specification Test,” forthcoming in *Journal of Statistics Planning and Inference*.
- Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*. Cambridge: Cambridge University Press.
- Guerre, E. and P. Lavergne (2001): “Rate-Optimal Data-Driven Specification Testing in Regression Models,” Université Paris 6.
- Guerre, E. and P. Lavergne (2002): “Optimal Minimax Rates for Nonparametric Specification Testing in Regression Models,” *Econometric Theory*, 18(5), 1139–1171.
- Hall, P. (1984): “Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators,” *Journal of Multivariate Analysis*, 14, 1–16.
- Hall, P. and J. L. Horowitz (1996): “Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators,” *Econometrica*, 64, 891–916.
- Hansen, B. (1996): “Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis,” *Econometrica*, 64, 413–430.
- Härdle, W. and E. Mammen (1993): “Comparing Nonparametric versus Parametric Regression Fits,” *Annals of Statistics*, 21, 1926–1947.
- Hart, J.D. (1997): *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer Verlag, New-York.
- Hong, Y. (1993): “Consistent Testing for Heteroscedasticity of Unknown Form,” U.C. San Diego.
- Hong, Y. and H. White (1995): “Consistent Specification Testing via Nonparametric Series Regressions,” *Econometrica*, 63, 1133–1160.
- Hong-Zhy, A and C. Bin (1991): “A Kolmogorov-Smirnov Type Statistic with Application to Test for Nonlinearity in Time Series,” *International Statistical Review*, 59, 287–307.
- Horowitz J., and W. Härdle (1994): “Testing a Parametric Model Against a Semiparametric Alternative,” *Econometric Theory*, 10, 821–848.
- Horowitz, J.L. and V.G. Spokoiny (2001): “An Adaptive, Rate-Optimal Test of a Parametric Model Against a Nonparametric Alternative,” *Econometrica*, 69(3), 599–631.
- Hart, J. D. (1997): *Nonparametric Smoothing and Lack-of-Fit Tests*. New-York: Springer-Verlag.

- Kieffer, J. (1959): "K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-von-Mises Tests", *Annals of Mathematical Statistics*, 30, 4420–4447.
- Koul, H. L. and Stute, W. (1999): "Nonparametric Model Checks for Time Series," *Annals of Statistics*, 27, 204–236.
- Kozek, A. S. (1991). "A Nonparametric Test of Fit of a Parametric Model," *Journal of Multivariate Analysis*, 37, 66–75
- Lavergne, P. and Q. H. Vuong (2000): "Nonparametric Significance Testing," *Econometric Theory*, 16, 576–601.
- Li, Q. (1999): "Consistent Model Specification Tests for Time Series Econometric Models," *Journal of Econometrics*, 92, 101–147.
- Li Q. and S. Wang (1998): "A Simple Consistent Bootstrap Test for a Parametric Regression Function," *Journal of Econometrics*, 87, 145–165.
- MacKinnon, J.G. and L. Magee (1990): "Transforming the Dependent Variable in Regression Models," *International Economic Review*, 31, 315–339.
- Newey, W. K. (1985a): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.
- Newey, W. K. (1985b): "Generalized Methods of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.
- Newey, W.K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.
- Prakasa Rao, B.L.S. (1983): *Nonparametric Functional Estimation*. Orlando: Academic Press.
- Powell, J. L., J. H. Stock and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.
- Robinson, P.M. (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies*, 56, 511–534.
- Stinchcombe, M.B. and H. White (1998): "Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295–325.
- Stute, W. (1997): "Nonparametric Model Checks for Regression," *Annals of Statistics*, 25, 613–641.
- Stute, W., W. Gonzalez-Manteiga and M. Presedo (1998): "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association*, 93, 141–149.
- Stute, W., Thies, S. and Zhu, L.-X. (1998): "Model Checks for Regression: an Innovation Process Approach," *Annals of Statistics*, 26, 1916–1934.
- Su, J. Q. and L. J. Wei (1991): "A Lack of Fit Test for the Mean Function in a Generalized Linear Model," *Journal of the American Statistical Association*, 86, 420–426.

- Tauchen, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.
- Whang, Y.J.(2000): "Consistent Bootstrap Tests of Parametric Regression Functions," *Journal of Econometrics*, 98, 27–46.
- Whang, Y.J.(2001): "Consistent Specification Testing for Conditional Moment Restrictions," *Economics Letters*, 71, 299–306.
- White, H. (1994): *Estimation, Inference and Specification Analysis*. New-York: Cambridge University Press.
- Wooldridge, J.M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory*, 6, 17–43.
- Zheng, X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.
- Zheng, X. (1998): "A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123–138.