

Consistent Tests of Conditional Moment Restrictions

Miguel A. DELGADO, Manuel A. DOMÍNGUEZ,
Pascal LAVERGNE

ABSTRACT. – We propose two classes of consistent tests in parametric econometric models defined through multiple conditional moment restrictions. The first type of tests relies on nonparametric estimation, while the second relies on a functional of a marked empirical process. For both tests, a simulation procedure for obtaining critical values is shown to be asymptotically valid. Finite sample performances of the tests are investigated by means of several Monte-Carlo experiments.

Tests convergents de restrictions de moments conditionnels

RÉSUMÉ. – Nous proposons deux classes de tests convergents de la spécification paramétrique de modèles économétriques définis par des restrictions de moments conditionnels. La première est basée sur l'estimation non-paramétrique, la seconde sur une fonctionnelle d'un processus empirique marqué. Pour les deux types de tests, une procédure de simulation permet d'obtenir des valeurs critiques asymptotiquement valides. Le comportement en petits échantillons de ces tests est étudié par des simulations.

We thank the editor and two referees for their comments. We acknowledge financial support from: the Spanish Ministry of Technology (Dirección General de Enseñanza Superior) under grant R000238212, Asociación Mexicana de Cultura, Consejo Nacional de Ciencia y Tecnología (CONACYT) under project grant J38276D, and the European Commission through research training grant ERBFMBICT961595.

Miguel A. DELGADO : Universidad Carlos III de Madrid, Spain

Manuel A. DOMÍNGUEZ : ITAM, Mexico

Pascal LAVERGNE: Simon Fraser University, 8888 University Drive, Burnaby, B.C., Canada, V5A 1S6; e-mail: pascall@sfu.ca

1 Introduction

Econometric models are frequently defined through conditional moment restrictions. This is the case for models simultaneously parameterizing different conditional moments (e.g. conditional mean and conditional variance) without specific distributional assumptions, transformation models, models identified by means of instrumental variables, and nonlinear-in-variables simultaneous equation models. Estimation under conditional moment restrictions is considered by CHAMBERLAIN [1987], NEWEY [1990], ROBINSON [1991], and DOMINGUEZ and LOBATO [2004] among others. Checking the validity of these moment restrictions is a central issue. Among the most popular specification tests, M-tests, as proposed by NEWEY [1985a, 1985b], TAUCHEN [1985] and WOOLDRIDGE [1990], aim at testing a finite number of arbitrary unconditional moment restrictions implied by the conditional moment restrictions. These tests are “directional” in the sense that they are unable to detect some misspecifications, though they may be optimal in the direction of precisely specified alternatives. Omnibus specification tests, which are consistent against any misspecification, are useful when the econometrician has no specific alternative in mind.

Two approaches have been developed in the recent literature to derive omnibus specification tests of a parametric regression function. The first approach compares the fitted parametric regression function with a nonparametric function estimated using smoothers, see EUBANK and SPIEGELMAN [1990], KOZEK [1991], HÄRDLE and MAMMEN [1993], HONG and WHITE [1995], FAN and LI [1996], ZHENG [1996], and ELLISON and ELLISON [2000] 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, there is a one-to-one relationship between the density and integral transforms such as the probability distribution function or 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 to build a specification test, see BIERENS [1982, 1990], DE JONG and BIERENS [1994] and BIERENS and PLOBERGER [1997]. The smooth test approach leads to asymptotically pivotal tests statistics but that depends on a smoothing parameter, while the integral-transform approach yields test statistics that have case-dependent limiting distributions but do not depend on the choice of a smoothing parameter. The two approaches can be interpreted as M-tests with an infinite number of moment restrictions.

Most work has focused on distribution and regression models, with few exceptions. ZHENG [1998] and BIERENS and GINTHER [2001] deal with quantile regression models. STINCHCOMBE and WHITE [1998], KOUL and STUTE [1999], and WHANG [2001] propose tests for an univariate conditional moment restriction. CHEN and FAN [1999] and DELGADO and GONZÁLEZ-MANTEIGA [2001] provide consistent procedures for testing some conditional moment restrictions in semiparametric and

nonparametric models, e.g. testing for omitted variables, but do not allow the null hypothesis to depend on unknown parameters.

Our aim is to propose tests for multiple conditional moment restrictions with unknown parameters and thus to provide useful procedures for econometric modeling. The innovative feature of our study with respect to previous work is to simultaneously follow and generalize the two approaches developed for specification testing of regression models. We restrict 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, see also DOMINGUEZ and LOBATO [2003], and LI [1999], who generalizes FAN and LI'S [1996] results. Here, we focus 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 explain 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 experiments, as there is little evidence in the econometric literature on the comparative small sample performances of the competing approaches.

The paper is organized as follows. In Section 2, we detail our general testing framework and we discuss examples of applications. In Section 3, we explain how the two testing approaches can be generalized and we study the asymptotics of some tests statistics based upon either nonparametric kernel estimation or integral-transform regression estimation. In Section 4, we explain the difficulties in applying known bootstrap methods in our general framework and we propose instead a simulation procedure to approximate critical values of each type of test. Section 5 reports the results of our Monte-Carlo study. Section 6 gives some directions for further research. Technical proofs are confined to Section 7.

2 Testing Framework

Let $\mathcal{Z}_n = \{Z_i, i = 1, \dots, n\}$ be a random sample drawn from a random vector $Z \in \mathbb{R}^s$, and let $X \in \mathbb{R}^q$ be a subvector of Z . Following NEWEY [1985a, 1985b], CHAMBERLAIN [1987], WOOLDRIDGE [1990] and WHITE [1994], we consider a particular parametric model indexed by $\theta \in \Theta \subset \mathbb{R}^p$, defined through conditional moment restrictions of the form

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

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 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*

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

where $\mu(\cdot, \cdot)$ and $\omega^2(\cdot, \cdot)$ are known functions up to the value of θ_0 , see e.g. WOOLDRIDGE [1990]. The parametric model is completely defined through restrictions (I) 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 the 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

$$E[(Y - \mu(X, \theta_0))^2 - \sigma_0^2 | X] = 0 \text{ a.s.}$$

where $\sigma_0^2 = E[U^2]$ is included in θ_0 .

EXAMPLE 2: *Consider the model*

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

where $\mu(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ are known functions. When λ_0 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 e.g. BURBIDGE, MAGEE and ROBB [1988]. When the distribution of U is unknown, the parametric model is simply defined through

$$E[U | X] = 0 \text{ a.s.},$$

that is through restrictions of the type (I). 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 then invalidates 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

$$\begin{aligned} Q &= a_0P + b_0I + U, & a_0 < 0 & \text{ (Demand)} \\ Q &= \alpha_0P + \beta_0W + V, & \alpha_0 > 0 & \text{ (Supply)}, \end{aligned}$$

where Q and P respectively denote quantity and price, I and W are exogenous variables, and U and V are the error terms. 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.

$$E \begin{pmatrix} Q - a_0P - b_0I \\ Q - \alpha_0P - \beta_0W \end{pmatrix} | X = \mathbf{0}.$$

These restrictions ensure identification of the coefficients and constitute the basis of 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]. Under the assumption that X is exogenous, rejecting these conditional moment restrictions means that the postulated functional forms of the demand and supply curves are incorrect.

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. STINCHCOMBE and WHITE [1998] proposed detecting misspecification of a conditional probability model by testing that the conditional score function is identically zero or that the information matrix equality holds conditionally upon the independent variables. Alternatively, for a correctly specified model, the latter restriction can be used to check whether a root of the likelihood equations is a *global* maximizer of the likelihood function, see GAN and JIANG [1999].

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 Z_n is an iid sample from a random variable Z with support on R^s . The subvector X_n has a distribution $F(\cdot)$ which admits an absolutely continuous density function $f(\cdot)$ with respect to the Lebesgue measure.
- 2.2 There is a \sqrt{n} -consistent estimator θ_n^* of the pseudo-parameter θ^* , where θ^* is a unique interior point of Θ . Under H_0 , $\theta^* = \theta_0$.

- 2.3** (i) $E \left\| \psi(Z, \theta^*) \right\|^2 < \infty$. 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^* , $E \left\| \dot{\psi}^{(k)}(Z, \theta^*) \right\| < \infty$ and $\sup_{\theta \in \mathcal{N}(\theta^*)} \left\| \ddot{\psi}^{(k)}(\cdot, \theta) \right\| < S(\cdot)$, with $E[S(Z)] < \infty$, where $\|\cdot\|$ denotes either a vector or matrix 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 almost everywhere.

Assumption 2.1 restricts our analysis to an iid context where the conditioning variables are continuous. Allowing some discrete components in X is not difficult, but would involve a more cumbersome notation¹. Assumption 2.2 says that we have at hand an estimator θ_n that is \sqrt{n} -consistent for some uniquely defined θ^* which is θ_0 under H_0 . Since we are in a testing problem, the uniqueness of θ^* could be dropped at the price of introducing other technicalities, see e.g. GUERRE and LAVERGNE [2005]. Our assumption 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]. Consider for instance an IV estimator θ_n solution to

$$(2) \quad \frac{1}{n} \sum_{i=1}^n M(X_i) \psi(Z_i, \theta) = \mathbf{0},$$

where $M(X)$ is a suitable vector of instruments. Then θ^* is defined as the unique solution to

$$E \left[M(X) \psi(Z, \theta^*) \right] = \mathbf{0}.$$

Assumption 2.3 imposes some regularity on the functions entering the restrictions, see e.g. NEWBY [1985b] or ROBINSON [1991] for similar conditions.

3 The Two Approaches: Rationale and Asymptotics

3.1 Smooth Tests

The first step consists in formulating the conditional moment restrictions as a completely equivalent (not arbitrary) unconditional moment restriction. Specifically, H_0 is equivalent to

$$(3) \quad H_0 : E \left[\psi'(Z, \theta_0) E(\psi(Z, \theta_0) | X) w_S(X) \right] = 0 \quad \text{for some } \theta_0 \in \Theta,$$

1. See ZHENG [1996] or STUTE [1997] for details.

for a suitable weight function $w_s(\cdot)$ which is strictly positive almost surely onto the support of $f(\cdot)$. The null hypothesis thus 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 but we here focus on the main arguments and delay our discussion on weighting until Section 3.3. A possible and convenient choice for $w_s(\cdot)$ is the density $f(\cdot)$ itself, see e.g. POWELL, STOCK and STOKER [1989], FAN and LI [1996], ZHENG [1996] and LAVERGNE and VUONG [2000] for an analogous device, and we thereafter concentrate on this choice. Define

$$T(\theta) = E \left[\psi'(Z, \theta) E(\psi(Z, \theta) | X) f(X) \right].$$

A test can then be built by checking whether an estimator of the latter quantity is significantly different from 0. The sample analog of $T(\theta_0)$ is

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

A feasible estimator thus requires on the one hand a consistent estimator for θ_0 and on the other hand a consistent estimator for $E(\psi(Z, \theta) | X) f(X)$. Substituting $\hat{\theta}_n$ for θ_0 and the kernel estimator of $E(\psi(Z, \theta) | X = \cdot) f(\cdot)$ into the last equation yields the estimator of $T(\theta_0)$ defined as

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

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 $\psi(Z_i, \hat{\theta}_n)$. It resembles the statistic proposed in FAN and LI [1996] and ZHENG [1996] for testing the specification of regression functions and is constructed similarly, with the generalized residual vector $\psi(Z_i, \hat{\theta}_n)$ in place of standard regression residuals.

The statistic T_n is a U -statistic. Its asymptotic behavior is determined through its Hoeffding decomposition and depends on whether H_0 holds, see SERFLING [1980]. Under misspecification, the first term of its Hoeffding decomposition determines the behavior of T_n , which is asymptotically \sqrt{n} -consistent for the unconditional moment in (3). Under H_0 however, the first term in the Hoeffding decomposition vanishes and higher order terms determine its asymptotic behavior. To state our formal result, we introduce some standard assumptions on the kernel function and the bandwidth parameter. We also impose some smoothness restrictions using the following definitions.

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

Let $\alpha_k(X) = E[\psi^{(k)}(Z, \theta^*) | X]$, $\gamma_k(X) = E[\psi^{(k)}(Z, \theta^*) | X]$, $\sigma_k^4(X) = E[(\psi^{(k)}(Z, \theta^*))^4 | X]$, $\sigma_{kl}(X) = E[\psi^{(k)}(Z, \theta^*) \psi^{(l)}(Z, \theta^*) | X]$ for $k, l = 1, \dots, m$, and let $\lambda(X) = E[S(Z) | X]$.

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

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

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 $E\left[\left\|\dot{\psi}^{(k)}(Z, \theta^*)\right\|^{8/3}\right] < \infty$.

THEOREM 1: Under H_0 and Assumptions 2.1 to 2.6,

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

To build an estimator for V , we note that

$$E[\sigma_{kl}^2 f(X)] = E\left\{\psi^{(k)}(Z, \theta_0) \psi^{(l)}(Z, \theta_0) E[\psi^{(k)}(Z, \theta_0) \psi^{(l)}(Z, \theta_0) | X] f(X)\right\}.$$

Hence, using a reasoning similar to the one leading to T_n , we can propose an estimator for V as

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

One could alternatively use $K_{ij} \int K^2(u) du$ rather than K_{ij}^2 in V_n . The key reason for our choice is that the selected form of the estimator gives better small sample properties for the test in preliminary simulations. Thus our 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 therefore a *one-sided* normal test and is asymptotically able to detect any violation of H_0 .

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 can be written as

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

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 simply relies on the addition of the latter two statistics. The asymptotic variance estimator is computed as

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

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 2 extends to the case of a random \hat{h} if there exists a deterministic h that fulfills the assumptions of Corollary 2 and such that $(\hat{h} - h)/h = o_p(1)$.

3.2 Integral-Transform Tests

The null hypothesis of interest H_0 is equivalent to the statement that $E[\psi(Z, \theta_0) w(X)] = 0$ for all real-valued $w(\cdot)$ such that this expectation exists. Checking such a statement is clearly difficult. Fortunately, as shown by STINCHCOMBE

and WHITE [1998], one can restrict to some particular class of functions. Therefore a possible alternative equivalent way to write H_0 is

$$(4) \quad H_0 : E \left[\psi(Z, \theta_0) w_p(X, x) \right] = \mathbf{0} \quad \forall x \in \mathbb{R}^q \text{ for some } \theta_0 \in \Theta,$$

where $w_p(\cdot, x)$ is a well-chosen function depending on a nuisance parameter x . STINCHCOMBE and WHITE [1998] show that there exists different choices that yield consistent tests, for instance $w_p(X, x) = G(X'x)$ with $G(\cdot)$ the logistic cumulative distribution function or the exponential 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. Hence we consider the integral-transform function $R(x, \theta) = E[\psi(Z, \theta) \mathbf{1}(X \leq x)]$, which is analogous to the distribution function $F(x) = E[\mathbf{1}(X \leq x)]$. Previous work where a similar choice is made includes STUTE [1997], ANDREWS [1997], KOUL and STUTE [1999], and WHANG [2000]. This is convenient because the expectations in (4) are easily estimated by sample analogs of the form

$$\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta_0) \mathbf{1}(X_i \leq x).$$

To get a feasible estimator, we replace θ_0 by θ_n and come up with

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

which is an empirical process of dimension m marked by the generalized residuals $\psi(Z_i, \theta_n)$. A test statistic for (4) 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 and we consider the Cramer-von-Mises type statistic²

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

where $F_n(\cdot)$ is the empirical distribution function of X . A Kolmogorov-Smirnov type test can also be constructed, but we focus on the Cramer-von-Mises type test

2. This terminology is used by STUTE [1997].

as WHANG [2000] reports better performances for this type of test in a regression context. Hence, our implicit null hypothesis is

$$H_0 : \int E[\psi'(Z, \theta_0) \mathbf{1}(X \leq x)] E[\psi(Z, \theta_0) \mathbf{1}(X \leq x)] dF(x) = 0$$

for some $\theta_0 \in \Theta$,

an unconditional moment restriction equivalent to the original conditional moment restrictions. We hereafter provide a functional central limit theorem for $R_n(\cdot)$ under the following assumption.

2.7 There is an estimator θ_n that admits the asymptotic expansion $\theta_n = \theta^* + n^{-1} \sum_{i=1}^n \ell(Z_i, \theta^*) + o_p(n^{-1/2})$ for some unique interior point θ^* of Θ , where the function $\ell(\cdot, \cdot)$ is such that $E[\ell(Z, \theta^*)] = 0$ and $E[\ell'(Z, \theta^*)]$ exist. Under H_0 , $\theta_0 = \theta^*$.

For instance, if θ_n is solution to (2), $\ell(Z, \theta^*) = -\left(E\left[M(X) \dot{\psi}(Z, \theta^*)\right]\right)^{-1} M(X) \psi(Z, \theta^*)$, where $\dot{\psi} = (\dot{\psi}^{(1)'}, \dots, \dot{\psi}^{(m)'})'$. We need Assumption 2.7 as the substitution of θ_0 by θ_n has a first-order effect on the behavior of $R_n(x)$. Indeed under H_0 , $\sqrt{n} R_n(x)$ is, as a process indexed by x , asymptotically equivalent to

$$(5) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i(x, \theta_0), \quad \text{where } r_i(x, \theta) = \psi(Z_i, \theta) \mathbf{1}(X_i \leq x) + E[\dot{\psi}(Z, \theta) \mathbf{1}(X \leq x)] \ell(Z_i, \theta).$$

It is apparent that the second term of $r_i(x, \theta)$ comes from the estimation error $\theta_n - \theta^*$.

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 Skorohod 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,*

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

$$n^{-1}c_n \rightarrow_p \int_{\mathbb{R}^q} E\left[\psi(Z, \theta^*) \mathbf{1}(X \leq x)\right] E\left[\psi(Z, \theta^*) \mathbf{1}(X \leq x)\right] dF(x) > 0$$

otherwise.

An asymptotic test can be based on the statistic c_n as soon as one can compute critical values. However the asymptotic distribution of c_n under H_0 is known only in special cases, see e.g. DELGADO [1993]. Hence, we propose in Section 4 a simulation method to approximate critical values. The resulting test is thus one-sided and consistent against any 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*

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

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 the generalized residuals $\psi(\cdot, \cdot)$ by $A(X)\psi(\cdot, \cdot)$, where $A(x)$ is a known $m \times m$ nonsingular matrix for almost all x . For the smooth test, when $A(X) = A_*(X) \{^{1/2}(X)$, the null hypothesis (3) writes

$$H_{0A} : E \left[\psi'(Z, \theta_0) W(X) E(\psi(Z, \theta_0) | X) \right] = 0$$

where $W(x) = A'(x) A(x) = A_*'(x) A_*(x) f(x)$, which is positive definite for almost all x . The asymptotic variance of the corresponding smooth test statistic would be

$$V_A = 2 \text{tr} \left\{ E \left[\Sigma_A^2(X) \right] \right\} \int_{\mathbb{R}^q} K^2(u) du$$

where $\Sigma_A(X) = A(X) \Sigma(X) A'(X)$ and $\Sigma(X)$ is the $m \times m$ matrix with generic element $\sigma_{kl}(X)$. Our weighting choice of Section 2.1 corresponds to an $A_*(X)$ equal to the identity matrix. For the integral-transform test, the null hypothesis writes

$$H_{0A} : E \left[\psi'(Z, \theta_0) A'(X) \mathbf{1}(X \leq x) \right] E \left[A(X) \psi(Z, \theta_0) \mathbf{1}(X \leq x) \right] dF(x) = 0$$

for some $\theta_0 \in \Theta$,

and the covariance structure of the corresponding limiting process $R_{A\infty}$ would be

$$\Omega_A(x, s) = E \left[A(X) r_1(x, \theta_0) r_1'(s, \theta_0) A'(X) \right], \quad \forall x, s \in \mathbb{R}^q.$$

Such expressions might be a starting point for investigating the optimality of the weighting matrix $A(X)$, but this is outside the scope of this paper³.

Second, as easily seen from our Example 1, one could consider testing (3) 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 asymptotically pivotal distribution of the sum of the standardized individual statistics. Yet another possibility 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. Indeed, it can reject the null hypothesis for negative values of T_{1n} and T_{2n} , which asymptotically arise under H_0 only, so that the level of the test is more difficult to control.

4 Simulated Critical Values

For both the smooth test and the integral-transform test, obtaining accurate small sample critical values is crucial. On the one hand, the practical implementation of the asymptotic smooth test 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, the integral-transform statistic has a limiting

3. The results in this section can be easily derived using our Theorems 1 and 4.

distribution under the null hypothesis that depends on the unknown data-generating process and therefore cannot be tabulated in general. One way of solving the latter problem is to derive case-independent upper-bounds of the asymptotic critical values as proposed by BIERENS and PLOBERGER [1997], but these are not sharp. Another possibility consists of applying a martingale transformation to $R_n(\cdot)$ that yields an asymptotically pivotal distribution under H_0 , as studied by KHMALADZE [1993], STUTE, THIES and ZHU [1998], KOUL and STUTE [1999] among others. However, the martingale transformation relies on some smoothing and trimming parameters, thus weakening the main advantage of the integral-transform test compared to the smooth test.

In what follows, we first briefly review bootstrap methods that has been 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 Bootstrap Methods for Approximating Critical Values

For smooth specification tests of regression models, wild bootstrap procedures have been proposed to compute accurate small sample critical values. When $\psi(Z, \theta) = Y - \mu(X, \theta)$ with Y scalar, HÄRDLE and MAMMEN [1993] propose to generate bootstrap resamples as $\left\{ (Y_i^*, X_i), i = 1, \dots, n \right\}$, where $Y_i^* = \mu(X_i, \theta_n) + \psi_{ni}^*$, with $\psi_{ni}^* = \psi(Z_i, \theta_n) \zeta_i$ and

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

The zero mean condition ensures that $Y_i^* - \mu(X_i, \theta_n)$ has zero mean conditional on the original sample. This condition yields resampled observations that fulfill the null hypothesis. The unit variance condition ensures that $Y_i^* - \mu(X_i, \theta_n)$ has the same conditional variance as $Y_i - \mu(X_i, \theta_n)$. Both conditions are crucial to obtain asymptotically valid critical values. These are then obtained in three steps: (i) generate several resamples (ii) for each resample, compute the corresponding estimate θ_n^* and test statistic t_n^* (iii) compute the empirical quantile of order $(1 - \alpha)$ of the statistics t_n^* . This provides a critical value to be compared to the original test statistic t_n . For given α the estimation accuracy of the critical value increases with the number of resamples. Such a wild bootstrap procedure is also applicable to statistics based on integral transforms of a regression, see STUTE, GONZÁLEZ-MANTEIGA and PRESEDO [1998] and WHANG [2000].

To apply such a wild bootstrap procedure in our framework, we should first figure out how to generate resamples with the same observations for the exogenous variables, but new observations for the endogenous variables. The above presentation of the bootstrap suggests to generate an artificial sample of generalized residuals as $\psi_{ni}^* = \psi(Y_i, X_i, \theta_n) \zeta_i$ and to find the values Y_i^* such that $\psi_{ni}^* = \psi(Y_i^*, X_i, \theta_n)$. From such resamples, one could then proceed as in Steps (ii) and (iii) above.

However, in some cases such a resampling method may be infeasible or difficult to implement from a practical point of view. First, practical difficulties arise when the model is nonlinear in the endogenous variable as in the transformation model of Example 2, so that resampling and subsequent computations can be time-consuming. Moreover, a reduced form for the endogenous variables may not be available, as in a nonlinear simultaneous equation model. Second, when one considers testing the specification of a single equation from a simultaneous equation system, say the demand equation from Example 3, we confront the issue of how to generate price observations: if we relied on the parametric supply equation, then we would actually test the whole specification of the system. LAVERGNE and THOMAS [2005] noted the same type of difficulty when testing the null hypothesis $E[Y|X] = 0$ in the partially linear model

$$Y = \alpha + X\beta + g(Z) + \varepsilon,$$

where it is unclear how to generate observations for the variable Z . Third, a theoretical problem arises as the equation $\psi_{ni}^* = \psi(y, X_i, \theta_n)$ may not have a unique solution in y almost surely. It is easily seen that such a solution does not exist in Example 1, where there are two generalized residuals for only one endogenous variable. In other instances, the solution may not be unique.

It may well be that a suitable resampling method can be tailored in some particular cases. Such an issue has been dealt with in some specific contexts, as in parametric tests based on Generalized Method of Moments estimators, see HALL and HOROWITZ [1996]. In each situation, it would remain to determine the conditions analogous to B1 necessary to obtain asymptotically valid critical values. Since we aim at a general method for determining critical values, we use a simulation technique, the conditional Monte-Carlo approach, proposed by SU and WEI [1991] and HANSEN [1996] and used by DE JONG [1996] and CHEN and FAN [1999] for integral-transform type statistics. This technique is simple and generally applicable but is not likely to be as good as one tailored for a specific model.

4.2 Smooth Tests

Because the statistic T_n is a function of $\mathcal{Y}_n = \{\psi(Z_i, \theta_n), X_i, i = 1, \dots, n\}$, 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 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

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

Alternatively, t_n^* could be computed using as a variance estimator

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

However, we keep the original variance estimator V_n for simplicity, since the asymptotic analysis is similar and unreported simulations results reveal that there seems to be no substantial change in the finite sample behavior of the test.

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 shows that such a simulated critical value is asymptotically valid.

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

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

where $\Phi(\cdot)$ is the standard normal distribution.

In the next section, we provide some evidence on the properties of the simulation technique in different setups. Our theoretical result does not ensure that the simulation method yields more accurate critical values than the asymptotic approximation. Such a study is outside the scope of our paper. Finer theoretical properties of simulation or bootstrap methods have been established only for testing a linear regression model. LI and WANG [1998] showed that moments up to order four are accurately matched by the wild bootstrap under the supplementary condition $E(\zeta_1^3) = 1$. FAN and LINTON [1999] further provide Edgeworth's expansions for the distribution of a similar test statistic in a regression model with symmetric errors.

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, $\sqrt{n}R_n$ has a first-order asymptotic expansion that depends on $r_i(x, \theta_0)$ only. This suggests to consider the statistic

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

$$r_{ni}^*(x) = \left(\psi(Z_i, \theta_n) \mathbf{1}(X_i \leq x) + \left[\frac{1}{n} \sum_{j=1}^n \psi(Z_j, \theta_n) \mathbf{1}(X_j \leq x) \right] \ell(Z_i, \theta_n) \right) \zeta_i.$$

It is easy to see that conditional on \mathcal{Z}_n , $n^{1/2}\bar{R}_n^*(\cdot)$ is centered at zero and with covariance structure $\Omega_n(x, s) = n^{-1} \sum_{i=1}^n r_i(x, \theta_n) r_i'(s, \theta_n)$. Critical values for testing H_0 at level α are obtained as for the smooth test. 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 solution to (2) then $\ell(Z_i, \theta_n)$ can be replaced by $-L_n^{-1} M(X_i) \psi(Z_i, \theta_n)$, where $L_n = n^{-1} \sum_{i=1}^n M(X_i) \dot{\psi}(Z_i, \theta_n)$. When $\psi(Z, \theta) = Y - X\theta$ and θ_n is the OLS estimator, $\ell(Z_i, \theta_n) = -\left[n^{-1} \sum_{j=1}^n X_j X_j' \right]^{-1} X_i \psi(Z_i, \theta_n)$ and

$$R_n^*(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\psi(Z_i, \theta_n) \mathbf{1}(X_i \leq x) - \left[\frac{1}{n} \sum_{j=1}^n X_j' \mathbf{1}(X_j \leq x) \right] \left[\frac{1}{n} \sum_{j=1}^n X_j X_j' \right]^{-1} X_i \psi(Z_i, \theta_n) \right) \zeta_i.$$

This is the wild bootstrap statistic proposed by STUTE, GONZÁLEZ-MANTEIGA and PRESEDO [1998].

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 $E[L(Z)] < \infty$.

THEOREM 7: Under the assumptions of Corollary 5 and Assumption 2.8, if $|\zeta_1| \leq c$ with probability 1 for some $c < \infty$

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

where c_∞ is as in Corollary 5 with θ^* in place of θ_0 .

5 Monte-Carlo Results

Even in the simplest case of testing a regression function, there is little evidence in the econometric literature on the comparative small sample performances of the different approaches. A notable exception is WHANG [2000] who compares KOLMOGOROV-SMIRNOV and CRAMER-VON-MISES type tests to the tests of HÄRDLE and

MAMMEN [1993] and BIERENS and PLOBERGER [1997]. In this section, we investigate the small sample behavior of the tests considered in this paper within the setup of our Examples 1 to 3.

Testing for linearity of a regression function. We first consider testing 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

$$H_0 : E[Y - \alpha_0 - \beta_0 X | X] = 0 \text{ a.s. for some } (\alpha_0, \beta_0).$$

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 and X_i are independent identically distributed $N(0, 1)$. 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 tests' behavior.

In each case two 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, where the ζ_i have a two-point distribution as in STUTE and alii [1998]. 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 kernel is the standard Gaussian density function. The results of each experiment are reported on a graph that shows 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 simulated critical values, and the dash line is the rejection probability for the smooth test using asymptotic critical values.

Figures 1 and 2 report results for DGP_0 and 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]. We also note that large bandwidths lead to large size distortions, a conclusion also reached by FAN and LINTON [2003]. However, the empirical size of the test becomes closer to its nominal size when the sample size increases. Using simulated critical values improves 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 $n = 50$.

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 very close to one when $n = 100$. When the frequency of the alternative increases, the rejection probability for Stute's test decreases. For DGP_2 it is less than 20% when $n = 50$ but increases to nearly 40% when $n = 100$. A sample size of 500 is required to get a power over 95%. As expected, power of the smooth test varies with the bandwidth and the frequency of

FIGURE 1
DGP_φ, n = 50

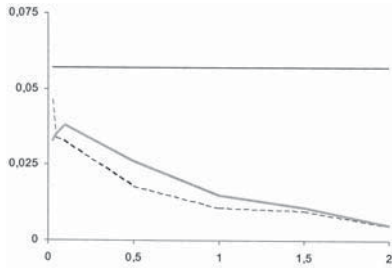


FIGURE 2
DGP_φ, n = 100

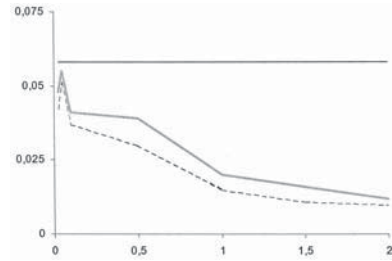


FIGURE 3
DGP_p, n = 50

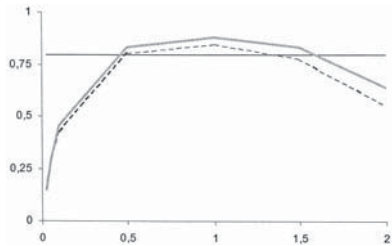


FIGURE 4
DGP_p, n = 100

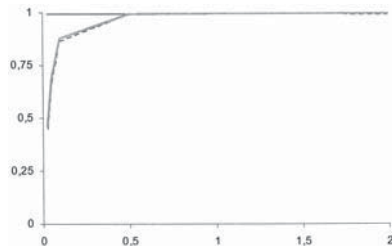


FIGURE 5
DGP_z, n = 50

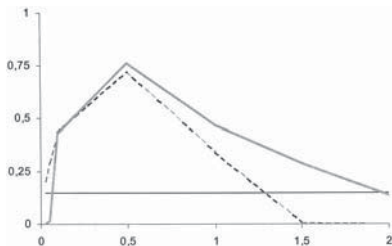


FIGURE 6
DGP_z, n = 100

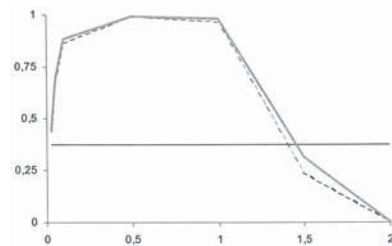


FIGURE 7
DGP_z, n = 50

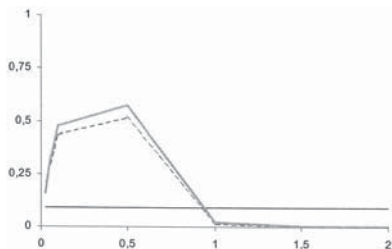
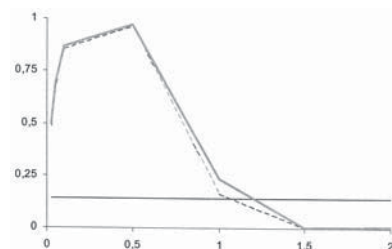


FIGURE 8
DGP_z, n = 100



the alternative. These phenomena have already been noted in other contexts, see e.g. LAVERGNE and VUONG [2000]. Small bandwidths correspond to low rejection frequencies, as also found by FAN and LINTON [2003], while large ones may yield low or high power depending upon the alternatives. The smooth test has reasonably high power for a range of bandwidths that narrows as δ increases. As was the case under the null hypothesis, using bootstrap critical values leads to some but limited improvement in most cases.

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

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

We consider the design

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

where the U_i and X_i are generated as before. The null hypothesis DGP'_0 corresponds to $\delta = 0$ and 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 experiment. 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 level 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 except for a very small bandwidth, but power of the latter is larger than 70% when $n = 250$ and larger than 95% for $n = 350$.

Testing the transformation model. We now consider a model with a well-known nonlinear transformation in the endogenous variable, namely the arcsinh transformation. The hypothesis of interest here can be written as

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

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 and X_i are generated as before, but the variance of the error term is 0.5. The parameters are estimated by one-step *GMM* with vector of instruments (1,

FIGURE 9

$DGP'_\rho, n = 100$

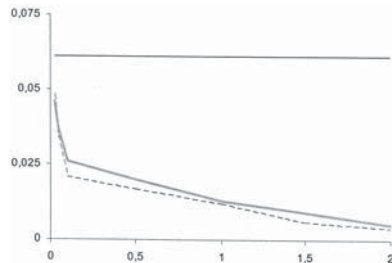


FIGURE 10

$DGP'_\rho, n = 250$

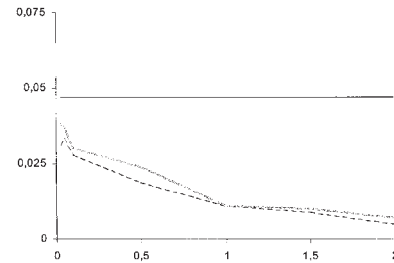


FIGURE 11

$DGP'_p, n = 100$

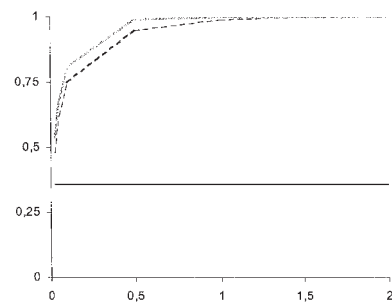
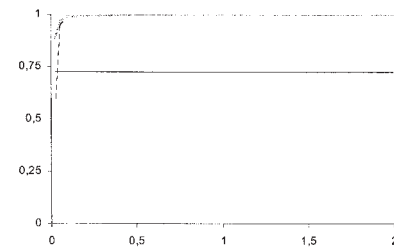


FIGURE 12

$DGP'_p, n = 250$



X, X^2). 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 previous experiments. 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 power function of the smooth test decreases sharply as the bandwidth increases for $n = 100$ but when $n = 250$, the power is essentially one over the whole range of considered bandwidths. Concerning the test based on c_n , it has low power for $n = 100$ but this power increases up to 95% when $n = 250$. For the higher frequency alternative DGP''_3 (Figures 19 and 20), the power of the smooth test is more dependent of the value of the smoothing parameter for both sample sizes, going from 1 to 0 for small and large bandwidths respectively. The test based on c_n has essentially no power for $n = 100$, but its rejection probability reaches 70% when $n = 250$.

FIGURE 13

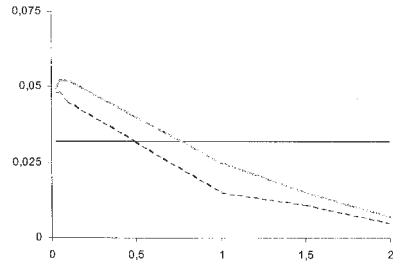
 $DGP''_{\phi}, n = 100$ 

FIGURE 14

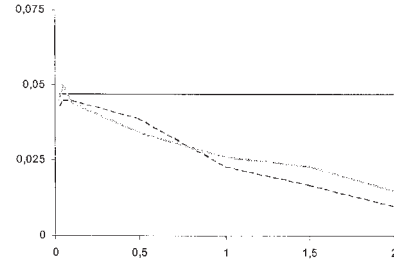
 $DGP''_{\phi}, n = 250$ 

FIGURE 15

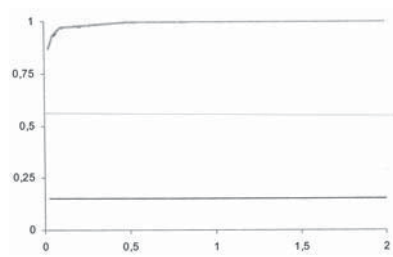
 $DGP''_p, n = 100$ 

FIGURE 16

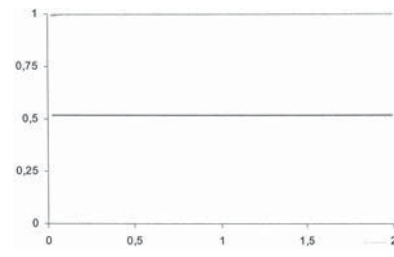
 $DGP''_p, n = 250$ 

FIGURE 17

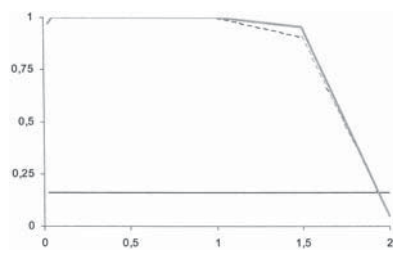
 $DGP''_z, n = 100$ 

FIGURE 18

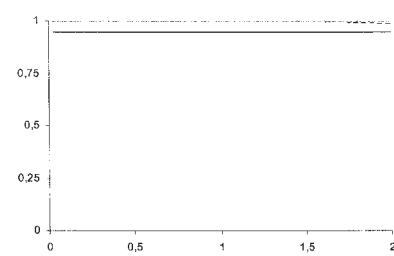
 $DGP''_z, n = 250$ 

FIGURE 19

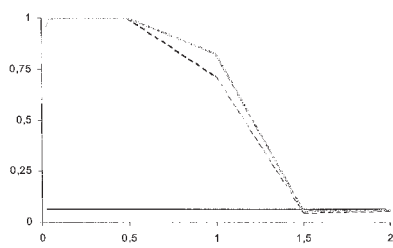
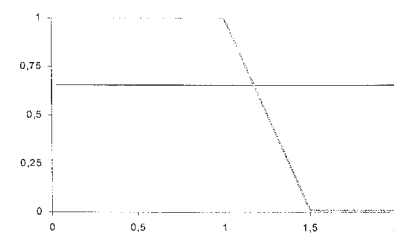
 $DGP''_3, n = 100$ 

FIGURE 20

 $DGP''_3, n = 250$ 

Testing the simultaneous equation model. We finally consider the classic exactly identified linear simultaneous equation model of Example 3

$$H_0 : \begin{pmatrix} E[Q - a_0P - b_0I | I, W] \\ E[Q - \alpha_0P - \beta_0W | I, W] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ a.s. for some } (\alpha_0, \beta_0, a_0, b_0).$$

We generate the data according to

$$\begin{aligned} Q_i &= \lambda_{11}I_i + \lambda_{12}W_i + V_{1i} \\ P_i &= \lambda_{21}I_i + \lambda_{22}W_i + \frac{\delta}{2}W_i^2 + V_{2i} \quad i = 1, \dots, n, \end{aligned}$$

where I_i and W_i are independent standard Gaussian random variables while V_{1i} and V_{2i} are correlated standard Gaussian random variables with $2^{-1/2}$ covariance and independent of (I_i, W_i) . Parameters are estimated using Indirect Least Squares, which is equivalent to a Two-Stage Least Squares procedure in this context. The null hypothesis corresponds to $\delta = 0$ and is denoted DGP_0''' , while DGP_δ''' denotes the alternative. Other details are otherwise unchanged.

Results are reported in Figures 21 to 24. The graphs show that under DGP_0''' the empirical level of the smooth test is very close to the nominal one for all bandwidths even when the sample size is as small as 100. On the other hand, the inte-

FIGURE 21
 DGP_φ''' , $n = 100$

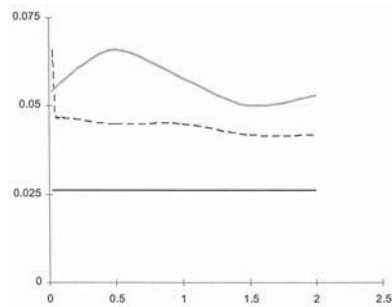


FIGURE 22
 DGP_φ''' , $n = 250$

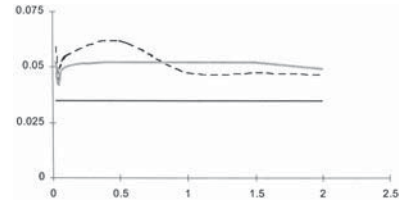


FIGURE 23
 DGP_p''' , $n = 100$

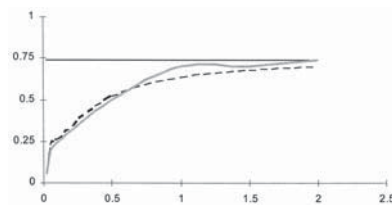
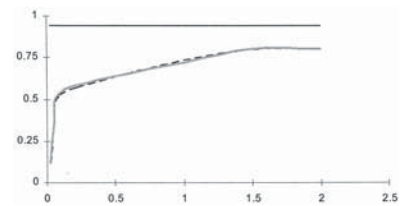


FIGURE 24
 DGP_p''' , $n = 250$



gral-transform test underrejects for $n = 100$ but the level is fairly close to its nominal value for $n = 250$. Under the alternative hypothesis, we keep the sample size fixed at $n = 100$ and allow δ to vary. The power of the smooth test increases with the bandwidth, but the integral-transform test dominates its competitor under both alternatives with a power as high as 95%.

Our results sheds light on the comparative behavior of the tests in varied situations. For the smooth test, the choice of the smoothing parameter is revealed crucial. Moreover, use of simulated critical values yields limited improvement for both size and power. Finally, no clear ranking emerges between the two tests. The overall performances of the tests depend on the particular experiment's features.

6 Conclusion

We have shown in this paper how the two approaches 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. Clearly, 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 of bandwidths' choice, as investigated by HOROWITZ and SPOKOINY [2001] and GUERRE and LAVERGNE [2005] for the regression model. Third, bootstrap procedures for computing critical values should be investigated. We have explained the difficulties related to resampling in a general context and we have subsequently proposed a simulation technique for computing critical values in small samples. Our experiments results illustrate that some limited improvement is expected using simulated critical values instead of asymptotic ones for the smooth test. Unreported simulations results suggest that more sophisticated methods can substantially improve upon our simulation technique, but the resampling scheme heavily depends on the model at hand. In this respect, a general accurate resampling method is still required.

7 Proofs

PROOF OF THEOREM 1: Henceforth, for $i = 1, 2, \dots, n$ and $k = 1, \dots, m$, $\psi_i^{(k)} = \psi^{(k)}(Z_i, \theta_0)$, $\widehat{\psi}_i^{(k)} = \widehat{\psi}^{(k)}(Z_i, \theta_n)$, $\dot{\psi}_i^{(k)} = \dot{\psi}^{(k)}(Z_i, \theta_0)$, $\ddot{\psi}_i^{(k)}(\theta) = \ddot{\psi}^{(k)}(Z_i, \theta)$, and $\Sigma_{i \neq j}$ stand for $\sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{i \neq j}$. We have

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

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

$$T_{1n}^{(k)} = \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{(k)} \left(\widehat{\psi}_j^{(k)} - \psi_j^{(k)} \right) K_{ij},$$

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

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,

$$T_{1n}^{(k)} = (\theta_n - \theta)' S_{1n} + (\theta_n - \theta_0)' S_{2n} (\theta_n - \theta_0) \quad \text{and} \quad T_{2n}^{(k)} = (\theta_n - \theta_0)' S_{3n} (\theta_n - \theta_0),$$

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

$$\text{with } \|\bar{\theta}_n - \theta_0\| \leq \|\theta_n - \theta_0\| \quad \text{and} \quad \|\tilde{\theta}_n - \theta_0\| \leq \|\theta_n - \theta_0\|.$$

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 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 $H_n(Z_1, Z_2) = h^{-q} \left(\psi_1^{(k)} \dot{\psi}_2^{(k)} + \psi_2^{(k)} \dot{\psi}_1^{(k)} \right) K_{12} / 2$ and

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

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

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

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 (6). We use of the following result for degenerate U -statistics.

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

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

The first term of (6) 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

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

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

by Assumptions 2.4-2.6. Moreover,

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

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

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 (6) 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,

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

and T_n converges to a strictly positive limit when H_0 does not hold. A similar reasoning shows that $V_n \rightarrow_p V$ whether H_0 holds or not.

PROOF OF THEOREM 3: Under the null hypothesis, we shall show that $n\widehat{h}^{q/2}T_n(\widehat{h}) - nh^{q/2}T_n(h) = o_p(1)$, 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 $n(vh)^{q/2}T_n(vh)$ for $v \in [B_1, B_2]$ with $0 < B_1 < 1 < B_2 < \infty$. It can be seen that the second and third term in (6) are both $O_p(n^{-1})$ uniformly for $v \in [B_1, B_2]$. Let $\tilde{T}_n(h)$ be the first term in (6). It is asymptotically normal at a fixed point and converges to the same limit for any v . Moreover for $v_1, v_2 \in [B_1, B_2]$,

$$\begin{aligned} & E \left[n(v_1h)^{q/2} \tilde{T}_n(v_1h) - n(v_2h)^{q/2} \tilde{T}_n(v_2h) \right]^2 \\ &= \frac{4n}{n-1} \sum_{k=1}^m \sum_{l=1}^m E \left\{ \sigma_{kl}(X_1) \sigma_{kl}(X_2) \left[(v_1h)^{-q/2} K \left(\frac{X_1 - X_2}{v_1h} \right) \right. \right. \\ & \qquad \qquad \qquad \left. \left. - (v_2h)^{-q/2} K \left(\frac{X_1 - X_2}{v_2h} \right) \right] \right\}^2 \end{aligned}$$

is $O[(v_1 - v_2)^2]$ by a Taylor expansion of $v_2^{-q/2}K[x/v_2]$ around v_1 , using Assumption 2.6 and Hölder's inequality. Hence $n(vh)^{q/2}\tilde{T}_n$ is tight for $v \in [B_1, B_2]$, see BILLINGSLEY [1968]. Under the alternative hypothesis, it is sufficient to show that $T_n(vh)$ is tight for $v_1, v_2 \in [B_1, B_2]$, which is shown similarly. An analogous result for V_n then implies the desired result.

PROOF OF THEOREM 4: We have 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

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

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 (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 Skorohod 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

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

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

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

for all $k = 1, \dots, m$, where μ is an arbitrary measure. Applying STUTE'S [1997] Lemma 1,

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

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 $\alpha_i \beta_i = 0$,

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

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 then an immediate consequence of the continuous mapping theorem.

PROOF OF THEOREM 6: Henceforth, for $i = 1, 2, \dots, n$ and $k = 1, \dots, m$, $\psi_i^{*(k)} = \psi^{(k)}(Z_i, \theta^*) \zeta_i$, $\widehat{\psi}_i^{*(k)} = \psi^{(k)}(Z_i, \theta_n) \zeta_i$, $\dot{\psi}_i^{*(k)} = \dot{\psi}^{(k)}(Z_i, \theta^*) \zeta_i$ and $E^* [\cdot] \equiv E[\cdot | Z_n]$. We have a decomposition similar to (6), that is,

$$(7) \quad T_n^* = \binom{n}{2}^{-1} \sum_{i \neq j} D_n^{*(k)}(Z_i, Z_j) + \sum_{k=1}^m 2T_{1n}^{*(k)} + T_{2n}^{*(k)},$$

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 (6), with $\psi_i^{*(k)}$ and $\widehat{\psi}_i^{*(k)}$ in place of $\psi_i^{(k)}$ and $\widehat{\psi}_i^{(k)}$. 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,

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

$$\text{where } S_{1n}^* = \frac{1}{n(n-1)h^q} \sum_{i \neq j} \psi_i^{*(k)} \dot{\psi}_j^{*(k)} K_{ij},$$

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

and $\ddot{\psi}_j^{*(k)}(\theta) = \ddot{\psi}^{(k)}(Z_j, \theta) \zeta_j$. 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 ,

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

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, $E[E^*[S_{2n}^{*2}]] = O(1)$. 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 (7). For the sake of simplicity, we treat the case where $m = 1$. We then consider

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

where $E^* [D_n^*(Z_i, Z_j) | \xi_i] = 0$, for all i . By Proposition 3.2 in DE JONG [1987], $\sigma_n^{-1} \widehat{T}_n^*$ converges in distribution to a $N(0, 1)$ in probability conditional upon \mathcal{Z}_n if G_1, G_2 , and G_3 are of lower order in probability than $(\sigma_n^2)^2$, where

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

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

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

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 Polya's theorem.

PROOF OF THEOREM 7: It is immediate that $\sup_{x,s} \|\Omega_n(x,s) - \Omega(x,s)\| = O_p(1)$ and $\sup_x \|\bar{R}_n^*(x) - \bar{R}_n^{0*}(x)\| = o_p(n^{-1/2})$ 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 by (5). 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 Cr amer-Wold device, the convergence of the finite distributions follows from the convergence of

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

Asymptotic normality is proved by showing the Lindeberg-Lévy condition, i.e., for each $\delta > 0$,

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

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 the W_i^2 are iid with finite first moment. Then Theorem 7 follows 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

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

The tightness of R_n^{01*} follows from the continuity of $G(\cdot, \cdot)$ 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 law of large numbers,

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

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

References

- ANDREWS D.W.K. (1997). – « A Conditional Kolmogorov Test », *Econometrica*, 65, pp. 1097-128.
- APOSTOL T.M. (1957). – « Mathematical Analysis », Reading: Addison-Wesley.
- BICKEL P.J. and WICHURA M.J. (1971). – « Convergence Criteria for Multiparameter Stochastic Processes and Some Applications », *Annals of Mathematical Statistics*, 42, pp. 1656-70.

- BIERENS H. (1982). – « Consistent Model Specification Test », *Journal of Econometrics*, 20, pp. 105-34.
- BIERENS H. (1990). – « A Consistent Conditional Moment Test of Functional Form », *Econometrica*, 58, 1443-58.
- BIERENS H. and GINTHER D. (2001). – « Integrated Conditional Moment Testing of Quantile Regression Models », *Empirical Economics*, 26, pp. 307-24.
- BIERENS H. and PLOBERGER W. (1997). – « Asymptotic Theory of Integrated Conditional Moment Tests », *Econometrica*, 65, pp. 1129-51.
- BILLINGSLEY P. (1968). – « Convergence of Probability Measures », New-York: John Wiley & Sons.
- BUCKLEY M.J. (1991). – « Detecting a Smooth Signal: Optimality of Cusum Based Procedures », *Biometrika*, 78, pp. 253-62.
- BURBIDGE J.B., MAGEE L. and ROBB A.L. (1988). – « Alternative Transformations to Handle Extreme Values of the Dependent Variable », *Journal of the American Statistical Association*, 83, pp. 123-27.
- CHAMBERLAIN G. (1987). – « Asymptotic Efficiency in Estimation With Conditional Moment Restrictions », *Journal of Econometrics*, 34, pp. 305-34.
- CHEN X. and FAN Y. (1999). – « Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Times Series », *Journal of Econometrics*, 91, pp. 373-401.
- DAROLLES S., FLORENS J-P. and RENAULT E. (1999). – « Nonparametric Instrumental Regression », Université Toulouse 1.
- DAVIDSON R. and MacKinnon J.G. (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, pp. 261-77.
- DE JONG R.M. (1996). – « The Bierens Test Under Data Dependence », *Journal of Econometrics*, 72, pp. 1-32.
- DE JONG R.M. and BIERENS H.J. (1994). – « On Limit Behavior of Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity », *Econometric Theory*, 9, pp. 70-90.
- DELGADO M. (1993). – « Testing the Equality of Nonparametric Regression Curves », *Statistics and Probability Letters*, 17, pp. 199-204.
- DELGADO M. and GONZÁLES-MANTEIGA W. (2001). – « Signification Testing in Nonparametric Regression Based on the Bootstrap », *Annals of Statistics*, 29, pp. 1469-1507.
- DOMINGUEZ M. and LOBATO I. (2003). – « Testing the Martingale Hypothesis », *Econometric Reviews*, 22, pp. 351-77.
- DOMINGUEZ M. and LOBATO I. (2004). – « Consistent Estimation of Models Defined by Conditional Moment Restrictions », *Econometrica*, 72, pp. 1601-15.
- ELLISON G. and ELLISON S. (2000). – « A Simple Framework for Nonparametric Specification Testing », *Journal of Econometrics*, 93, pp. 1-23.
- 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, pp. 387-92.
- FAN Y. and LI Q. (1996). – « Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms », *Econometrica*, 64, pp. 865-90.
- FAN Y. and LINTON O. (2003). – « Some Higher-Order Theory For a Consistent Non-Parametric Model Specification Test », *Journal of Statistics Planning and Inference*, 109, pp. 125-54.
- GAN L. and JIANG J. (1999). – « A Test of Global Maximum », *Journal of the American Statistical Association*, 94, pp. 847-54.
- GUERRE E. and LAVERGNE P. (2005). – « Data-Driven Rate-Optimal Specification Testing in Regression Models », *Annals of Statistics*, forthcoming.
- HALL P. (1984). – « Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators », *Journal of Multivariate Analysis*, 14, pp. 1-16.

- HALL P. and HOROWITZ J.L. (1996). – « Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators », *Econometrica*, 64, pp. 891-916.
- HANSEN B. (1996). – « Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis », *Econometrica*, 64, pp. 413-30.
- HÄRDLE W. and MAMMEN E. (1993). – « Comparing Nonparametric versus Parametric Regression Fits », *Annals of Statistics*, 21, pp. 1926-47.
- HART J.D. (1997). – « Nonparametric Smoothing and Lack-of-Fit Tests », New-York: Springer-Verlag.
- HONG Y. (1993). – « Consistent Testing for Heteroscedasticity of Unknown Form », U.C. San Diego.
- HONG-ZHY A. and BIN C. (1991). – « A Kolmogorov-Smirnov Type Statistic with Application to Test for Nonlinearity in Time Series », *International Statistical Review*, 59, pp. 287-307.
- HOROWITZ J.L. and SPOKOINY V.G. (2001). – « An Adaptive, Rate-Optimal Test of a Parametric Model Against a Nonparametric Alternative », *Econometrica*, 69(3), pp. 599-631.
- KHMALADZE E.V. (1993). – « Goodness of Fit Problem and Scanning Innovation Martingales », *Annals of Statistics*, 21, pp. 798-829.
- KIEFFER J. (1959). – « K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-von-Mises Tests », *Annals of Mathematical Statistics*, 30, pp. 420-47.
- KOUL H.L. and STUTE W. (1999). – « Nonparametric Model Checks for Time Series », *Annals of Statistics*, 27, pp. 204-36.
- KOZEK A.S. (1991). – « A Nonparametric Test Fit of a Parametric Model », *Journal of Multivariate Analysis*, 37, pp. 66-75.
- LAVERGNE P. and THOMAS A. (2005). – « Semiparametric Estimation and Testing in a Model of Environmental Regulation with Adverse Selection », *Empirical Economics*, forthcoming.
- LAVERGNE P. and VUONG Q.H. (2000). – « Nonparametric Significance Testing », *Econometric Theory*, 16, pp. 576-601.
- LI Q. (1999). – « Consistent Model Specification Tests for Time Series Econometric Models », *Journal of Econometrics*, 92, pp. 101-47.
- LI Q. and WANG S. (1998). – « A Simple Consistent Bootstrap Test for a Parametric Regression Function », *Journal of Econometrics*, 87, pp. 145-65.
- MACKINNON J.G. and MAGEE L. (1990). – « Transforming the Dependent Variable in Regression Models », *International Economic Review*, 31, pp. 315-39.
- NEWBY W.K. (1985a). – « Maximum Likelihood Specification Testing and Conditional Moment Tests », *Econometrica*, 53, pp. 1047-70.
- NEWBY W.K. (1985b). – « Generalized Methods of Moments Specification Testing », *Journal of Econometrics*, 29, pp. 229-56.
- NEWBY W.K. (1990). – « Efficient Instrumental Variables Estimation of Nonlinear Models », *Econometrica*, 58, pp. 809-37.
- PRAKASA RAO, B.L.S. (1983). – « Nonparametric Functional Estimation », Orlando: Academic Press.
- POWELL J.L., STOCK J.H. and STOKER T.M. (1989). – « Semiparametric Estimation of Index Coefficients », *Econometrica*, 57, pp. 1403-30.
- ROBINSON P.M. (1991). – « Best Nonlinear Three-Stage Least Squares of Certain Econometric Models », *Econometrica*, 59, pp. 1329-63.
- SERFLING R.J. (1980). – « Approximations Theorems of Mathematical Statistics », New-York: John Wiley & Sons.
- STINCHCOMBE M.B. and WHITE H. (1998). – « Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative », *Econometric Theory*, 14, pp. 295-325.
- STUTE W. (1997). – « Nonparametric Model Checks for Regression », *Annals of Statistics*, 25, pp. 613-41.
- STUTE W., GONZALEZ-MANTEIGA W. and PRESEDO M. (1998). – « Bootstrap Approximations in Model Checks for Regression », *Journal of the American Statistical Associations*, 93, pp. 141-9.

- STUTE W., THIES S. and ZHU L.-X. (1998). – « Model Checks for Regression: an Innovation Process Approach », *Annals of Statistics*, 26, pp. 1916-34.
- SU J.Q. and WEI L.J. (1991). – « A Lack of Fit Test for the Mean Function in a Generalized Linear Model », *Journal of the American Statistical Associations*, 86, pp. 420-26.
- TAUCHEN G. (1985). – « Diagnostic Testing and Evaluation of Maximum Likelihood Models », *Journal of Econometrics*, 30, pp. 415-43.
- WHANG Y.J. (2000). – « Consistent Bootstrap Tests of Parametric Regression Functions », *Journal of Econometrics*, 98, pp. 27-46.
- WHANG Y.J. (2001). – « Consistent Specification Testing for Conditional Moment Restrictions », *Economics Letters*, 71, pp. 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, pp. 17-43.
- ZHENG X. (1996). – « A Consistent Test of Functional Form via Nonparametric Estimation Techniques », *Journal of Econometrics*, 75, pp. 263-89.
- ZHENG X. (1998). – « A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions », *Econometric Theory*, 14, pp. 123-38.

