Econometric Theory, **31**, 2015, 891–910.
doi:10.1017/S0266466614000644

A SIMPLE OMNIBUS OVERIDENTIFICATION SPECIFICATION TEST FOR TIME SERIES ECONOMETRIC MODELS

MANUEL A. DOMÍNGUEZ
Universidad Complutense de Madrid

IGNACIO N. LOBATO
Instituto Tecnológico Autónomo de México

Despite their theoretical advantages, Integrated Conditional Moment (ICM) specification tests are not commonly employed in the econometrics practice. An important reason is that the employed test statistics are nonpivotal, and so critical values are not readily available. This article proposes an omnibus test in the spirit of the ICM tests of Bierens and Ploberger (1997, *Econometrica* 65, 1129–1151) where the test statistic is based on the minimized value of a quadratic function of the residuals of time series econometric models. The proposed test falls under the category of overidentification restriction tests started by Sargan (1958, *Econometrica* 26, 393–415). The corresponding projection interpretation leads us to propose a straightforward wild bootstrap procedure that requires only linear regressions to estimate the critical values irrespective of the model functional form. Hence, contrary to other existing ICM tests, the critical values are easily calculated while the test preserves the admissibility property of ICM tests.

1. INTRODUCTION

In econometrics it is common to assess the suitability of a model using the minimized value of the objective function employed for estimation. This practice links naturally both aspects of inference, estimation and specification testing. Sargan's (1958) test in two-stage least squares estimation and Hansen's (1982) overidentification test in a GMM context are two leading examples. However, it is well known that these tests are not omnibus.

Omnibus specification testing of econometric models has a relatively short history, starting in Bierens (1982) (see Delgado, Domínguez, and Lavergne, 2006 for

We thank an associate editor and a referee for useful comments, and Marine Carrasco, Miguel Delgado, and Juan Carlos Escanciano for useful conversations. Lobato acknowledges financial support from the Mexican CONACYT, reference number 151624, and from Asociación Mexicana de Cultura. This article was also supported by Spanish Ministry of Education, Plan Nacional I+D+i, SEJ2007-62908. Address correspondence to Ignacio Lobato, Centro de Investigación Económica, Instituto Tecnológico Autónomo de México (ITAM), Av. Camino a Santa Teresa #930, 10700 México, D.F., Mexico; e-mail: ilobato@itam.mx.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

an overview). A very fruitful approach to omnibus testing is the one based on the integrated regression function. This approach leads to tests with a superior local power compared to other procedures. However, despite their attractive theoretical properties, these omnibus tests are not frequently employed in the econometrics practice. One reason may be that, as shown in Escanciano (2009), despite their omnibus nature, these tests have substantial local power only against alternatives in a finite-dimensional space. This could be a reason when the sample size is small or moderate, but even in studies with huge sample sizes, applied researchers rarely report omnibus specification tests. In our opinion, the reasons are that omnibus tests are stated in unfamiliar terms for applied econometricians, are difficult to apply, and require complicated adjustments to get critical values of the test statistics. The target of this article is to propose an omnibus test that overcomes these difficulties.

In the spirit of overidentification tests, this article introduces a new omnibus specification test, based on the integrated conditional expectation, which employs as test statistic the minimized value of the objective function considered in Domínguez and Lobato (2004, DL hereafter). Although the value minimizing the objective function is not an efficient estimator of the parameters, we show that the test is still omnibus. Contrary to the standard practice that emphasizes the use of efficient estimators, we point out that the method employed for estimating the unknown parameter value is not critical for the admissibility of the test as long as the estimator is root-n consistent. Therefore, from the theoretical point of view, the test is comparable to other tests based on the integrated regression function.

The proposed test enjoys some practical advantages over existing ICM tests. First the test is easier to apply than existing ICM tests since its critical values can be easily calculated, using a novel bootstrap procedure that only involves running linear regressions irrespective of the complexity of the nonlinear model. The bootstrap procedure is the main practical contribution of this article because it avoids level distortions derived from employing upper bound critical values, as in Bierens and Ploberger (1997), complicated bootstrap resampling schemes, as in Delgado et al. (2006), or restrictive transformations, as in Koul and Stute (1999). The second advantage is that the test recovers Sargan's and Hansen's approach and allows to write the whole inferential procedure in very familiar terms to econometrics practitioners, which will encourage the application of these methods in empirical research. The proposed test statistic can be thought in terms of projections and in terms of identifying and overidentifying restrictions, which leads to a better understanding of the testing procedure and of the local alternatives that the test is able to detect. In addition, the test statistic is easily calculated since it is a by-product of the parameter estimation stage.

The plan of this article is the following. Section 2 introduces the proposed specification test, establishes its asymptotic properties, and presents it as an overidentification goodness of fit test. Section 3 proposes the simple bootstrap procedure for estimating its critical values, and Section 4 presents a Monte Carlo study to

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

illustrate the irrelevance in finite samples of plugging into the test statistic an efficient or an inefficient estimator. Section 5 concludes.

2. THE TEST

Concerning notation and assumptions, we follow DL. Then, z_t is a p-dimensional strictly stationary ergodic time series vector and $\{y_t, x_t\}$ are two subvectors of z_t (that could have common coordinates), where y_t is a m-dimensional time series vector and x_t is a d-dimensional time series vector that contains the exogenous variables. The coordinates of z_t are related by the following conditional moment restriction (CMR) model

$$E(u(y_t, \theta_0) \mid x_t) = 0, \quad a.s. \tag{1}$$

for a unique value $\theta_0 \in \Theta$, where $\Theta \subset \mathbb{R}^k$ is the parametric space. Equation (1) defines the unknown true value θ_0 for the parameter θ. For completeness we state

Assumption A (see DL)**:**

(a) $u(y, \cdot)$ is continuous in Θ for each y in $\mathbb{R}^k$, $|u(y_t, \theta)| < k(y_t)$ with $Ek(y_t) < \infty$ and $E(u(y_t, \theta) \mid X_t) = 0$ $a.s.$ (almost surely) if and only if $\theta = \theta_0$.

(b) Z_t is ergodic and strictly stationary.

(c) $\Theta \subset \mathbb{R}^m$ is compact.

(d) $u(y, \cdot)$ is once continuously differentiable in a neighborhood of θ_0 and satisfies that $E\left[\sup_{\theta \in \aleph_0} |\dot{u}(Y_t, \theta)|\right] < \infty$ where $\dot{u}(y, \theta) = \frac{\partial u(y, \theta)}{\partial \theta}$, and $\aleph_0$ denotes a neighborhood of θ_0.

(e) $u(y_t, \theta_0)$ is a martingale difference sequence with respect to $\{Z_s, s \leq t\}$.

(f) $\theta_0 \in int(\Theta)$.

(g) $E\left[u^4(y_t, \theta_0) \| X_t \|^{1+\delta}\right] < \infty$.

(h) The density of the conditioning variables given the past is bounded and continuous.

For regularity we require that the function u that maps $\mathbb{R}^m \times \Theta$ into $\mathbb{R}^l$ is known and its fourth moment is bounded. We also assume that u is smooth in the second argument: in particular, it is partially differentiable and its first derivatives have uniformly bounded first moments in a neighborhood of θ_0 (note that smoothness on u can be substituted by smoothness on its expected value at the cost of some additional technical details). The condition (h) is a technical assumption required in DL to obtain the asymptotic distribution of the CMM estimator. For i.i.d. (independent and identically distributed) data assumptions (g) and (h) can be removed and replaced by the existence of the first two moments. In addition, from a statistical point of view, the function $u(y_t, \theta_0)$ is a martingale difference sequence

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

with respect to the increasing sequence of sigma fields defined by the sequence of conditioning variables. In general, $u(y_t, \theta_0)$ can be understood as the errors in a multivariate nonlinear dynamic regression model; for instance, $u(y_t, \theta_0)$ are called generalized residuals in Wooldridge (1990). This model has been repeatedly considered in the econometrics literature and several estimators have been proposed (see references in DL for estimation and Delgado et al., 2006 for a specification testing framework). For simplicity, in this article we consider the case $l = 1$.

For any random vector a we denote P_a and $P_{a,n}$ to the distribution and empirical distribution function of a respectively. For $\mathtt{x} \in \mathbb{R}^d$, let $I(x_t \leq \mathtt{x})$ denote the indicator function that equals one when x_t is less or equal than $\mathtt{x}$ coordinatewise; and equals zero otherwise. Define the integrated errors[1]

$$U(\mathtt{u}, \theta) = \int_{\mathbb{R}^p} u(y, \theta) I(x \leq \mathtt{u}) P_{z_t}(dz). \tag{2}$$

The CMR (1) is equivalent to the set of unconditional moment restrictions

$$U(\mathtt{u}, \theta_0) = 0, \text{ for almost all } \mathtt{u} \in \mathbb{R}^d, \tag{3}$$

see Billingsley (1995, Thm. 16iii) and Stute (1997), and θ_0 is the unique value minimizing

$$Q(\theta) = \int_{\mathbb{R}^d} U(\mathtt{u}, \theta)^2 P_{x_t}(d\mathtt{u}). \tag{4}$$

Note that $Q(\theta)$ equals zero at θ_0 if and only if the CMR holds. Hence, $\theta_0 = \arg\min_\theta Q(\theta)$ is a minimum distance parameter. The consistent method of moments (CMM) estimator of θ_0 is defined by its sample analog, i.e., it is a minimum distance estimator. Formally, for a sample of size n, define the sample analogs of $U(\mathtt{u}, \theta)$ and $Q(\theta)$ by

$$U_n(\mathtt{u}, \theta) = \int_{\mathbb{R}^p} u(y, \theta_0) I(x \leq \mathtt{u}) P_{z_t,n}(dz) = \frac{1}{n} \sum_{t=1}^{n} u(y_t, \theta) I(x_t \leq \mathtt{u}), \tag{5}$$

and

$$Q_n(\theta) = \int_{\mathbb{R}^d} U_n(\mathtt{u}, \theta)^2 P_{x_t,n}(d\mathtt{u}) = \frac{1}{n} \sum_{\ell=1}^{n} U_n(x_\ell, \theta)^2. \tag{6}$$

Then, the CMM estimator is defined by $\hat{\theta} = \arg\min_{\theta \in \Theta} Q_n(\theta)$.

For this model, DL showed that $\hat{\theta}$ is consistent (see DL for further details). Define also

$$\dot{U}(\mathtt{u}, \theta) = \int_{\mathbb{R}^p} [\dot{u}(y, \theta) I(x \leq \mathtt{u})] P_{z_t}(dz),$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

and its sample analog

$$\dot{U}_n(\mathtt{u},\theta)=\int_{\mathbb{R}^p}[\dot{u}(y,\theta)\,I(x\leq\mathtt{u})]\,P_{z_t,n}(dz)=\frac{1}{n}\sum_{t=1}^{n}\dot{u}(y_t,\theta)\,I(x_t\leq\mathtt{u}), \tag{7}$$

is the $1\times k$ vector of the first order derivatives. DL also showed that

$$\sqrt{n}(\widehat{\theta}-\theta_0)=-\left\{\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,\dot{U}_n(x_\ell,\theta_0)\right\}^{-1}\times\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,U_n(x_\ell,\theta_0)+o_p(1) \tag{8}$$

and used this linear representation to prove that $\sqrt{n}(\hat{\theta}-\theta_0)$ is asymptotically a centered Gaussian random variable whose covariance matrix is

$$\left\{\int_{\mathbb{R}^d}\dot{U}'(\mathtt{u},\theta_0)\,\dot{U}(\mathtt{u},\theta_0)\,P_{x_t}(d\mathtt{u})\right\}^{-1}\Gamma_2\left\{\int_{\mathbb{R}^d}\dot{U}'(\mathtt{u},\theta_0)\,\dot{U}(\mathtt{u},\theta_0)\,P_{x_t}(d\mathtt{u})\right\}^{-1}, \tag{9}$$

where

$$\Gamma_2=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\dot{U}'(\mathtt{u},\theta_0)\,\Gamma(\mathtt{u},\mathtt{v})\,\dot{U}(\mathtt{v},\theta_0)\,P_{x_t}(d\mathtt{u})\,P_{x_t}(d\mathtt{v}),$$

$$\Gamma(\mathtt{u},\mathtt{v})=E\left(u^2(y_t,\theta_0)I(x_t\leq\mathtt{u}\wedge\mathtt{v})\right).$$

Once the parameters of interest are estimated, we propose using the scaled minimized value of the objective function $Q_n(\hat{\theta})$ for testing whether the model is correctly specified. Formally, consider the null hypothesis

$$H_0: E(u(y_t,\theta_0)\mid x_t)=0\quad a.s., \tag{10}$$

whereas the alternative hypothesis is that for any θ

$$H_1: P(E(u(y_t,\theta)\mid x_t)=0)<1.$$

In this respect, note that H_1 is the negation of H_0 under Assumption A and we rule out a set identification framework. We propose testing model adequacy using the test statistic

$$T_n\;=\;nQ_n(\hat{\theta}). \tag{11}$$

2.1. Behavior of T_n Under the Null Hypothesis

In order to derive the asymptotic null distribution of T_n, we consider first the behavior of $U_n(\mathtt{u},\hat{\theta})$ under H_0. Starting from equation (5), using (8), and applying results in DL to the developments in Stute (1997), it follows that, see DL and Escanciano (2006) for details, uniformly in $\mathtt{u}$

$$\sqrt{n}U_n(\mathtt{u},\hat{\theta})=M_n\sqrt{n}U_n(\mathtt{u},\theta_0),$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

where M_n is the projection operator defined by

$$M_n r(\mathtt{u}) = r(\mathtt{u}) - \dot{U}_n(\mathtt{u}, \theta_0) \left\{ \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0) \right\}^{-1} \times \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) r(x_\ell, \theta_0). \tag{12}$$

Note M_n projects r into the orthogonal space of $\dot{U}_n$ and hence, it is a continuous operator. Since, $\sqrt{n} U_n(\mathtt{u}, \theta_0) \to_d B_\Gamma(\mathtt{u})$, where convergence is in the space of cadlag functions, $D[-\infty, \infty]^d$, B_Γ stands for a centered Gaussian process with covariance structure Γ, and $M_n \to M$, where M is the projection operator defined by

$$Mr(\mathtt{u}) = r(\mathtt{u}) - \dot{U}(\mathtt{u}, \theta_0) \left\{ \int_{\mathbb{R}^d} \dot{U}'(\mathtt{v}, \theta_0) \dot{U}(\mathtt{v}, \theta_0) P_{x_t}(d\mathtt{v}) \right\}^{-1} \times \int_{\mathbb{R}^d} \dot{U}'(\mathtt{v}, \theta_0) r(\mathtt{v}) P_{x_t}(d\mathtt{v}).$$

Then,

$$\sqrt{n} U_n(\mathtt{u}, \hat{\theta}) \to_d B_\Phi(\mathtt{u}) \equiv M B_\Gamma(\mathtt{u}), \tag{13}$$

where B_Φ stands for a centered Gaussian process with corresponding covariance structure Φ. As a result we establish the following proposition.

PROPOSITION 1. *Under the null hypothesis and Assumption A,*

$$T_n \to_d T = \int_{\mathbb{R}^d} B_\Phi^2(\mathtt{u}) P_{x_t}(d\mathtt{u}).$$

2.2. Behavior of T_n Under the Alternative Hypothesis

Concerning the behavior under the alternative, it is straightforward to show that the test statistic T_n diverges under any fixed alternative. Before formally stating the result in the next proposition, we introduce

Assumption B:

(a) $\hat{\theta} \to_{a.s.} \theta_1$.

(b) $u(y, \cdot)$ is continuous at θ_1 for each y and either (b1) or (b2) holds: $\forall \theta \in \aleph_1$ neighborhood of θ_1
(b1) $\exists k(\cdot)$ such that $|u(y, \theta)| \leq k(y)$ and $Ek(Y_t) < \infty$;
(b2) $\exists k(\cdot)$ such that $|u(y, \theta) - u(y, \theta_1)| \leq k(y) |\theta - \theta_1|$ and $Ek(Y_t) < \infty$.

(c) *Z_t is ergodic and strictly stationary.*

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

PROPOSITION 2. *Under the alternative hypothesis and Assumption B, the proposed test is consistent, that is,*

$$P[T_n > \zeta] \to 1 \text{ as } n \to \infty, \quad \text{for all } \zeta \in \mathbb{R}.$$

Next we consider the behavior against local alternatives. Specifically, let

$$H_{1,n} : E(u(y_t, \theta_0) \mid x_t) = \frac{g(x_t)}{\sqrt{n}} \quad a.s. \tag{14}$$

be the sequence of local alternatives of interest. Establishing the properties of the test in this setting demands properties of the estimator in the same setting. We introduce first

Assumption C:

(a) Replace u by $v_n(z_t, \theta) \equiv u(y_t, \theta) - n^{-1/2} g(x_t)$ in Assumption A.

(b) $E|g(x_t)| < \infty$.

Note that under $H_{1,n}$ and Assumption C, it follows that $U_n(\mathtt{u}, \theta)$ equals

$$\frac{1}{n}\sum_{t=1}^{n} [u(y_t, \theta) - u(y_t, \theta_0)] I(x_t \le \mathtt{u}) + \frac{1}{n}\sum_{t=1}^{n} v_n(z_t, \theta_0) I(x_t \le \mathtt{u})$$
$$+ \frac{1}{n}\sum_{t=1}^{n} n^{-1/2} g(x_t) I(x_t \le \mathtt{u}) \to_{as} U(\mathtt{u}, \theta_0) - U(\mathtt{u}, \theta),$$

uniformly in $\mathtt{u}$, where $U(\mathtt{u}, \theta)$ is defined in (2). Hence,

$$Q_n(\theta) \to_{as} \int_{\mathbb{R}^d} [U(\mathtt{u}, \theta_0) - U(\mathtt{u}, \theta)]^2 P_{x_t}(d\mathtt{u}).$$

As a result, θ_0 still minimizes (4), which implies the following result.

PROPOSITION 3. *Under $H_{1,n}$ and Assumption C, $\hat{\theta} \to_{a.s.} \theta_0$.*

In this local alternative framework, the local linear approximation for $\hat{\theta}$ switches from (8) to

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{\frac{1}{n}\sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0)\right\}^{-1} \frac{1}{\sqrt{n}}\sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) V_n(x_\ell, \theta_0)$$
$$-\left\{\frac{1}{n}\sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0)\right\}^{-1} \frac{1}{n}\sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) G_n(x_\ell)$$
$$+ o_p(1), \tag{15}$$

where

$$V_n(\mathtt{u}, \theta) = \frac{1}{n}\sum_{t=1}^{n} v_n(z_t, \theta) I(x_t \le \mathtt{u}), \text{ and } G_n(\mathtt{u}) = \frac{1}{n}\sum_{t=1}^{n} g(x_t) I(x_t \le \mathtt{u}).$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

Since $v_n(z_t, \theta)$ is a martingale difference in this local alternative framework, we can apply the functional central limit theorem to V_n, just as we applied it to U_n, leading to $V_n \to_d B_\Gamma$. Hence, although $\hat{\theta}$ is still consistent, its asymptotic distribution is not centered at 0, as the next proposition states.

PROPOSITION 4. *Under $H_{1,n}$ and Assumption C, $\sqrt{n}(\widehat{\theta} - \theta_0)$ is asymptotically normal with mean given by $\left\{\int_{\mathbb{R}^d} \dot{U}'(\mathbf{u}, \theta_0)\dot{U}(\mathbf{u}, \theta_0) P_{x_t}(d\mathbf{u})\right\}^{-1}\left\{\int_{\mathbb{R}^d} \dot{U}'(\mathbf{u}, \theta_0) G(\mathbf{u}) P_{x_t}(d\mathbf{u})\right\}$ where $G(\mathbf{u}) = \int_{\mathbb{R}^p}[g(x) I(x \le \mathbf{u})] P_{z_t}(dx)$ and variance given by expression (9) where now $\Gamma(\mathrm{u}, \mathrm{v}) = E(v_n^2(y_t, \theta_0) I(x_t \le \mathrm{u} \wedge \mathrm{v}))$.*

These calculations allow us to derive the asymptotic distribution of T_n under $H_{A,n}$ (see Escanciano, 2006). Note that under $H_{A,n}$, $\sqrt{n}U_n(\mathbf{u}, \hat{\theta})$ equals

$$\begin{aligned}\sqrt{n}V_n(\mathrm{u}, \hat{\theta}) + G_n(\mathrm{u}) &= \sqrt{n}V_n(\mathrm{u}, \theta_0) + \dot{U}_n(\mathrm{u}, \theta_0)\sqrt{n}\left(\hat{\theta} - \theta_0\right) + G_n(\mathrm{u}) + o_p(1) \\ &= M_n(\sqrt{n}V_n(\mathrm{u}, \theta_0) + G_n(\mathrm{u})). \end{aligned} \tag{16}$$

The first addend behaves as the linear approximation to U_n under H_0, and the second introduces a mean term in the process U_n. Then, the asymptotic distribution of T_n under $H_{1,n}$ is the same Gaussian process obtained under the null, but centered at the function C where

$$\begin{aligned} C(\mathrm{v}) &= G(\mathrm{v}) - \dot{U}(\mathrm{v}, \theta_0)\left\{\int_{\mathbb{R}^d} \dot{U}'(\mathrm{u}, \theta_0)\dot{U}(\mathrm{u}, \theta_0) P_{x_t}(d\mathrm{u})\right\}^{-1} \\ &\quad \times \int_{\mathbb{R}^d} \dot{U}'(\mathrm{u}, \theta_0) G(\mathrm{u}) P_{x_t}(d\mathrm{u}) \\ &= MG(\mathrm{v}). \end{aligned} \tag{17}$$

Formally, we establish the following proposition.

PROPOSITION 5. *Under the local alternative hypothesis (14) and Assumption C,*

$$T_n \to_d T = \int_{\mathbb{R}^d} [B_\Phi(\mathrm{u}) + C(\mathrm{u})]^2 P_{x_t}(d\mathrm{u}).$$

Remark 1. Proposition 5 points out that T_n converges to a different limit under $H_{1,n}$ except when $C = 0$. In such case, the limit distribution under H_0 and $H_{1,n}$ is the same and T_n cannot detect the deviation from H_0. Since $C(\mathrm{v})$ can be seen as the projection of $G(\mathrm{v})$ onto the orthogonal space of $\dot{U}(\mathrm{v}, \theta_0)$, the case of $C = 0$ occurs only when the $G(\mathrm{v})$ is a linear combination of $\dot{U}(\mathrm{v}, \theta_0)$, that is, when $g_t = \dot{u}_t\beta$ for some β. For the linear regression model, such alternative model means $y_t = x_t\theta_0 + x_t\beta/\sqrt{n} + u_t$, and $H_{1,n}$ is not really a deviation from the linear model. For more general regression models, where $u_t = y_t - m(x_t, \theta)$, such a g_t leads to an alternative model equivalent to the null up to linear terms and, since first order asymptotics are based on linear approximations, that alternative cannot be detected. When u_t does not come from a regression model, and the

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

alternative depends only on the conditioning variables, the condition $g_t = \dot{u}_t\beta$ cannot hold since $\dot{u}_t$ will depend on y_t, and the alternative is always detected. Finally, we stress that the lack of local power that occurs when $C = 0$ is not a particular problem associated to the estimator $\hat{\theta}$. This problem is a side effect of the necessity of estimating θ_0.

Once it has been established that T_n behaves differently under $H_{1,n}$ because of the mean term C, the following proposition states that the power of T_n against $H_{1,n}$ is nontrivial.

PROPOSITION 6. *Under the local alternative hypothesis (14) and* $g_t \neq \dot{u}_t\beta$ *for all* β, *and Assumption C,*

$$P\left[T_n > F_T^{-1}(1-\alpha)\right] > \alpha,$$

where F_T *denotes the distribution of* T.

This result is similar to the one obtained by Stute (1997) in an i.i.d. framework. In that paper, he considers a related test statistic and allows for general linear estimators, such as

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} l_t(z_t,\theta_0) + o_p(1).$$

In the Appendix it is shown that CMM estimator satisfies the linear representation with $l_t(z_t,\theta_0) = D(x_t)\,u(y_t,\theta_0)$, where

$$D(x) = \left\{\int_{\mathbb{R}^d} \dot{U}'(\mathtt{u},\theta_0)\,\dot{U}(\mathtt{u},\theta_0)\,P_{x_t}(d\mathtt{u})\right\}^{-1} \int_{\mathbb{R}^d} \dot{U}'(\mathtt{u},\theta_0)\,1(x \leq \mathtt{u})\,P_{x_t}(d\mathtt{u}).$$

Hence, the result is similar to Stute (1997) but for our general martingale framework.

Neyman–Pearson optimality can also be studied. For instance, Stute (1997) obtained optimal specification tests for the simple linear regression model where the parameters are estimated by OLS. In that restrictive framework, Stute decomposed B_Φ into the principal components given by the eigenfunctions of Φ and showed how to build Neyman–Pearson optimal tests using a finite number of these components. However, these Neyman–Pearson optimal tests are inconsistent since they do not have any power in the direction of the omitted components.

Those optimality results are of limited interest for us since they do not apply to T_n, which contains all the components. Instead of that approach, the admissibility of the full principal component representation of T_n test can be easily proved by showing that there does not exist a test that is uniformly more powerful than T_n. Admissibility is a weak optimality concept, which is proved, as in Bierens and Ploberger (1997), by showing that T_n is asymptotically equivalent to a likelihood ratio test for an alternative model that is a weighted average of local alternative

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

models defined as linear combinations of the Fourier coefficients of the functions $1(x_t < \mathtt{u}) - \dot{U}(\mathtt{u})\left\{\int_{\mathbb{R}^d} \dot{U}'(\mathtt{u},\theta_0)\,\dot{U}(\mathtt{u},\theta_0)\,P_{x_t}(d\mathtt{u})\right\}^{-1} \dot{u}(y_t,\theta_0)$ with respect to the eigenvectors of the covariance structure Φ. Note that this local alternative is tailored to the CMM estimator. The proof is omitted since it follows identically from Bierens and Ploberger, noticing that their differentiability assumption in A5 is not employed in their admissibility proof.

PROPOSITION 7. *Under Assumptions A and C, T_n is asymptotically admissible for testing H_0.*

2.3. The Test as an Overidentifying Restriction Test

According to Proposition 1, the asymptotic distribution of the test statistic T_n is not distribution free and therefore the test based on asymptotic critical values is not feasible. We propose to use bootstrap critical values to run the test. Before describing the resampling scheme, we derive an alternative expression for T_n that motivates the bootstrap.

The objective function in matrix notation can be written as a sum of squared residuals as follows: $Q_n(\theta) = n^{-1}\mathbb{U}(\theta)'\mathbb{U}(\theta)$, where $\mathbb{U}(\theta)$ is the $n \times 1$ vector whose t-th element is $U_n(x_t,\theta)$, that is, the integrated generalized errors. In addition, let $\dot{\mathbb{U}}(\theta)$ be the $n \times k$ matrix whose t-th row is the $1 \times k$ vector $\dot{U}_n(x_t,\theta)$. Then,

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \sqrt{n}\left(\dot{\mathbb{U}}(\theta_0)'\dot{\mathbb{U}}(\theta_0)\right)^{-1}\dot{\mathbb{U}}(\theta_0)'\mathbb{U}(\theta_0) + o_p(1). \tag{18}$$

Expression (18) is similar to the linearization for the nonlinear least squares (NLS) estimator. Hence, in our case the parameter estimators can be interpreted as the values that define the orthogonal projection of the "integrated errors" ($\mathbb{U}(\theta_0)$) onto the subspace generated by the "integrated derivatives" ($\dot{\mathbb{U}}(\theta_0)$).

The test (11) can be written as

$$T_n = \mathbb{U}(\hat{\theta})'\mathbb{U}(\hat{\theta}),$$

so that further intuition can be gained by employing a straightforward asymptotic expansion of the objective function

$$\mathbb{U}(\theta_0)'\mathbb{U}(\theta_0) = \mathbb{U}(\hat{\theta})'\mathbb{U}(\hat{\theta}) + (\hat{\theta} - \theta_0)'\left(\dot{\mathbb{U}}(\hat{\theta})'\dot{\mathbb{U}}(\hat{\theta})\right)(\hat{\theta} - \theta_0) + o_p(1). \tag{19}$$

Equation (19) points out that $\mathbb{U}(\hat{\theta})'\mathbb{U}(\hat{\theta})$ underestimates $\mathbb{U}(\theta_0)'\mathbb{U}(\theta_0)$, just as in the linear regression model the sample variance underestimates the population variance, which motivates dividing the sum of squared residuals by $n-k$ instead of n. More interesting is to rewrite our proposed test statistic as

$$T_n = \mathbb{U}(\theta_0)'\mathbb{U}(\theta_0) - \mathbb{U}(\theta_0)'\dot{\mathbb{U}}(\theta_0)\left(\dot{\mathbb{U}}(\theta_0)'\dot{\mathbb{U}}(\theta_0)\right)^{-1}\dot{\mathbb{U}}(\theta_0)'\mathbb{U}(\theta_0) + o_p(1). \tag{20}$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

Now define

$$\mathbb{P}(\theta_0) = \dot{\mathbb{U}}(\theta_0)\left(\dot{\mathbb{U}}(\theta_0)'\dot{\mathbb{U}}(\theta_0)\right)^{-1}\dot{\mathbb{U}}(\theta_0)',$$
$$\mathbb{M}(\theta_0) = \mathbb{I} - \mathbb{P}(\theta_0),$$

where the projection matrix $\mathbb{M}(\theta_0)$ can be seen as the matrix version of the L^2-projection operator M_n defined in (12).

Then, we can write (20) as

$$T_n = \mathbb{U}(\theta_0)'\mathbb{M}(\theta_0)\mathbb{U}(\theta_0) + o_p(1), \tag{21}$$

which is the familiar sum of squares of the projection of the integrated errors ($\mathbb{U}(\theta_0)$) onto the orthogonal space generated by the integrated derivatives ($\dot{\mathbb{U}}(\theta_0)$). Compared to NLS, we can understand that the asymptotic distribution of T_n is not distribution free since the variables involved are not stationary but integrated.

Equation (21) shows that T_n can be seen as an overidentifying restriction test, analogous to Hansen's J test or to the Sargan test, see Sections 3.3 and 5.1 in Hall (2005). Note that the n population conditions $E\mathbb{U}(\theta_0) = 0$ can be divided between k locally identifying restrictions, $E\mathbb{P}(\theta_0)\mathbb{U}(\theta_0) = 0$, and $n-k$ locally overidentifying restrictions, $E\mathbb{M}(\theta_0)\mathbb{U}(\theta_0) = 0$. Then, equation (21) just shows that T_n is testing the $n-k$ locally overidentifying restrictions while the test is blind to violations of the other k restrictions. Compared to Hansen or Sargan tests, we can understand that our test cannot have an asymptotic standard distribution since the number of overidentifying restrictions is not fixed.

The interpretation of our test as an overidentification goodness of fit test for conditional moment models recovers the relation between parameter estimation and model checking that naturally links both stages of inference. Hence, these two procedures jointly make up a methodology for performing inference in models defined by conditional moment restrictions (see Domínguez and Lobato, 2010).

Finally, the fact that we can express T_n in terms of projections, as in expression (21), is what allows us to propose the simple bootstrap procedure to estimate the critical value in the next section.

3. THE BOOTSTRAP TEST

Since Φ depends on the DGP, the asymptotic distributions of both $U_n(\mathfrak{u}, \hat{\theta})$ and T_n also do. To overcome this nonpivotality we propose to use a simple bootstrap to estimate the asymptotic distribution T_n. Alternative approaches to the bootstrap are: first, the use of critical values based on bounds, such as in Bierens (1982) or Bierens and Ploberger (1997), and, second, the use of transformed test statistics with asymptotic pivotal distributions, such as in Koul and Stute (1999). The first approach does not control the type I error, whereas the second can only be performed in restrictive frameworks, and it is case specific. Moreover, an application of the second approach would also change the test statistics and lose the

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

interpretation in terms of projections. Instead of these approaches, we propose to estimate critical values using the bootstrap for its simplicity and wide applicability. Contrary to the current applications of the bootstrap in specification testing, which are computationally complicated (see Delgado et al., 2006), we propose a bootstrap test that is computationally very simple to implement because it just involves running linear regressions (see Remark 2). Furthermore, the bootstrap test is valid under heteroskedasticity of any form and it is not a case specific procedure.

From expression (21), we propose to estimate consistently the distribution of T_n with the distribution of T_n^* where

$$T_n^* = \mathbb{U}^*(\hat{\theta})'\mathbb{M}(\hat{\theta})\mathbb{U}^*(\hat{\theta}), \tag{22}$$

where the t-th element of the n-dimensional vector $\mathbb{U}^*(\hat{\theta})$ is $U_n^*(x_t, \hat{\theta})$ and

$$U_n^*(\mathtt{u}, \hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} u(y_t, \hat{\theta}) I(x_t \leq \mathtt{u}) w_t, \tag{23}$$

where $\{w_t\}$ is a sequence of independent random variables with zero mean, unit variance, and bounded support. This procedure has been called a wild or external bootstrap (see Davidson and MacKinnon, 2006 for applications in econometrics). The proposed bootstrap procedure is consistent because

$$\sqrt{n}U_n^*(\mathtt{u}, \hat{\theta}) \Rightarrow_* B_\Gamma(\mathtt{u}) \ \ a.s.,$$

where $\Rightarrow_* a.s.$ denotes weak convergence almost surely under the bootstrap law, that is,

$$P(\sqrt{n}U_n^*(x, \hat{\theta}) \leq \mathtt{u} \mid \mathcal{X}_n) \to_{a.s.} P(B_\Gamma(x) \leq \mathtt{u}) \text{ as } n \to \infty$$

plus tightness a.s. Then, since any projection is a linear continuous operator, projections of $\sqrt{n}U_n(\mathtt{u}, \theta_0)$ and $\sqrt{n}U_n^*(\mathtt{u}, \hat{\theta})$ also enjoy the same limit distribution. Hence, consistency of bootstrap distribution trivially follows. Formally, we introduce Assumption D before establishing the next proposition.

Assumption D:

(a) Assumption B.

(b) Assumption A, replacing θ_0 by θ_1.

PROPOSITION 8. *Under either H_0 and Assumption A, or $H_{A,n}$ and Assumption C, or under H_A and Assumption D,*

$$T_n^* \Rightarrow_* \int (B_\Phi(\mathtt{u}))^2 P_{x_t}(d\mathtt{u}) \ a.s., \tag{24}$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

where under H_1, *the* θ_0 *must be replaced by* θ_1 *in the definition of* Φ, Γ_2, *and* Γ.

Remark 2. Note that the standard bootstrap approach, based on constructing a bootstrap sample (y_t^*, x_t) from resampling the residuals, cannot be followed. The reason is that y_t^* would be defined as the implicit solution of the equation $w_t u(y_t^*, \hat{\theta}) = 0$. However, since u is nonlinear this solution may not exist or may not be unique.

Remark 3. The wild bootstrap proposed in (22) is original in specification testing in this general context. Different authors have proposed wild bootstrap procedures in more restrictive contexts (see, for instance, Domínguez, 2004 or Delgado et al., 2006). In these references, the bootstrap is asymptotically equivalent to resampling a process based on the decomposition

$$U_n(\mathrm{u}, \tilde{\theta}) = U_n(\mathrm{u}, \theta_0) + R_n(\mathrm{u}, \tilde{\theta}),$$

where the remainder term R_n accounts for the estimation effect of θ_0. This remainder term causes the bootstrap procedure to be computationally involved. In contrast, in our case both the estimator and the test statistic are defined in terms of the same process $U_n(\mathrm{u}, \theta_0)$, which is just the marked empirical process that one would consider in case the parameters were known. Consequently, the effects of the errors in T_n are fully summarized in $U_n(\mathrm{u}, \theta_0)$, and hence, only this simple marked empirical process has to be resampled. As a result, in order to bootstrap T_n, the wild bootstrap only involves $U_n^*(\mathrm{u}, \hat{\theta})$, which implies a great computational advantage. Note that in the linear regression case, our bootstrap test, as well as those in Domínguez (2004) and Delgado et al. (2006), coincides with that in Stute, González-Manteiga, and Presedo-Quindimil (1998).

Remark 4. Since under $H_{A,n}$ the bootstrap still fits the null distribution, the bootstrap test preserves the admissibility property of the original asymptotic test, as in Domínguez (2004).

Proposition 8 justifies the estimation of the asymptotic critical values of T_n by those of T_n^*. In practice, the critical values of T_n^* are approximated by the simulation procedure described next. In detail, one would test the null hypothesis (10) as follows:

Step 1: Minimize the objective function (6) and calculate the test statistic $T_n = n Q_n(\hat{\theta})$.

Step 2: Obtain the integrated residuals, $U_n(x_t, \hat{\theta})$ from equation (5), and the integrated derivatives, $\dot{U}_n(x_t, \hat{\theta})$ from expression (7) and construct the $n \times k$ matrix $\dot{\mathbb{U}}(\hat{\theta})$.

Step 3: Generate $\{w_t\}$ a sequence of n bounded independent random variables with zero mean and unit variance. This sequence is serially independent and is also independent of the original sample. Then, compute the bootstrap integrated

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

errors $n \times 1$ vector, $\mathbb{U}^*(\hat{\theta})$, where the t-th coordinate of this vector is $U_n^*(x_t, \hat{\theta})$ given by (23).

Step 4: Run a linear regression of $\mathbb{U}^*(\hat{\theta})$ onto $\dot{\mathbb{U}}(\hat{\theta})$. Call T_n^* to the residual sum of squares of this regression.

Step 5: Repeat steps 3 and 4 B times, where in step 3 each sequence $\{W_t\}$ is independent of each other. This produces a set of B independent (conditionally in the sample) values of T_n^* that share the asymptotic distribution of T_n.

Step 6: Let $T^*_{[1-\alpha]}$ be the $(1-\alpha)$-quantile of the empirical distribution of the B values of T_n^*. The proposed test of nominal level α rejects the null hypothesis if $T_n > T^*_{[1-\alpha]}$.

Remark 5. Notice that, although this bootstrap procedure targets the estimation of the critical values from the asymptotic null distribution of T_n, each time we carry out step 4 we obtain a realization from the distribution of $\hat{\theta}$. Hence, as a side product we obtain B realizations from the distribution of $\hat{\theta}$ which could be employed to construct a confidence interval for θ.

4. SIMULATIONS

This section illustrates the irrelevance of plugging in efficient or inefficient estimators into omnibus test statistics in terms of finite sample power with two simple models: the linear and the Box–Cox transformation model.

First, we consider the simple linear case that establishes that

$$y_t = \alpha_0 + \beta_0 x_t + u_t,$$

with $E(u_t \mid x_t) = 0$. For a comparison between the performance of the CMM estimator and the efficient GMM (EGMM) estimator in the same set up we refer to Domínguez and Lobato (2010). Here, we just note that the CMM estimator is more robust than the EGMM estimator in the sense that it performs better (lower MSE) in nonstandard contexts like under heteroskedasticity. We set $\alpha_0 = 1$ and $\beta_0 = 2$, the x_t are i.i.d. draws from a $N(0, 5)$, and we consider three cases for u_t : an i.i.d. sequence of N(0,1) variates, an i.i.d. sequence of standardized χ^2_1, and a heteroskedastic case where $u_t \mid x_t$ follows $\exp(0.25x_t)$ times the standardized χ^2_1. The sample sizes are 50, 100, and 200. The number of bootstrap replications $B = 99$, and we conducted 2,000 experiments. This design has been partially motivated by Shin (2008). In this section $\{\omega_t\}$ is an i.i.d sequence of Bernoulli variates W where $P(W = 0.5(1-\sqrt{5})) = (1+\sqrt{5})/2\sqrt{5}$ and $P(W = 0.5(1+\sqrt{5})) = 1-(1+\sqrt{5})/2\sqrt{5}$. As noted in Stute et al. (1998), the second and third moment of W are equal to 1, and hence, this selection of $\{\omega_t\}$ guarantees that the finite dimensional distributions of the original process U_n and the bootstrap process U_n^* share the same second and third moments.

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

In this section we report the results for the proposed bootstrap test explained in Section 3 and for the test based on the test statistic that plugs in the EGMM estimator

$$T_n^{(2)} = n\, Q_n(\widehat{\theta}_{EGMM}),$$

where θ represents the vector of parameters. This test is motivated by Stute (1997) and extended by Delgado et al. (2006) who also proposed bootstrap estimators for the critical values. We follow Delgado et al. (2006) to implement the wild bootstrap for $T_n^{(2)}$ and employ the same selection for $\{\omega_t\}$.

Table 1 reports the empirical rejection percentages under the null of the proposed specification test T_n and of the $T_n^{(2)}$ test for testing the null hypothesis that the linear model is correct for three nominal levels: 0.10, 0.05, and 0.01. Similar to the estimation message, Table 1 indicates that for specification testing the CMM methodology is more robust, since it controls well the type I error for any sample size under both homoskedasticity and heteroskedasticity. On the contrary, the $T_n^{(2)}$ test presents substantial size-distortions for moderate sample sizes in the heteroskedastic case. These distortions eventually disappear as Table 1 shows, where we have added the results for two additional sample sizes, 500 and 1,000, for the heteroskedastic case.

TABLE 1. Size results (percentage). Linear model

	$T_n^{(2)}$			T_n		
n	10	5	1	10	5	1
Homoskedastic normal noise						
50	11.3	5.56	1.20	11.0	5.78	1.14
100	10.5	5.62	1.40	11.1	5.42	1.22
200	10.4	5.23	1.07	9.93	5.13	1.07
Homoskedastic chi-square noise						
50	11.6	6.18	1.28	11.5	6.16	1.46
100	10.2	4.82	.88	9.92	4.86	.96
200	10.6	5.30	.83	10.3	5.23	1.07
Heteroskedastic noise						
50	15.8	9.34	2.54	11.4	5.86	1.40
100	13.6	8.14	2.36	9.56	4.74	1.04
200	13.6	8.12	2.46	10.1	5.18	.92
500	12.8	7.43	2.63	10.4	5.23	1.13
1000	11.1	6.13	1.80	10.1	5.33	1.00

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

Next, we consider a power comparison for the linear model. We consider two alternative nonlinear regression models. First, a quadratic:

$$E(y_t \mid x_t) = \alpha_0 + \beta_0 x_t + \gamma_0 x_t^2 \tag{25}$$

with $\gamma_0 = 0.05$. Second, a discontinuous regression function:

$$E(y_t \mid x_t) = \begin{cases} \alpha_0 + \gamma_0 + \beta_0 x_t & \text{if } -0.2 \le x_t \le 0.2, \\ \alpha_0 + \beta_0 x_t & \text{otherwise,} \end{cases} \tag{26}$$

with $\gamma_0 = 3.5$.

Table 2 reports the percentage power for both tests for a 5% nominal level. For the quadratic alternative with homoskedastic noise, Table 2 shows that the $T_n^{(2)}$ test presents some power advantage over the proposed T_n test. However, for the heteroskedastic quadratic case, and for the discontinuous regression case, the proposed T_n test clearly dominates the $T_n^{(2)}$ test.

The second model we consider is motivated by the Box–Cox transformation given by

$$y_t^{(\lambda_0)} = \begin{cases} \frac{y_t^{\lambda_0} - 1}{\lambda_0} & \text{if } \lambda_0 \neq 0 \\ \log y_t & \text{if } \lambda_0 = 0, \end{cases}$$

so that the linear regression model for the transformed dependent variable establishes that

$$E\left(y_t^{(\lambda_0)} \mid x_t\right) = \alpha_0 + \beta_0 x_t. \tag{27}$$

Shin (2008) has also compared the finite sample behavior of the CMM estimator versus alternative estimators for a transformation related to the Box–Cox.

TABLE 2. Power results (percentage). Linear model. Nominal level is 5%

	$T_n^{(2)}$			T_n		
n	50	100	200	50	100	200
	Quadratic alternative					
Homoskedastic normal noise	31.4	60.6	88.3	25.6	52.3	84.2
Homoskedastic chi-square noise	43.6	63.7	89.3	38.6	57.8	83.5
Heteroskedastic noise	21.4	34.6	63.8	24.7	42.9	74.0
	Break alternative					
Homoskedastic normal noise	30.6	55.5	84.9	36.7	64.1	93.9
Homoskedastic chi-square noise	28.6	55.2	86.1	34.1	66.3	95.0
Heteroskedastic noise	30.2	42.1	63.8	35.7	58.6	84.7

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

We generate the data for the $\lambda_0 = 0$ case, which allow the implementation of the unfeasible efficient GMM (UEGMM) estimator under homoskedasticity. The UEGMM estimator employs the unfeasible instruments

$$\begin{pmatrix} -1 \\ -x_t \\ \frac{1}{2}\left[(\alpha_0+\beta_0 x_t)^2+\sigma_0^2\right] \end{pmatrix},$$

where σ_0^2 denotes the variance of $u_t \equiv y_t^{(0)} - E\left(y_t^{(0)} \mid x_t\right)$.

Since $\lambda_0 = 0$, we can write the DGP as

$$y_t = \exp\{\alpha_0 + \beta_0 x_t + u_t\}.$$

We set similar specifications for the parameters and the distributions as in the linear case.

Domínguez and Lobato (2010) compare the UEGMM and the CMM estimators; here we just focus on the comparison between the specification tests, T_n and $T_n^{(2)}$, where now $T_n^{(2)} = n Q_n(\widehat{\theta}_{UEGMM})$. Table 3 presents the empirical rejection percentages under the null of the proposed specification test T_n and of the $T_n^{(2)}$

TABLE 3. Size results (percentage). Box–Cox model

	$T_n^{(2)}$			T_n		
n	10	5	1	10	5	1
Homoskedastic normal noise						
50	11.94	5.88	1.00	10.18	4.84	0.90
100	10.66	5.54	0.92	9.84	4.62	0.90
200	10.70	5.38	1.20	9.74	4.96	1.02
Homoskedastic chi-square noise						
50	8.58	3.58	0.36	6.98	2.88	0.32
100	9.76	4.34	0.76	8.80	3.64	0.64
200	10.02	4.66	0.74	9.06	4.12	0.78
Heteroskedastic noise						
50	12.85	5.45	0.80	11.85	4.85	0.75
100	13.30	6.70	1.28	11.06	5.32	1.04
200	13.62	7.30	1.58	11.92	5.70	1.02
600	14.37	7.93	1.97	10.40	5.57	1.33
1500	12.73	6.87	2.10	10.01	4.93	0.70
3000	11.10	5.73	1.80	10.60	5.87	1.03

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

test for testing the null hypothesis that there exists some α_0, β_0, and λ_0 such that model (27) is correct, that is,

$$H_0 : E\left(y_t^{(\lambda_0)} \mid x_t\right) = \alpha_0 + \beta_0 x_t.$$

Table 3 shows that for the homoskedastic cases both tests control properly the type I error, although the proposed test T_n is somewhat conservative for the standardized χ_1^2 case and small sample sizes. However, for the heteroskedastic case, the consistent test based on the UEGMM estimator presents size distortions that seems to increase as the sample size increases. Since the UEGMM estimator does not show any bias, this size distortion has to be a finite sample issue. In order to confirm it, in Table 3 we consider three additional sample sizes, $n = 600$, $1,500$, and $3,000$ for the heteroskedastic case. This table shows that for this case, the $T_n^{(2)}$ test needs rather large sample sizes to control the type I error.

The final Monte Carlo results in Table 4 refer to a power comparison between both tests for two cases motivated by (25) and (26). Specifically, we consider that

$$y_t = \exp\{g(x_t) + u_t\}.$$

In the first case g is quadratic,

$$g(x_t) = \alpha_0 + \beta_0 x_t + c x_t^2,$$

with $c = 0.2$. In the second case g is discontinuous,

$$g(x_t) = \begin{cases} \alpha_0 + \gamma_0 + \beta_0 x_t & \text{if } -0.2 \le x_t \le 0.2 \\ \alpha_0 + \beta_0 x_t & \text{otherwise} \end{cases}$$

with $\gamma_0 = 2$. The u_t follows the same three sequences as above. We just report the empirical rejection percentages for the 0.05 nominal tests. Table 4 shows that for

TABLE 4. Power results (percentage). Box–Cox model. Nominal level is 5%

	$T_n^{(2)}$			T_n		
n	50	100	200	50	100	200
	Quadratic alternative					
Homoskedastic normal noise	44.3	78.6	97.6	22.2	52.7	81.7
Homoskedastic chi-square noise	56.2	83.0	98.5	29.6	59.8	88.7
Heteroskedastic noise	67.6	91.5	99.9	33.2	64.9	94.8
	Break alternative					
Homoskedastic normal noise	19.1	46.9	79.6	19.7	51.7	86.6
Homoskedastic chi-square noise	18.2	40.1	77.6	19.1	43.0	84.9
Heteroskedastic noise	16.6	24.0	42.3	21.6	38.0	63.3

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

the quadratic alternative $T_n^{(2)}$ dominates, whereas for the discontinuous alternative T_n presents higher empirical power. These results are especially interesting given that Table 3 showed that the $T_n^{(2)}$ presents in all cases higher rejection probabilities under the null.

To summarize this section, Tables 1–4 have illustrated that there is no advantage of plugging in an efficient estimator into the considered omnibus test statistic. Under the null, our proposed test controls better the type I error, and under the alternative, which test dominates depends on the alternative.

5. CONCLUSIONS

This article has introduced an ICM type test for correct specification of nonlinear models defined by conditional moment restrictions. The key difference with standing ICM tests is that we propose to plug in the inefficient CMM estimator into the test statistic. The main advantage is the simplicity of estimating the critical values by just running linear regressions. In addition, the test keeps the admissibility property of ICM tests, and it can be interpreted as an overidentification restriction test. Simulations confirm that in finite samples plugging in the efficient estimator does not guarantee a more powerful ICM test.

NOTE

1. We use capital letters to denote integrated functions.

REFERENCES

Bierens, H. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment test. *Econometrica* 65, 1129–1151.

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Davidson, R. & J.G. MacKinnon (2006) Bootstrap methods in econometrics. In K. Patterson & T.C. Mills (eds.), *Econometric Theory*. Palgrave Handbook of Econometrics, vol. 1. MacMillan.

Delgado, M., M. Domínguez, & P. Lavergne (2006) Consistent specification testing of conditional moment restrictions. *Annales D'Economie et de Statistique* 81, 33–67.

Domínguez, M.A. (2004) On the power of bootstrapped specification tests. *Econometric Reviews* 23, 215–228.

Domínguez, M.A. & I.N. Lobato (2004) Consistent estimation of models defined by conditional moment restrictions. *Econometrica* 72, 1601–1615.

Domínguez, M.A. & I.N. Lobato (2010) Consistent inference in models defined by conditional moment restrictions: An alternative to GMM. Working paper, ITAM, Available at http://ftp.itam.mx /pub/ academico/inves/ lobato/10-05.pdf.

Escanciano, J.C. (2006) A consistent diagnostic test for regression models using projections. *Econometric Theory* 22(06), 1030–1051.

Escanciano, J.C. (2009) On the lack of power of omnibus specification tests. *Econometric Theory* 25, 162–194.

Hall, A. (2005) *Generalized Method of Moments*. Oxford University Press.

Hansen, L.P. (1982) Large-sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press

Koul, H.L. & W. Stute (1999) Nonparametric model checks for time series. *Annals of Statistics* 27, 204–236.
Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393–415.
Shin, Y. (2008) Semiparametric estimation of the Box–Cox transformation model. *Econometrics Journal* 11, 517–537.
Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25, 613–641.
Stute, W., W.G. González-Manteiga, & M. Presedo-Quindimil (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* 83, 141–149.
Wooldridge, J.M. (1990) A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.

APPENDIX: Linearity of the CMM Estimator

Recall that

$$\sqrt{n}(\widehat{\theta}-\theta_0) = -\left\{\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\dot{U}_n(x_\ell,\theta_0)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)U_n(x_\ell,\theta_0) + o_p(1).$$

Note that the last factor can be written as

$$\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)U_n(x_\ell,\theta_0) = \frac{1}{n\sqrt{n}}\sum_{\ell=1}^{n}\sum_{t=1}^{n}\dot{U}_n'(x_\ell,\theta_0)u(x_t,\theta_0)1(x_t\le x_\ell) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u(x_t,\theta_0)H_n(x_t),$$

where $H_n(x,\theta_0)=\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)1(x\le x_\ell)$. Use the continuous mapping theorem to get that $H_n(x,\theta_0)\to_{a.s.} H(x,\theta_0)=\int\dot{U}'(\mathtt{u},\theta_0)1(x\le\mathtt{u})dP_{x_t}(x)$. Hence

$$\sqrt{n}(\widehat{\theta}-\theta_0) = -\left\{\int_{\mathbb{R}^d}\dot{U}'(\mathtt{u},\theta_0)\dot{U}(\mathtt{u},\theta_0)P_{x_t}(d\mathtt{u})\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H(x_t,\theta_0)u(x_t,\theta_0) + o_p(1).$$

Finally, to get the result define

$$l_t(z_t,\theta_0) = -\left\{\int_{\mathbb{R}^d}\dot{U}'(\mathtt{u},\theta_0)\dot{U}(\mathtt{u},\theta_0)P_{x_t}(d\mathtt{u})\right\}^{-1}H(x_t,\theta_0)u(x_t,\theta_0).$$

https://doi.org/10.1017/S0266466614000644 Published online by Cambridge University Press