# A CONSISTENT SPECIFICATION TEST FOR MODELS DEFINED BY CONDITIONAL MOMENT RESTRICTIONS [1]

Manuel A. Domínguez and Ignacio N. Lobato

Instituto Tecnológico Autónomo de México (ITAM)
Av. Camino a Santa Teresa #930
Col. Héroes de Padierna
10700 México, D.F., MEXICO

**Abstract**

This article proposes to use the minimized value of the objective function of the minimum distance estimator proposed in Domínguez and Lobato (2004) as a consistent specification test for models defined by conditional moment restrictions. Inference in these models is typically performed using the generalized method of moments (GMM) methodology, which has been criticized because its inconsistency. Our test is more general than the currently available consistent tests and presents three additional advantages. First, because the employed estimator is robust, that is, it is squared-root-n consistent under the null under general conditions, the resulting test is also robust in the sense of controlling the type I error in general settings. Second, by using this test, we recover the spirit of GMM, namely, the use of a single tool for estimation and model checking. Third, from a practical point of view, the test is very simple because the bootstrap, which is required for estimating the asymptotic null distribution, does not demand re-estimating the parameters in each bootstrap sample, but only resample a simple marked empirical process.

Keywords and Phrases: Generalized Method of Moments, Identification, Unconditional Moments, Marked Empirical Process, Integrated Regression Function.

[1] We thank M. Carrasco, M.A. Delgado, J.C. Escanciano and A. Marcet for useful conversations. We also thank seminar participants at Universidad de Alicante, Universidad de Bilbao, Universidad de Navarra, Universidad Carlos III de Madrid, Universidad Autónoma de Barcelona, Universidad Pompeu Fabra, CEMFI and CIREQ for useful comments. This research was partially financed by the Spanish Ministerio de Educación y Ciencia, Ref. no. SEJ2004-04583/ECON. Part of this research was carried out while Lobato was visiting Universidad Carlos III de Madrid thanks to the Spanish Secretaría de Estado de Universidades e Investigación, Ref. no. SAB2004-0034. Domínguez acknowledges financial support from Asociación Mexicana de Cultura. Lobato acknowledges financial support from Asociación Mexicana de Cultura.

# 1 Introduction

Most successful methodologies for performing statistical inference in econometric models are based on the use of a compatibility index between the model and the data. Formally, this compatibility index is expressed in terms of an Objective Function (OF, hereinafter) that takes an ideal value when there is full agreement between the model assumptions and the data. Once the OF is defined, all inferential procedures are related to it. Parameter estimators are the parameter values that make the OF closest to the ideal value. Tests for correct specification are based on the difference between the ideal value of the OF and the value it takes on the data. Tests for parameter restrictions are based on the change in the OF derived from the imposition of these restrictions.

The paradigm of such methodology is the Generalized Method of Moments (GMM, henceforth) approach employed in models defined by Conditional Moment Restrictions (CMR, hereinafter). In these procedures, the OF measures the distance between some empirical and theoretical unconditional moments, which ideally should be zero. Therefore, the GMM is a minimum distance methodology. Unfortunately, it yields inconsistent tests and estimators since it does not use the full information contained in the definition of the model. This information involves an infinite number of unconditional moments and so, consistent tests and estimators require the use of such amount of restrictions in the distance measure. For specification testing, this problem was early noticed, see Newey (1985) or Tauchen (1985). For estimation, Domínguez and Lobato (2004, DL hereinafter) recently have proposed the Consistent Method of Conditional Moments (CMCM, henceforth).

This article presents two purposes. First, we point out that the identification problem that leads to the inconsistency of GMM estimators is more frequent that the examples in DL suggest. In particular, in the next section we present the identification problem in a basic asset allocation model. Second, once the importance of the CMCM estimator has been stressed, we propose the simplest and, up to date, most general specification test for CMR models. The test statistic is just the minimized value of DL's OF, which is based on an infinite number of unconditional moment restrictions that fully imposes the CMR. This article complements DL and establishes a new inferential unified methodology for CMR models. In this way, we recover the unified approach to inference, and relate in a natural way both parts of inference, estimation and diagnostic testing.[2]

[2]"All the usual optimization estimators share the feature that the value of the expected criterion function at the minimum is an indicator of goodness of fit" (Davidson, 2000, p.221).

Furthermore, the resulting specification test presents three additional advantages over the existing ones. The first one, as previously mentioned, is the use of an unified methodology for inference in CMR models, where the specificaction test is just a by-product of the estimation procedure. The second advantage is the robustness, namely, under the null hypothesis the proposed test properly controls the type I error under fairly general conditions. Note that all specification tests regard the model parameters as nuisance, and need to replace them by consistent estimators. However, the existing tests are very careful imposing the full model definition only at the model checking stage, and not at the estimation stage. As a result, the estimators are consistent only under additional assumptions on the distribution of the exogenous variables. If these assumptions do not hold, the tests will not control the type I error. The third, and main practical, advantage is the simplicity of the bootstrap procedure that estimates the critical values. In addition, the proposed test has power comparable to the alternative consistent tests.

The plan of this note is the following. Section 2 introduces a simple motivating example to stress the importance of the identification problem. In Section 3 we present the testing framework and introduce our test statistic. Section 4 states the asymptotic properties of the test. Since the asymptotic distribution of the proposed test is case dependent, it cannot be automatically implemented. In Section 5 we propose a feasible implementation of the test that employs critical values obtained by a simple bootstrap procedure.

# 2 Example

Consider the following simplification of the basic capital asset pricing model

$$E(X_t^{-\alpha_0} \mid \mathcal{F}_{t-1}) = K, \tag{1}$$

where $X_t$ is a positive random variable, namely the ratio between current and past consumption, which in these models is the representative agent's marginal rate of substitution between current and past consumptions. In addition, $\mathcal{F}_{t-1}$ represents the relevant information for the agent at period $t-1$, and $\alpha_0$ and $K$ are parameters, namely the coefficient of relative risk aversion and the inverse of the agent's intertemporal discount rate. This simplified model is enough for illustrating the main points, later we will comment on extensions of this model. Assume that $K$ is known and the main interest is performing statistical inference about $\alpha_0$ and about the suitability of the model.

Consider that $X_t$ takes only three positive values $\{x_1, x_2, x_3\}$ and evolves according to a Markov Chain, with $\{p_{ij},\ i, j = 1, 2, 3\}$ denoting the transition probability of moving from state $i$ to state $j$. Note that the ergodic stationary distribution is the normalized eigenvector associated to the unit eigenvalue, call it $(p_1, p_2, p_3)$. For concreteness, also assume that $\{x_1, x_2, x_3\} = \{0.8, 1, 1.2\}$, so that consumption can decrease, remain the same or increase.

The traditional approach to make inference is the GMM approach, which is based on the selection of one (or more) unconditional moments, such as

$$E(X_t^{-\alpha}) = K. \tag{2}$$

The key insight is that GMM assumes that the unconditional moment (2) identifies the true $\alpha_0$. That is, $\alpha_0$ is the unique $\alpha$ that satisfies (2). However, condition (2) just imposes that

$$p_1 0.8^{-\alpha} + p_2 + p_3 1.2^{-\alpha} = K. \tag{3}$$

Because $0.8^{-\alpha}$ is an increasing function, whereas $1.2^{-\alpha}$ is decreasing, the function $p_1 0.8^{-\alpha} + p_2 + p_3 1.2^{-\alpha}$ has an asymmetric $U-$shape. In Figure 1 we have plotted this function for the case $(p_1, p_2, p_3) = (1/3, 1/3, 1/3)$. It shows that equation (3) has no solutions when $K < 1$, has a unique minimum at $K = 1$ and, for each $K > 1$, there are two values of $\alpha$ that verify condition (2). Therefore, for this unconditional distribution, when the discount factor is less than one (the usual case), GMM based on condition (2) delivers inconsistent estimators because equation (2) has two solutions for $\alpha$. Furthermore, for alternative unconditional distributions the same problem occurs because the effect of changing the values for $(p_1, p_2, p_3)$ is to shift the location of the minimum, but $E(X_t^{-\alpha})$ still has an asymmetric $U-$shape, so that the identification problem remains. In addition, the same phenomenon happens even when $X$ is continuous. By writing the unconditional expectation as

$$E(X_t^{-\alpha}) = E(X_t^{-\alpha} \mid X_t > 1)P(X_t > 1) + E(X_t^{-\alpha} \mid X_t < 1)P(X_t < 1),$$

and noticing that $E(X_t^{-\alpha} \mid X_t > 1)$ is decreasing in $\alpha$, whereas $E(X_t^{-\alpha} \mid X_t < 1)$ is increasing in $\alpha$, the function $E(X_t^{-\alpha})$ has also an asymmetric $U-$shape.

Hence, for this simplest case where the selected instrument is the constant, GMM provides inconsistent inference. One could think that the identification problem could be solved by selecting the optimal instrument or by employing some additional instruments. However, as DL showed, these approaches would still fail in general. In particular, for our example, if the optimal instrument is employed, the three conditional restrictions and the unconditional restriction define a system of equations with 9 unknowns and only 4 restrictions[3], apart from the positiveness of $p_{ij}$. So, there are plenty of solutions. For example, when $p_{11}$ =.291, $p_{12}$ = .266, $p_{21}$ = .282, $p_{22}$ = .294, $p_{31}$ = .282, $p_{32}$ = .294, $K$ =1.038, and the true $\alpha$ is 2.19, there are two additional values for $\alpha$ that satisfy the unconditional restriction imposed by the optimal instrument. In Figure 2 we have plotted the unconditional restriction that satisfies the GMM optimal instrument with the three solutions.

The example given in equation (1) may appear to be too simple, but many financial models have a similar structure. For instance, the basic asset allocation model presents the following first order condition

$$\beta_0 E\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0} R_{t+1} \mid \digamma_{t-1}\right] = 1 \tag{4}$$

where $\beta_0$ is the discount factor, and $R$ is the gross return to bonds. Note that (4) is basically model (1) with $K = 1/\beta_0$, and where we have constrained $R$ to be constant. Note that the Markovian structure imposed in the example is just for simplicity and makes the problem of lack of identification less likely since the conditional distribution determines the unconditional one, and so it decreases the number of unknowns in the system of equations commented above.

It is important to stress that there is nothing special about the previous example. The main message from this section is that the identification problem just outlined is a general problem that occurs when the parameters are defined through conditional restrictions but they are estimated using a finite number of unconditional restrictions. If $E(h(Y,\theta_0) \mid X) = 0$ then $E(h(Y,\theta_0)W(X)) = 0$ for 'any' $W$, but there are many other parameters that may also verify this unconditional relation. Consider a different parameter value $\theta_1$ and let $E(h(Y,\theta_1)W(X) \mid X) = D(X)$. The important point to emphasize is that as soon as the function $D(X)$ takes positive and negative values, the identification problem arises. Of

[3] The four equations are: $0.8^{\alpha}p_{11} + p_{12} + 1.2^{\alpha}(1 - p_{11} - p_{12}) - K = 0$,
$0.8^{\alpha}p_{21} + p_{22} + 1.2^{\alpha}(1 - p_{21} - p_{22}) - K = 0$,
$0.8^{\alpha}p_{31} + p_{32} + 1.2^{\alpha}(1 - p_{31} - p_{32}) - K = 0$, and

$$\begin{aligned}
&(0.8^{\beta}p_{11} + p_{12} + 1.2^{\beta}(1 - p_{11} - p_{12}) - K)(0.8^{\beta}\log(0.8)p_{11} + 1.2^{\beta}\log(1.2)p_{13})p_1 \\
&+(0.8^{\beta}p_{21} + p_{22} + 1.2^{\beta}(1 - p_{21} - p_{22}) - K)(0.8^{\beta}\log(0.8)p_{21} + 1.2^{\beta}\log(1.2)p_{23})p_2 \\
&+(0.8^{\beta}p_{31} + p_{32} + 1.2^{\beta}(1 - p_{31} - p_{32}) - K)(0.8^{\beta}\log(0.8)p_{31} + 1.2^{\beta}\log(1.2)(1 - p_{31} - p_{32}))(1 - p_1 - p_2) = 0.
\end{aligned}$$

The nine unknowns are $p_{11}, p_{12,}p_{21,}p_{22}, p_{31,}p_{32}, \alpha, \beta$, and $K$, where $\beta$ denotes the alternative value for the relative risk aversion coefficient.

course, the more values of $\theta$ that leads to the function $D(X)$ taking positive and negative values, the more severe the identification problem would be in practice.

# 3 Framework

In this section we will formally introduce our consistent specification test and compare it with related tests procedures. We follow the notation in DL. Then, $Z_t$ is a time series vector and $\{Y_t, X_t\}$ are two subvectors of $Z_t$ (that could have common coordinates), where $Y_t$ is a $k$-dimensional time series vector that may contain endogenous and exogenous variables and a finite number of these lagged variables and $X_t$ is a $d-$dimensional time series vector that contains the exogenous variables (again, a finite number of these lagged variables can be included). The coordinates of $Z_t$ are related by an econometric model which establishes that the true distribution of the data satisfies the following conditional moment restrictions

$$E(h(Y_t, \theta_0) \mid X_t) = 0, \quad a.s. \tag{5}$$

for a unique value $\theta_0 \in \Theta$, where $\Theta \subset \mathbb{R}^m$. Equation (5) defines the parameter $\theta_0$ which is unknown to the econometrician. The function $h$ that maps $\mathbb{R}^k \times \Theta$ into $\mathbb{R}^l$ is supposed to be known. In general, $h(Y_t, \theta_0)$ can be understood as the errors in a multivariate nonlinear dynamic regression model; for instance, $h(Y_t, \theta_0)$ are called generalized residuals in Wooldridge (1990). In this paper, for simplicity, we will consider the case where $l = 1$. This model has been repeatedly considered in the econometrics literature and several estimators have been proposed, see references in DL.

In this article, we focus on testing whether model (5) is correctly specified. Specifically, we consider (5) as the null hypothesis ($H_0$), and the alternative hypothesis is that for any $\theta$

$$P(E(h(Y_t, \theta) \mid X_t) = 0) < 1.$$

As mentioned previously, the GMM overidentifying restriction test is not consistent for our null hypothesis because it just tests the validity of an arbitrary finite number of unconditional restrictions (from the infinite implied by the conditional expectation (5)). In order to avoid this problem, Bierens (1982, 1990), Bierens and Ploberger (1997) or Stute (1997), among others, proposed tests which employ an infinite number of unconditional moments. However, those references do not consider inference as a whole, but just focus on the model check stage. Since the parameters of the model are nuisance for model checking, they propose to replace them by consistent estimators, without discussing carefully the estimation stage. The examples in DL and the example in the previous section show that this discussion can not be overlooked. In particular, plugging in the typical GMM estimator, even the efficient one, would lead to an oversized test. For example, assume that $H_0$ holds but the true value $\theta_0$ is estimated with an inconsistent estimator $\hat{\theta}$ which converges in probability to the random variable $S$. Then, these tests check whether $E(h(Y_t, S) \mid X_t) = 0, \quad a.s.$, which may be false, although $H_0$ holds. Therefore, under $H_0$, these tests will reject asymptotically more often than the specified theoretical level. In particular, the commented examples can be worked out further to show that the GMM estimator converges to $\theta_0$ with probability $p$ and to some other parameter values $\theta \neq \theta_0$ with probability $1-p$. In this particular case, the asymptotic type I error of the tests that employ the GMM estimator is $\alpha p + (1-p)$ where $\alpha$ is the desired nominal size.

The framework we are considering is more general than the one in previous specification tests. For instance, Carrasco and Florens (2000), Koul and Ni (2004), Escanciano (2006a) and Delgado, Domínguez and Lavergne (2006) considered general specification testing in i.i.d. settings. Koul and Stute (1999) considered specification testing in a generalized regression model in a Markovian framework. Bierens and Ploberger (1997) considered specification testing for time-series regression models.

Next, we describe our test procedure. Let $P_X$ be the probability law of $X_t$ and let $I(X_t \leq x)$ denote the indicator function that equals 1 when each component in $X_t$ is less or equal than the corresponding component in $x$; and equals 0 otherwise. DL used the compatibility index

$$Q(\theta) = \int_{\mathbb{R}^d} E(h(Y_t, \theta) I(X_t \leq x))^2 dP_X(x) \tag{6}$$

which is 0 at $\theta_0$ only if the conditional moment restrictions hold. The CMCM estimator of $\theta_0$ is defined by

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} Q_n(\theta),$$

where $Q_n(\theta)$ is the sample analog of $Q(\theta)$, namely,

$$Q_n(\theta) = \frac{1}{n^3} \sum_{\ell=1}^{n} \left( \sum_{t=1}^{n} h(Y_t, \theta) I(X_t \leq X_\ell) \right)^2 .$$

DL showed that this estimator is consistent. Following the discussion in the introduction, a natural goodness of fit test procedure uses $Q_n(\theta)$ at its minimized value as a test statistic, specifically the proposed test statistic is $T_n = nQ_n(\widehat{\theta})$.

As an additional point, notice that instead of plugging in $\widehat{\theta}$, that is consistent but inefficient, one could propose to employ as test statistic $T_n^{'} = nQ_n(\widetilde{\theta})$, where $\widetilde{\theta}$ is an efficient estimator. The comparison between $T_n$ and $T_n^{'}$ should be carried out under the null and under the alternative. Under the null, $T_n^{'}$ may not control the type I error. Even when both tests control properly the type I error, under the alternative, $T_n^{'}$ does not lead to a more powerful test. The reason is clear: efficiency of $\widetilde{\theta}$ is a property derived under the null hypothesis, assuming that the specified model is correct, whereas power refers to the behavior of the statistic under the alternative hypothesis, see Neyman (1969).

In order to derive the asymptotic theory, it is useful to rewrite the statistic in terms of the rescaled integrated regression function that can be seen as a marked empirical process with marks given by $h(Y_t, \theta)$. Introducing the following marked empirical process

$$R_n(\theta, x) = n^{-1/2} \sum_{t=1}^{n} h(Y_t, \theta) I(X_t \leq x),$$

our statistic can be seen as a Cramer von Mises statistic applied to the process $R_n(\widehat{\theta}, x)$, namely

$$T_n = \frac{1}{n} \sum_{\ell=1}^{n} R_n(\widehat{\theta}, X_l)^2.$$

# 4 Asymptotic Results

Under general assumptions, DL established that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \rightarrow_d -\Sigma^{-1}_{\dot{H}\dot{H}'} \Sigma_{\dot{H}B_\Gamma}, \tag{7}$$

where $\Sigma_{\dot{H}\dot{H}'} = \int \dot{H}\dot{H}' dP_X$, $\Sigma_{\dot{H}B_\Gamma} = \int \dot{H} B_\Gamma dP_X$, with $\dot{H}(x) = E(\dot{h}(Y_t, \theta_0) I(X_t \leq x))$, $\dot{h}(Y_t, \theta) = \partial h(Y_t, \theta)/\partial\theta$, $B_\Gamma$ denotes a centered Gaussian process in $D[R]^d$ (where $D[R]^d$ is the space of real functions that are continuous from above and with limits from below), with covariance structure given by $\Gamma(r, s) = E(h^2(Y_t, \theta_0) I(X_t \leq r \wedge s))$, and where $\wedge$ denotes minimum. DL also showed that

$$-\Sigma^{-1}_{\dot{H}\dot{H}'} \Sigma_{\dot{H}B_\Gamma} \stackrel{d}{=} N(0, V)$$

for a certain $V$ that can be easily estimated. When $h$ is homoskedastic and $d = 1$, $B_\Gamma$ particularizes to a scaled Brownian motion. In addition, notice that (7) reminds similar properties satisfied by popular estimators such as nonlinear least squares or GMM estimators. The difference with them is that, in our case the involved variables (”regressors” and errors or generalized residuals) are partial sum processes instead of raw variables. Note that, if instead of integrating the marked empirical process with respect to an estimator of the distribution of the exogenous variables, the integration is done with respect to an absolutely continuous distribution, the asymptotic theory could be studied in an alternative framework that would require only finiteness of second moments, see Escanciano (2006b).

Next, we present the properties of this test, which follow from the results in DL and Bierens and Ploberger (1997). First, under the null hypothesis

$$T_n \to_d \int \left(B_\Gamma - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}B_\Gamma}\right)^2 dP_X.$$

Note that the covariance structure of the process $B_\Gamma - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}B_\Gamma}$ is given by

$$\begin{aligned}\Phi(t,s) &= \Gamma(t,s) - \dot{H}'(t)\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma'_{\dot{H}\Gamma}(s) - \Sigma_{\dot{H}'\Gamma}(t)\Sigma^{-1}_{\dot{H}\dot{H}'}\dot{H}'(s) \\ &\quad + \dot{H}'(t)\Sigma^{-1}_{\dot{H}\dot{H}'}\left[\int \dot{H}'(u)\Gamma(u,v)\dot{H}(v)P_X(du)P_X(dv)\right]\Sigma^{-1}_{\dot{H}\dot{H}'}\dot{H}(s),\end{aligned}$$

where $\Sigma_{\dot{H}'\Gamma}(s) = \int \dot{H}'(u)\Gamma(u,s)\,dP_X(u)$. Therefore, the critical values of the test statistic $T_n$ depend on the data generating process (DGP), complicating statistical inference.

Second, concerning the behavior under the alternative, it is straightforward to show that the test statistic $T_n$ diverges under any fixed alternative. Therefore, the test is consistent. Moreover, under a sequence of local alternatives, such as

$$H_{A,n} : E(h(Y_t,\theta_0) \mid X_t) = \frac{g(X_t)}{\sqrt{n}} \qquad a.s.$$

note that $\theta_0$ still minimizes (6). However, the distribution of $\widehat{\theta}$ is now

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to_d -\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}(B_\Gamma+G)},$$

where $\Sigma_{\dot{H}(B_\Gamma+G)} = \int \dot{H}(B_\Gamma + G)dP_{X_1}$, $G(x) = E(g(X)I(X \leq x))$. Then,

$$T_n \to_d \int \left(B_\Gamma + G - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}(B_\Gamma+G)}\right)^2 dP_X.$$

Hence, the asymptotic distribution of $T_n$ under $H_{A,n}$ is the same Gaussian process obtained under the null, but now centered at the function $G - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}G}$, where $\Sigma_{\dot{H}G} = \int \dot{H}GdP_X$. Note that the structure of the asymptotic distribution is essentially equivalent to the structure of the tests proposed in Bierens and Ploberger (1997) and Stute (1997). As showed there, the existence of this bias is all that is needed to show that the probability of rejecting under $H_{A,n}$ is larger than $\alpha$, and therefore, to show that the test has nontrivial power against the sequence of local alternatives $H_{A,n}$. Along the lines of those papers, it can be straightforwardly shown that the test enjoys optimality properties.

# 5 Bootstrap test

Since $\Phi$ depends on the DGP, the asymptotic distributions of both $R_n(\theta, x)$ and $T_n$ generally also do. Hence, the theory established in the previous section cannot be automatically applied for statistical inference because there are not generally valid critical values. There are three approaches to constructing feasible tests: first, to estimate the asymptotic null distribution by estimating its spectral decomposition, see Horowitz (2006) or Carrasco and Florens (2007); second, to use the bootstrap to estimate this distribution; and third, to transform the test statistic via a martingalization that yields an asymptotically distribution free statistic. The first approach is computationally involved and requires the selection of a truncation number. Comparing the second and third approaches, Koul and Sakhanenko (2005) report that in finite samples, tests based on the bootstrap control worse the type I error, although they have more empirical power.

We prefer to follow the bootstrap approach for two reasons. First, the bootstrap test is valid under heteroskedasticity of any form and it is not a case specific procedure. Second, it is unclear whether the martingalization approach would lead to abandon the unifying inference approach advocated in this article.

Next, we explain and justify the proposed bootstrap-based test procedure. Recall

$$R_n(\widehat{\theta}, x) = n^{-1/2} \sum_{t=1}^{n} h(Y_t, \widehat{\theta}) I(X_t \leq x),$$

so that,

$$R_n(\widehat{\theta}, x) = R_{1n}(\theta_0, x) + R_{2n}(\widehat{\theta}, x),$$

where

$$R_{1n}(\theta, x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(Y_t, \theta) I(X_t \leq x),\ R_{2n}(\theta, x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{h}(Y_t, \overline{\theta}) I(X_t \leq x)(\theta - \theta_0)$$

and $\overline{\theta}$ is intermediate between $\widehat{\theta}$ and $\theta_0$. $R_{1n}$ is the sample process that a test would use for model checking when the parameters are known, while $R_{2n}$ corrects $R_{1n}$ for the effect of the estimation of the model parameters. Therefore, using the asymptotic expansion of $\widehat{\theta}$, we have that

$$R_{2n}(\widehat{\theta}, x) = -\dot{H}(x)' \widehat{\Sigma}^{-1}_{\dot{H}\dot{H}'} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{H}(X_t) R_{1n}(\theta_0, X_t) + o_p(1).$$

So, we can define

$$R_n^*(\widehat{\theta},x) = R_{1n}^*(\widehat{\theta},x) + R_{2n}^*(\widehat{\theta},x)$$

where

$$R_{1n}^*(\widehat{\theta},x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(Y_t,\widehat{\theta})I(X_t \le x)W_t, \text{ and } R_{2n}^*(\widehat{\theta},x) = -\widehat{\dot{H}}(x)'\widehat{\Sigma}^{-1}_{\dot{H}\dot{H}'}\frac{1}{n}\sum_{k=1}^{n}\widehat{\dot{H}}(X_k)R_{1n}^*(\widehat{\theta},X_k),$$

where

$$\widehat{\dot{H}}(x) = \frac{1}{n}\sum_{s=1}^{n}\dot{h}(Y_s,\widehat{\theta})I(X_s \le x),$$

$$\widehat{\Sigma}_{\dot{H}\dot{H}'} = \frac{1}{n}\sum_{k=1}^{n}\widehat{\dot{H}}(X_k)\widehat{\dot{H}}'(X_k),\ \widehat{\Sigma}^*_{\dot{H}B_\Gamma} = \frac{1}{n}\sum_{k=1}^{n}\widehat{\dot{H}}(X_k)R_{1n}^*(\widehat{\theta},X_k),$$

and where $\{W_t\}$ is a sequence of independent random variables with zero mean, unit variance and bounded support. The main idea is to estimate the distribution of $\sqrt{n}R_n\left(\widehat{\theta},x\right)$ by the distribution of $\sqrt{n}R_n^*\left(\widehat{\theta},x\right)$, and hence to estimate the distribution of $T_n$ by the distribution of $T_n^*$, defined by

$$T_n^* = \frac{1}{n}\sum_{\ell=1}^{n} R_n^*(\widehat{\theta},X_l)^2. \tag{8}$$

This procedure has been called a wild or external bootstrap, see Wu (1986), Mammen (1993) and Delgado and Fiteni (2002) for applications in econometrics.

**Remark 1**. If $\Pi_{\dot{H}}$ denotes the orthogonal projection on $\dot{H}$ in $L_2(P_X)$, the Hilbert space of all real-valued and $P_X-$square integrable functions, then, as a referee has pointed out, the limit of the first order asymptotic expansion of the process $R_n(\widehat{\theta},x)$ is given by $\Pi_{\dot{H}}B_\Gamma = B_\Gamma - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\Sigma_{\dot{H}B_\Gamma}$, see Escanciano (2006a) for a similar finding. This interpretation is important for understanding the proposed bootstrap procedure. Informally speaking, the limit distribution of $\Pi_{\dot{H}}R_n(\theta_0,\cdot)$ can be approximated by the bootstrap distribution of $\Pi_{\dot{H}}R_{1n}^*(\widehat{\theta},\cdot)$, because the distribution of $R_{1n}^*(\widehat{\theta},\cdot)$ approximates the distribution of $R_n(\theta_0,\cdot)$ and the projection is a continuous functional. This argument represents further motivation for the use of the proposed estimator over other estimators for which this "projection" idea does not work.

**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 $W_t h(y,\widehat{\theta}) = 0$. However, this solution may not exist or may not be unique.

**Remark 3**. The wild bootstrap proposed in (8) is original in specification testing. Different authors have proposed wild bootstrap procedures in similar contexts, see for instance, Stute, González-Manteiga and Presedo-Quindimil (1998), Domínguez (2004) or Delgado, Domínguez and Lavergne (2006). In these references, the bootstrap procedure is asymptotically equivalent to resampling a complicated process, which can not be regarded as a marked empirical process because of the particular form of the corresponding term $R_{2n}$. On the contrary, in our case both the estimator and the test statistic are defined in terms of the same process $R_{1n}(\theta, x)$. Consequently, the effects of the errors in $T_n$ are fully summarized in $R_{1n}(\theta, x)$, 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 $R^*_{1n}(\widehat{\theta}, x)$, which is just the marked empirical process that one would consider in case the parameters were known. In practical terms, it means a great advantage from a computational point of view.

Then, under the assumptions commented in Section 3,

$$\sqrt{n}R_n^*\left(\widehat{\theta}, x\right) \Rightarrow_* B_\Gamma - \dot{H}'\,\Sigma^{-1}_{\dot{H}\dot{H}'}\,\Sigma_{\dot{H}B_\Gamma} \quad a.s.,$$

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

$$P(\sqrt{n}R_n^*\left(\widehat{\theta}, x\right) \leq s \mid \mathcal{X}_n) \rightarrow_{a.s.} P(B_\Gamma - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\,\Sigma_{\dot{H}B_\Gamma} \leq s) \text{ as } n \rightarrow \infty$$

plus tightness a.s..

Therefore, the asymptotic distribution of $\sqrt{n}R_n\left(\widehat{\theta}, x\right)$ can be estimated with that of $\sqrt{n}R_n^*\left(\widehat{\theta}, x\right)$. Similarly, the asymptotic distribution of $T_n$ can be estimated with that of $T_n^*$. In fact, a straightforward application of the Continuous Mapping yields that

$$T_n^* \Rightarrow_* \int \left(B_\Gamma - \dot{H}'\Sigma^{-1}_{\dot{H}\dot{H}'}\,\Sigma_{\dot{H}B_\Gamma}\right)^2 dP_X \; a.s.. \tag{9}$$

This result 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 simulations. Equation (9) establishes that under the null hypothesis, $T_n$ and $T_n^*$ share the same asymptotic distribution for almost all samples. Hence, under the null, the rejection probability of the bootstrap test converges to the theoretical level. In addition, using arguments similar to Domínguez (2004), it can be shown that the proposed bootstrap preserves the critical region even under local alternatives. Therefore, the optimal original properties of the test still hold.

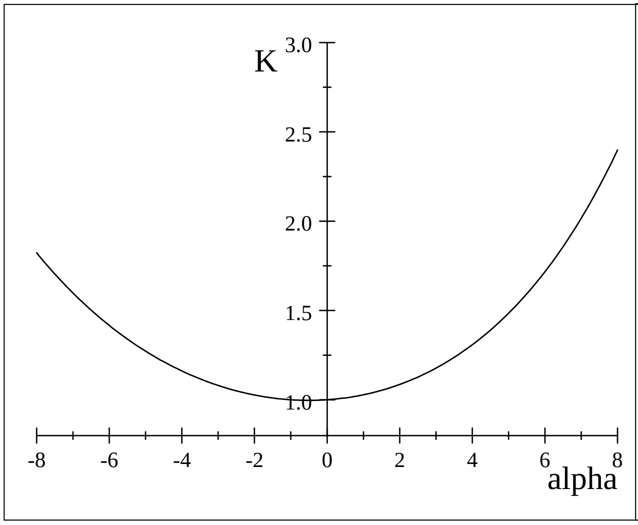


Figure 1

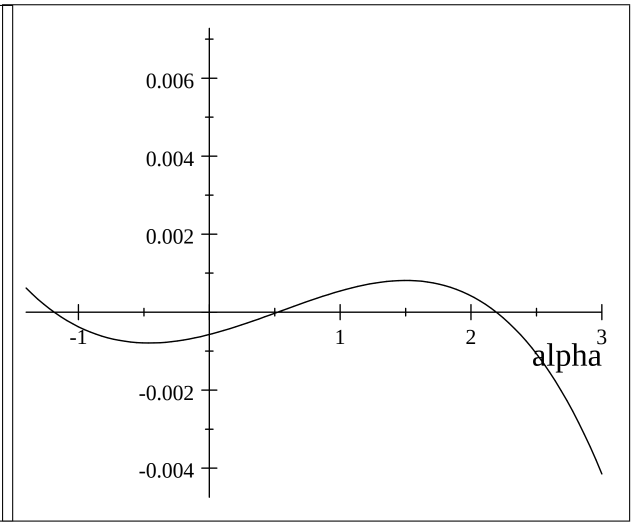


Figure 2

# References

Bierens, H. (1982), ”Consistent Model Specification Tests”, Journal of Econometrics, 20, 105-134.

_____ (1990), ”A consistent conditional moment test of functional form”, Econometrica, 58, 1443-1458.

_____ and Ploberger, W. (1997), ”Asymptotic theory of integrated conditional moment test”, Econometrica, 65, 1129-1151.

Carrasco, M. and Florens J.-P. (2000), ”Generalization of GMM to a continuum of moment conditions”, Econometric Theory, 16, 797-834.

_____ and E. Renault (2007), " Linear Inverse Problems in Structural Econometrics", Handbook of Econometrics, Vol. 6, eds. J.J. Heckman and E.E. Leamer.

Davidson, J. (2000), Econometric Theory, Oxford University Press.

Delgado M., M. Domínguez, and P. Lavergne (2006), "Consistent Specification Testing of Conditional Moment Restrictions", Annales D'Economie et de Statistique 81, 33-67.

Delgado, M. A. and Fiteni, I. (2002), ”External bootstrap tests for parameter stability”, Journal of Econometrics, 109, 275-303.

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

_____ and I.N. Lobato (2004), ”Consistent Estimation of Models Defined by Conditional Moment Restrictions”, Econometrica, 1601-1615.

Escanciano, J.C. (2006a), "A Consistent Diagnostic Test for Regression Models Using Projections", Econometric Theory, 2006, 22, 1030-1051.

_____ (2006b), "Goodness-of-fit Tests for Linear and Nonlinear Time Series Models", Journal of the American Statistical Association 101, 531 - 541.

Horowitz, J. (2006) "Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables", Econometrica, 74, 521-538.

Koul, H. L. and Ni, P. (2004), "Minimum distance regression model checking", Journal of Statistical Planning and Inference, 119, 109–144.

Koul, H. and Sakhanenko, L. (2005), "Goodness-of-fit testing in regression: A finite sample comparison of bootstrap methodology and Khamaladze transformation", Statistics and Probability Letters, 74 (3), 290-302.

Koul, H. L. and W. Stute (1999), "Nonparametric model checks for time series", Annals of Statistics, 27, 204-236.

Mammen, E. (1993), "Bootstrap and Wild Bootstrap for High Dimensional Linear Models", Annals of Statistics, 21, 255-285.

Newey, W. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", Econometrica, 53, 1047-1070.

Neyman, J. (1969), "Statistical Problems in Science. The Symmetric Test of a Composite Hypothesis", Journal of the American Statistical Association, 64, 1154-1171.

Stute, W. (1997), "Nonparametric model checks for regression", Annals of Statistics, 25, 613-641.

_____, González-Manteiga, W.G., and Presedo-Quindimil, M. (1998), "Bootstrap approximations in model checks for regression", Journal of the American Statistical Association, 83, 141-149.

Tauchen, G. (1985), "Diagnostic testing and evaluation of maximum likelihood models", Journal of Econometrics, 30, 415–443.

Wooldridge, J.M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests", Econometric Theory, 6, 17–43.

Wu, C. F. J. (1986), "Jacknife, bootstrap and other resampling methods in regression analysis (with discussion)", Annals of Statistics, 14, 1261-1350.