INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

CENTRO DE INVESTIGACIÓN ECONÓMICA

Discussion Paper Series

Consistent Inference in Models Defined by Conditional Moment Restrictions: An Alternative to GMM

Manuel Domínguez
Universidad Complutense de Madrid

and

Ignacio N. Lobato
Instituto Tecnológico Autónomo de México

November 2010
Discussion Paper 10-05

Av. Camino a Santa Teresa # 930
México, D.F. 10700
M É X I C O
Tel. +52(55) 5628 4197

CONSISTENT INFERENCE IN MODELS DEFINED BY CONDITIONAL MOMENT RESTRICTIONS: AN ALTERNATIVE TO GMM

Manuel A. Domínguez
Universidad Complutense de Madrid

Ignacio N. Lobato[1]
Instituto Tecnológico Autónomo de México

Abstract

This article introduces a unified methodology for estimating and testing nonlinear econometric models defined by conditional moment restrictions. These models are very common in econometrics, such as nonlinear rational expectation models. The current approach for inference in these models is the generalized method of moments (GMM) methodology, as proposed by Hansen and Singleton (1982). Although GMM provides a unified methodology for statistical inference that is simple to implement, it may yield inconsistent statistical procedures because it just employs a finite number of moments. This is a very important theoretical and applied problem, as illustrated by a simplified consumption-based asset pricing model. Contrary to GMM, the methodology proposed in this article delivers consistent statistical procedures because it employs an infinite number of moments that characterizes the conditional moment. In addition, the proposed methodology is widely applicable for general time series data and easy to implement. In particular, the proposed specification test relies on a novel and very simple wild bootstrap procedure.

Keywords and Phrases: Generalized Method of Moments, Identification, Unconditional Moments, Consistency, Minimum Distance, Marked Empirical Process.

JEL classification numbers: C12 and C52

[1]Corresponding author. Address: Centro de Investigación Económica, Instituto Tecnológico Autónomo de México (ITAM), Av. Camino a Santa Teresa #930, 10700 México, D.F., MEXICO, ilobato@itam.mx

1 Introduction

A sensible approach for performing statistical inference in parametric econometric models involves the definition of a compatibility index that measures the distance between the model and the data. Then, any inference is related to this index: parameters' estimators are defined as the parameter values that maximize the compatibility according to that index, that is, the minimizers of the distance between the model and the data; specification tests are based on the compatibility index value, that is, on the minimum distance between the model and the data; and, finally, tests for parameter restrictions are based on the change in the compatibility index derived from imposing these restrictions. This approach is called Minimum Distance (MD, hereinafter) and traces its history back to Pearson's chi-square goodness of fit test and the corresponding Fisher's minimum chi-square estimator. The approach is sensible because it fully exploits the process of optimizing the measure of distance, using both the minimizer and the minimized value of the compatibility index.

An application of this general principle is the Generalized Method of Moments (GMM, henceforth) approach based on the GMM estimator, and its companion Overidentifying Restrictions Specification (ORS, hereinafter) test. The GMM approach has become the standard inference methodology for models defined by Conditional Moment Restrictions (CMR, hereinafter), that is, models defined by assuming that some parametric functions of interest have zero conditional mean when evaluated at the so called 'true parameter values'. Equivalently, CMR models imply that those parametric functions are uncorrelated to any other auxiliary function that depends only on the conditioning variables. The GMM approach chooses some of these auxiliary functions, which are called instruments in econometrics, and defines the compatibility index in terms of the magnitude of the correlations between the parametric functions and the instruments.

Since GMM selects only some of such auxiliary functions, ignoring most of them, it does not exploit the full definition of the model, and so, it does not yield consistent inference unless additional assumptions are imposed. In particular, as an estimation method, GMM fails in identifying the true parameter values and hence, it is inconsistent unless one restricts either the parametric set or the distribution of the conditioning variables, see Domínguez and Lobato (2004). Similarly, as a specification test, the ORS test is inconsistent unless one restricts either the conditioning variables distributions or the set of alternatives of interest, see Newey (1985), Tauchen (1985), Bierens (1990) or Stute (1997). Therefore, in CMR models the GMM approach presents the conceptual problem that its consistency depends on the distribution of the conditioning variables, which is typically considered to be irrelevant. Moreover, the additional assumptions required by the GMM approach to yield consistent inference procedures are case dependent, in many cases unknown, and in practice never tested. GMM practitioners just assume that the selected Unconditional Moment Restrictions (UMR, hereinafter) identify the parameters of interest.

In order to employ all the information contained in the CMR, the MD methodology can be carried out in two ways, see Domínguez and Lobato (2003) and Delgado, Domínguez and Lavergne (2006) for a comparison between both approaches. In the first way, the compatibility index measures the distance between the empirical and the theoretical (zero) conditional expectation functions. This approach involves the use of nonparametric techniques to estimate the conditional expectation functions, see Lavergne and Patilea (2008).

The second approach exploits the same fact as GMM, namely, that the conditional expectation of a random variable is zero if and only if this random variable is uncorrelated to any function of the conditioning variables, but instead of just checking a few correlations as GMM does, it checks all of them. Compared to GMM, this second approach presents the challenging feature of involving an infinite number of UMR's. For specification testing this idea was followed by Bierens (1982), who proposed an alternative to the ORS test based on a compatibility index that accounts for an infinite number of UMR. Bierens' test is the first example of a consistent specification test for CMR models. Since then, a number of specification tests have been proposed based on compatibility indices that impose the whole information provided by the CMR's, see for instance, Bierens and Ploberger (1997), Stute (1997), Koul and Stute (1999), Carrasco and Florens (2000) or Escanciano (2006).

Following the second approach, Domínguez and Lobato (2004, DL hereinafter) addressed the parameter estimation for CMR models and showed that minimizing the compatibility index that employs an infinite number of UMR yields a consistent estimator of the parameters of interest, without any artificial additional assumption on the conditioning variables distribution. We call this estimator the Conditional Method of Moments (CMM, henceforth) estimator. The CMM estimator is very simple to implement and consistent under general conditions. DL proposed a modification of the CMM estimator to achieve an efficient estimator. However, constructing an efficient estimator involves computationally costly nonparametric estimation techniques, which can be arbitrary because of the selection of the smoothing parameter. Hence, before embarking on efficient estimation, it is sensible to test first whether the model is properly specified.

Therefore, the current situation regarding the second approach is that there are some inferential procedures for estimation and others for testing that are apparently unrelated. However, both parts should be linked since estimation is a fundamental part of model checking. For instance, plugging an inconsistent estimator into the specification test implies that the type I error is not controlled. In addition, plugging in an efficient estimator, with the idea of improving the power, is useless since efficiency is a property studied for correctly specified models, which is not maintained under the alternative.

Hence, the purpose of this article is to propose a global methodology for performing consistent statistical inference on CMR models by extending the results in DL. In particular, once the CMM estimator has been computed, we propose to use the minimized value of its objective function as a test statistic for testing whether the model is correctly specified. In this way, we recover the unified approach for inference, and relate in a natural way both parts of inference, estimation and diagnostic testing.[2] Moreover, we provide expressions for both the estimator and the specification test that relate them to other popular estimators and tests and allow a familiar interpretation of the obtained results. An additional contribution of this article is the discussion of the properties of the CMM estimator when the model is not correct. This discussion not only allow us to study optimality of the test but also to assess that optimality is preserved when the test is implemented using bootstrap critical values.

In addition to the methodological contribution, the test also presents practical and technical advantages over the existing ones. The first one is its simplicity: the test statistic is a by-product of the estimation procedure. Secondly, the novel bootstrap procedure that we propose for estimating the critical values is simpler than those previously employed because it just involves running linear regressions. Finally, we stress that the framework considered in this article is very general compared to the previous literature. In particular, as opposed to Bierens and Ploberger (1997), our framework not only covers regression models but also CMR models defined by general functional forms. In addition, as opposed to Koul and Stute (1999), instead of Markovian conditions, we prefer to restrict the data dependence by using martingale difference conditions, which naturally appear in rational expectations models, and consider general CMR models instead of autorregressions.

The plan of this article is the following. Since we are introducing a new methodology for CMR models, we consider it necessary to remark the inconsistency of the previous GMM methodology when applied to this framework. This is the aim of Section 2, where

[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).

we discuss an example motivated by the consumption-based asset pricing model. This example, which is more realistic than those in DL, shows that the inconsistency of GMM estimators is a common problem in nonlinear economic models. Section 3 presents the proposed methodology, first summarizing DL results concerning the CMM estimator and then introducing the specification test and the bootstrap procedure for estimating its critical values. We stress the originality and simplicity of the bootstrap that can be widely applied. For a basic consumption-based asset pricing model calibrated for the US economy, Section 4 presents a Monte Carlo study that compares our approach with the GMM methodology. In addition, in this section we also consider the linear and the Box-Cox transformation model. Section 5 concludes. In the Appendix we collect the assumptions.

2 Example

In this section we present a simple example of the identification problem the GMM methodology faces in CMR models. The example is a simplified version of the consumption asset pricing model. The identification problem not only leads to the inconsistency of GMM estimators but also implies that any specification test that plugs in its value will fail in controlling the type I error.

Initially, consider the basic intertemporal consumption-based asset pricing model whose first order condition is given by

$$\beta_0 E_t \left[\left(\frac{c_{t+1}}{c_t} \right)^{-\alpha_0} r_{t+1} \right] = 1, \tag{1}$$

where β_0 is the discount parameter, c_t is the representative agent's consumption at time t, r_t is the gross return to bonds at time t, and E_t denotes the conditional expectation given the information set publicly available at time t. The main parameter of interest is α_0 that corresponds to the representative agent's relative risk aversion coefficient. In Section 4 we employ model (1) calibrated for the US economy to compare the finite sample behavior of our methodology with GMM procedures. However, in this section, we will simplify model (1) to clearly point out the identification problem. Consider the CMR model

$$E_{t-1}(x_t^{-\alpha_0}) = K_0, \tag{2}$$

where x_t is a positive random variable, and α_0 and K_0 are the true parameter values. This can be regarded a simplification of model (1) where x_t denotes the ratio between current and past consumption, and where we have constrained r to be constant ($\bar{r}$, say) so that $K_0 = 1/\bar{r}\beta_0$. We focus on inference on the risk aversion parameter, α_0, and hence we assume that K_0 is known. Model (2) has been employed by Gregory, Lamarche and Smith (2002) to compare the finite sample performance of GMM estimators and estimators based on the Kullback-Leibler information criterion, see their equation (16).

Suppose one is interested in performing statistical inference about α_0 and about the suitability of the model. For the time being, 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. Hence, the effect of the infinite past on the present value is completely summarized in the last period. Note that the ergodic stationary distribution is the normalized eigenvector associated to the unit eigenvalue, call it (p_1, p_2, p_3). In addition, assume that $\{x_1, x_2, x_3\} = \{0.8, 1, 1.2\}$, so that consumption can decrease, stay at the same level or increase. Notice that this simple example allows for complicated dependence structure, such as conditional heteroskedasticity, which is a common feature of financial series.

The standard econometric approach to make inference on α_0 is GMM, which is based on the selection of one (or more) UMR, such as

$$E(x_t^{-\alpha}) = K_0. \tag{3}$$

That is, GMM assumes that there is a single value of the parameter α that solves (3). Although this assumption is generally true in a neighborhood of α_0, a close look to the condition shows that this is not globally true. Indeed, notice that condition (3) just imposes that

$$p_1 0.8^{-\alpha} + p_2 + p_3 1.2^{-\alpha} = K_0. \tag{4}$$

Because $0.8^{-\alpha}$ is an increasing function of α, whereas $1.2^{-\alpha}$ is decreasing, the function $p_1 0.8^{-\alpha} + p_2 + p_3 1.2^{-\alpha}$ exhibits an asymmetric $U-$shape. In Figure 1 we have plotted this function for the case $(p_1, p_2, p_3) = (1/3, 1/3, 1/3)$. Figure 1 shows that equation (4) has no solutions when $K_0 < 1$, has a unique solution if $K_0 = 1$ and, for each $K_0 > 1$, there are two values of α that verify condition (3). Therefore, for this unconditional distribution, when the discount factor times the fixed bond return is less than one, GMM based on condition (3) delivers inconsistent estimators because equation (3) has two solutions for α. 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 preserves an asymmetric $U-$shape, so that the identification problem remains. In addition, the problem does not vanish if x has a continuous distribution as one can notice by writing the unconditional expectation as

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

Since $E(x_t^{-\alpha} \mid x_t > 1)$ is decreasing in α, whereas $E(x_t^{-\alpha} \mid x_t < 1)$ is increasing in α, the function $E(x_t^{-\alpha})$ still has an asymmetric $U-$shape.

So far we have considered the simplest case where the selected instrument is the constant, and showed that 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 fact, for the model in hand the optimal instrument does not change the nature of the identification problem since the UMR associated to the optimal instrument may be written as

$$w_1 0.8^{-\alpha} + w_2 + w_3 1.2^{-\alpha} = C_0, \tag{5}$$

where $w_i = p_i z_i$, C_0 is some constant, and z_i denotes in this section the value of the optimal instrument for the state of nature i-th. In the unfeasible optimal instrument case, z_i is known and the argument concerning the $U-$shape still applies to (5), so the same conclusions hold. Furthermore, in the feasible optimal instrument case, where the values for z_i also depend on the α parameter, equation (5) can have more than two solutions. For instance, when p_{11} =.291, $p_{12} = .266$, $p_{21} = .282$, $p_{22} = .294$, $p_{31} = .282$, $p_{32} = .294$, K_0 =1.038, and the true $\alpha_0 = 2.19$, there are two additional values for α that satisfy the unconditional restriction imposed by the optimal instrument, namely, 0.518 and -1.198. In Figure 2 we have plotted the UMR that satisfies the feasible GMM optimal instrument with the three solutions. In addition, we have verified that the CMR model is properly identified, that is, for those specific $p_{ij}'s$ the true $\alpha_0 = 2.19$ is the only value for α that verifies the CMR.

Finally, we just mention that there is nothing special about this example. The Markovian structure was imposed in the example just for simplicity, and for more complicated dependence structures the lack of global identification of GMM also appears. As commented above, the key insight is the difference between global and local identification. GMM focuses on local identification, and assumes the equivalence between local and global identification, see condition (iii) in Theorem 2.1 in Hansen (1982). On the contrary, CMM focuses on global identification, and uses the equivalence between a conditional expectation and an infinite number of unconditional expectations (see Billingsley,1995, Theorem 16.10iii) to globally identify the parameters of interest. Note that in this simple example, since the domain of the conditioning variable is just three points, one can globally identify the true parameter by properly selecting three unconditional restrictions. Similarly, if the domain of the conditioning variable were K points, one could achieve global identification by properly selecting K unconditional restrictions. By extension, with continuous conditioning variables one would need an infinite number of unconditional restrictions to achieve global identification. Therefore, the main message from this example is that global identification is not the rule but the exception in nonlinear CMR models which are estimated using a finite number of UMR.

In summary, this section shows that given a CMR and a set of UMR's, there can exist some data generating process (DGP, hereinafter) such that the UMR's fail to globally identify the true parameter values. In other words, given a CMR and a set of UMR's, one should not expect global identification uniformly over a large set of DGP's. More formally, if $E(h(y_t, \theta_0) \mid x_t) = 0$ then $E(h(y_t, \theta_0)W(x_t)) = 0$ for *any* W, but if one checks only *some* $W's$, there are many other parameters that may also verify these particular UMR's. For instance, consider a different parameter value θ_1 and call $D_{\theta_1}(x_t) = E(h(y_t, \theta_1)W(x_t) \mid x_t)$. Note that, by the Law of Iterated Expectations, $E(h(y_t, \theta_1)W(x_t)) = 0$ is equivalent to $ED_{\theta_1}(x_t) = 0$. Therefore, as soon as the variable $D_{\theta_1}(x_t)$ takes positive and negative values, the lack of identification problem arises. In fact, θ is identified only if the distribution of x_t does not centers $D_\theta(x_t)$ for any value of θ different from θ_0.

3 Theory

3.1 The model

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 an econometric model defined by the following CMR

$$E(u(y_t, \theta_0) \mid x_t) = 0, \quad a.s. \tag{6}$$

for a unique value $\theta_0 \in \Theta$, where $\Theta \subset \mathbb{R}^k$ is the parametric space. Equation (6) defines the unknown true value θ_0 for the parameter θ. 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 twice partially differentiable and its second derivatives have uniformly bounded second 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). In addition, from a statistical point of view, the function $u(y_t, \theta_0)$ is a martingale difference sequence with respect to the increasing sequence of sigma fields defined by the conditioning variables. For more details, see DL.

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. In this article we consider the case $l = 1$.

3.2 The CMM estimator

As commented, in order to consider the full information provided by the CMR, an infinite number of UMR's must be imposed. There are many possible UMR's sets with infinite elements that are equivalent to the CMR, see Stinchcombe and White (1998). For the choice of the set we follow DL, although analogous results follow for other sets. Let P_{z_t} and P_{x_t} denote the probability laws of the random vectors z_t and x_t, respectively, and let $I(x_t \leq x)$ denote the indicator function of the orthant defined by x, that is, it equals one when each coordinate in x_t is less or equal than the corresponding coordinate in x; and equals zero otherwise. DL used the set of UMR[3]

$$\left\{ U(v, \theta_0) = \int_{\mathbb{R}^p} u(y, \theta_0) I(x \leq v) dP_{z_t}(z) = 0, \ v \in \mathbb{R}^d \right\} \tag{7}$$

and the compatibility index

$$Q(\theta) = \int_{\mathbb{R}^d} U(v, \theta)^2 dP_{x_t}(v), \tag{8}$$

which is zero at θ_0 if and only if the CMR holds. Hence, $\theta_0 = \arg\min_\theta Q(\theta)$. The CMM estimator of θ_0 is defined by its sample analog. Formally, for a sample of size n, define the sample analogs of $U(v, \theta)$ and $Q(\theta)$

$$U_n(v, \theta) = \frac{1}{n} \sum_{t=1}^{n} u(y_t, \theta) I(x_t \leq v), \tag{9}$$

[3] A note regarding notation: along the article we employ lower case letters for random variables and capital letters for the corresponding integrated variables.

and

$$Q_n(\theta) = \frac{1}{n}\sum_{\ell=1}^{n} U_n(x_\ell,\theta)^2.$$

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 and asymptotically normal. Specifically, they show that under Assumption A in the Appendix

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

where

$$\dot{U}_n\left(v,\theta\right) = \frac{1}{n}\sum_{t=1}^{n}\dot{u}\left(y_t,\theta\right)I\left(x_t\leq v\right), \tag{10}$$

and

$$\dot{u}\left(y,\theta\right) = \frac{\partial u(y,\theta)}{\partial\theta}$$

is the $1\times k$ vector of the first order derivatives. Applying Lemmas 1 and 2 in DL it follows that

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \left\{\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}'\left(x_\ell,\theta_0\right)\dot{U}\left(x_\ell,\theta_0\right)\right\}^{-1}\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}'\left(x_\ell,\theta_0\right)B_\Gamma\left(x_\ell\right)+o_p\left(1\right) \tag{11}$$

where

$$\dot{U}\left(v,\theta\right) = \int_{\mathbb{R}^p}\left[\dot{u}\left(y,\theta\right)I\left(x\leq v\right)\right]dP_{z_t}(z)$$

and B_Γ stands for a Gaussian process in the space of cadlag functions, $D\left[-\infty,\infty\right]^d$, centered at zero and with covariance structure Γ, where $\Gamma(r,s) = E(u^2(y_t,\theta_0)I(x_t\leq r\wedge s))$, where $\wedge$ denotes minimum. In addition, note that Γ accounts asymptotically for all the dependence between the exogenous variables x and the generalized errors u.

Since $n^{-1}\sum_{\ell=1}^{n}\dot{U}'\left(x_\ell,\theta_0\right)\dot{U}\left(x_\ell,\theta_0\right) = \int_{\mathbb{R}^d}\dot{U}'\left(v,\theta_0\right)\dot{U}\left(v,\theta_0\right)dP_{x_t}\left(v\right) + o_p\left(1\right)$, then (11) can be written as

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \frac{1}{n}\sum_{\ell=1}^{n}L\left(x_\ell\right)B_\Gamma\left(x_\ell\right)+o_p\left(1\right), \tag{12}$$

where $L\left(x_\ell\right) = \left\{\int_{\mathbb{R}^d}\dot{U}'\left(v,\theta_0\right)\dot{U}\left(v,\theta_0\right)dP_{x_t}\left(v\right)\right\}^{-1}\dot{U}'\left(x_\ell,\theta_0\right)$. The first addend in (12) converges to the stochastic integral of the Gaussian process B_Γ. Therefore, $\sqrt{n}(\hat{\theta}-\theta_0)$ is asymptotically a centered Gaussian random variable whose covariance matrix is

$$\left\{\int_{\mathbb{R}^d}\dot{U}'\left(v,\theta_0\right)\dot{U}\left(v,\theta_0\right)dP_{x_t}\left(v\right)\right\}^{-1}\Gamma_2\left\{\int_{\mathbb{R}^d}\dot{U}'\left(v,\theta_0\right)\dot{U}\left(v,\theta_0\right)dP_{x_t}\left(v\right)\right\}^{-1}, \tag{13}$$

where

$$\Gamma_2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \dot{U}'(v_1, \theta_0) \Gamma(v_1, v_2) \dot{U}(v_2, \theta_0) \, dP_{x_t}(v_1) \, dP_{x_t}(v_2).$$

Alternatively, we can write

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} D(x_t) \, u(y_t, \theta_0) + o_p(1), \tag{14}$$

where

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

Representation (14) is useful because it allows the application of Hall and Heyde's (1980) martingale Central Limit Theorem to obtain the asymptotic distribution and to directly apply Stute's (1997) results to the test proposed below (see Proposition 5).

3.3 Specification testing

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

$$H_0 : E(u(y_t, \theta_0) \mid x_t) = 0 \quad a.s., \tag{15}$$

for a unique θ_0, whereas the alternative hypothesis is that for any θ

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

We propose testing model adequacy using the test statistic

$$T_n \; = \; nQ_n(\hat{\theta}). \tag{16}$$

In order to derive the asymptotic null distribution of T_n, we consider first the behavior of $U_n(v, \hat{\theta})$ under the null hypothesis. Applying results in DL to the developments in Stute (1997), it follows that under the null hypothesis

$$\begin{aligned} & \sqrt{n} U_n(v, \hat{\theta}) = \sqrt{n} U_n(v, \theta_0) + \dot{U}_n(v, \theta_0) \sqrt{n}(\hat{\theta} - \theta_0) + o_p(1) \\ = \; & \sqrt{n} U_n(v, \theta_0) \\ & - \dot{U}_n(v, \theta_0) \left\{ \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) U_n(x_\ell, \theta_0) + o_p(1). \end{aligned}$$

Therefore,

$$\sqrt{n}U_n(v,\hat{\theta}) \rightarrow_d B_\Phi\left(v\right),$$

where

$$B_\Phi\left(w\right) = B_\Gamma(w) - \dot{U}\left(w,\theta_0\right)\left\{\int_{\mathbb{R}^d} \dot{U}'\left(v,\theta_0\right)\dot{U}\left(v,\theta_0\right)dP_{x_t}\left(v\right)\right\}^{-1}\int_{\mathbb{R}^d} \dot{U}'\left(v,\theta_0\right)B_\Gamma(v)dP_{x_t}\left(v\right).$$

Note that $B_\Phi\left(w\right)$ is a Gaussian process in $D\left[-\infty,\infty\right]^d$ that can be seen as an integral transformation (actually an L^2 residual) of the process B_Γ

$$B_\Phi\left(w\right) = B_\Gamma\left(w\right) - \int_{\mathbb{R}^d} K\left(w,v\right)B_\Gamma(v)dP_{x_t}\left(v\right) \tag{17}$$

with integral kernel

$$K\left(w,v\right) = \dot{U}\left(w,\theta_0\right)\left\{\int_{\mathbb{R}^d} \dot{U}'\left(\xi,\theta_0\right)\dot{U}\left(\xi,\theta_0\right)dP_{x_t}\left(\xi\right)\right\}^{-1}\dot{U}'\left(v,\theta_0\right).$$

So that $B_\Phi\left(w\right)$ is a Gaussian process centered at zero and whose covariance operator can be written as

$$\begin{aligned}\Phi\left(t,s\right) &= \Gamma\left(t,s\right) - \int_{\mathbb{R}^d} K\left(t,w\right)\Gamma(w,s)dP_{x_t}\left(w\right) - \int_{\mathbb{R}^d}\Gamma\left(t,w\right)K(w,s)dP_{x_t}(w) \\ &\quad + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K\left(t,w\right)\Gamma(w,v)K\left(v,s\right)dP_{x_t}\left(w\right)dP_{x_t}\left(v\right).\end{aligned}$$

As a result we establish the following proposition.

Proposition 1: Under the null hypothesis and Assumption A in the Appendix

$$T_n \rightarrow_d \int_{\mathbb{R}^d} B_\Phi^2(v)dP_{x_t}(v).$$

Therefore, the critical values of the test statistic T_n depend on the DGP, complicating statistical inference.

Concerning the behavior under the alternative, it is straightforward to show that the test statistic T_n diverges under any fixed alternative. We formally state the result in the following proposition.

Proposition 2: Under the alternative hypothesis and Assumption B in the Appendix, the proposed test is consistent.

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

$$H_{A,n} : E(u(y_t,\theta_0) \mid x_t) = \frac{g(x_t)}{\sqrt{n}} \qquad a.s., \tag{18}$$

be the sequence of local alternatives of interest. Before stating the properties of the test in this setting we first study the effect of such alternatives on the properties of the estimator. Denote $v(z_t, \theta) = u(y_t, \theta) - (g(x_t)/\sqrt{n})$. Note that under $H_{A,n}$ and assumption C in the appendix, it follows that

$$\begin{aligned} U_n(v, \theta) &= \frac{1}{n} \sum_{t=1}^{n} u(y_t, \theta) I(x_t \leq v) \\ &= \frac{1}{n} \sum_{t=1}^{n} [u(y_t, \theta_0) - u(y_t, \theta)] I(x_t \leq v) \\ &\quad + \frac{1}{n} \sum_{t=1}^{n} n^{-1/2} g(x_t) I(x_t \leq v) + \frac{1}{n} \sum_{t=1}^{n} v(z_t, \theta_0) I(x_t \leq v) \\ &\to_{as} U(v, \theta_0) - U(v, \theta), \end{aligned}$$

uniformly in v, where $U(\theta, v)$ is defined in (7). Hence,

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

As a result, θ_0 still minimizes (8). Hence, we can state the following result.

Proposition 3: Under $H_{A,n}$ and Assumption C in the Appendix, $\hat{\theta} \to_{a.s.} \theta_0$.

Despite this result, the estimator $\hat{\theta}$ exhibits a bias in the asymptotic distribution since, in this local alternative framework, the local linear approximation for $\hat{\theta}$ switches to

$$\begin{aligned} \sqrt{n}(\widehat{\theta} - \theta_0) &= \left\{ \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) V_n(x_\ell, \theta_0) \\ &\quad + \left\{ \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) \dot{U}_n(x_\ell, \theta_0) \right\}^{-1} \frac{1}{n} \sum_{\ell=1}^{n} \dot{U}_n'(x_\ell, \theta_0) G_n(x_\ell) \\ &\quad + o_p(1), \end{aligned}$$

where

$$V_n(x, \theta) = \frac{1}{n} \sum_{t=1}^{n} \left[u(x_t, \theta) - g(x_t)/\sqrt{n} \right] 1(x_t \leq x),$$

and

$$G_n(x) = \frac{1}{n} \sum_{t=1}^{n} g(x_t) I(x_t \leq x).$$

Since V_n 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 $H_{A,n}$ and Assumption C, $\sqrt{n}(\widehat{\theta}-\theta_0)$ is asymptotically normal with mean given by $\left\{\int_{\mathbb{R}^d}\dot{U}'(\xi,\theta_0)\,\dot{U}(\xi,\theta_0)\,dP_{x_t}(\xi)\right\}^{-1}\left\{\int_{\mathbb{R}^d}\dot{U}'(\xi,\theta_0)\,G(\xi)\,dP_{x_t}(\xi)\right\}^{-1}$ where $G(\xi)=\int_{\mathbb{R}^p}[g(x)\,I(x\leq\xi)]\,dP_{z_t}(x)$ and variance given in (13).

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

$$
\begin{aligned}
& \sqrt{n}V_n(v,\hat{\theta})+G_n(v) \\
= \;& \sqrt{n}V_n(v,\theta_0)-\dot{U}_n(v,\theta_0)\sqrt{n}\left(\hat{\theta}-\theta_0\right)+G_n(v)+o_p(1) \\
= \;& \sqrt{n}V_n(v,\theta_0)-\dot{U}_n(v,\theta_0)\left\{\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,\dot{U}_n(x_\ell,\theta_0)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,V_n(x_\ell,\theta_0) \\
& +G_n(v)-\dot{U}_n(v,\theta_0)\left\{\frac{1}{n}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,\dot{U}_n(x_\ell,\theta_0)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}\dot{U}_n'(x_\ell,\theta_0)\,G_n(x_\ell)+o_p(1)\,.
\end{aligned}
$$

The first two addends behave as the linear approximation to U_n under H_0. The next two addends are a bias term. Then, the asymptotic distribution of T_n under $H_{A,n}$ is the same Gaussian process obtained under the null, but centered at the function C where

$$
C(w)=G(w)-\dot{U}(w,\theta_0)\left\{\int_{\mathbb{R}^d}\dot{U}'(v,\theta_0)\,\dot{U}(v,\theta_0)\,dP_{x_t}(v)\right\}^{-1}\int_{\mathbb{R}^d}\dot{U}'(v,\theta_0)\,G(v)dP_{x_t}(v)\,.
$$

Formally, we establish the following proposition.

Proposition 5: Under the local alternative hypothesis (18) and Assumption C in the Appendix,

$$
T_n\rightarrow_d\int_{\mathbb{R}^d}\left[B_\Phi(v)+C(v)\right]^2dP_{x_t}(v).
$$

Therefore, unless $C=0$, the rejection probability under $H_{A,n}$ is larger than the nominal level as in Bierens and Ploberger (1997) and Stute (1997), that is, the test has nontrivial power against the sequence of local alternatives $H_{A,n}$. Moreover, similarly to these references, by recalling representation (14) it can be straightforwardly shown that the test enjoys optimality properties because of the term C.

3.4 Understanding the methodology

3.4.1 The linear case

In order to provide further intuition on the structure of the CMM methodology, consider the linear model

$$E(y_t \mid x_t) = x_t\beta_0,$$

where β_0 represents the $k \times 1$ vector of the true unknown parameters of interest. Note that in this case there is no need of using the CMM estimator since any k linearly independent UMR's identify β_0, both locally and globally.

The CMM estimator is given by

$$\widehat{\beta} = \arg\min_b \sum_s \left(\sum_t (y_t - x_t b) I (x_t \leq x_s) \right)^2 .$$

Notice that this objective function can be written as

$$\sum_s (Y_s - X_s b)^2$$

where we use capital letters to denote the integrated variables

$$X_s = \sum_t x_t I (x_t \leq x_s) , \text{ and } Y_s = \sum_t y_t I (x_t \leq x_s) .$$

Hence, by considering $Y_s - X_s b$ as the residuals of the linear least squares regression on integrated variables, the objective function can be seen as a linear least squares objective function, which employs integrated variables.

The first order conditions are

$$0 = \sum_s X_s' \left(Y_s - X_s \widehat{\beta} \right) ,$$

which can be written in matrix notation as

$$0 = \mathbb{X}' \left(\mathbb{Y} - \mathbb{X}\widehat{\beta} \right) ,$$

where the $n \times k$ matrix $\mathbb{X}$ has X_s as its $s - th$ row, and the n-dimensional vector $\mathbb{Y}$ has Y_s as its $s - th$ component. Therefore, $\widehat{\beta}$ takes the habitual OLS form

$$\widehat{\beta} = (\mathbb{X}'\mathbb{X})^{-1} \mathbb{X}'\mathbb{Y}$$

with the only difference that now the variables have been integrated first.

Similarly, the specification test statistic is the optimized value of the objective function, that is, the residual sum of squares of the regression with the integrated variables. Denoting the integrated errors and residuals by $\mathbb{U}(\beta_0) = \mathbb{Y} - \mathbb{X}\beta_0$ and $\mathbb{U}(\widehat{\beta}) = \mathbb{Y} - \mathbb{X}\widehat{\beta}$, respectively, and using that

$$\widehat{\beta} - \beta_0 = (\mathbb{X}'\mathbb{X})^{-1}\,\mathbb{X}'\mathbb{U}(\beta_0),$$

we derive

$$\mathbb{U}(\widehat{\beta}) = \mathbb{M}_{\mathbb{X}}\mathbb{Y} = \mathbb{M}_{\mathbb{X}}\mathbb{U}(\beta_0),$$

where, following the conventional econometrics notation,

$$\mathbb{M}_{\mathbb{X}} = \mathbb{I}_n - \mathbb{X}\,(\mathbb{X}'\mathbb{X})^{-1}\,\mathbb{X}',$$

so that the specification test statistic is just

$$T_n = \mathbb{U}(\widehat{\beta})'\mathbb{U}(\widehat{\beta}) = \mathbb{U}(\beta_0)'\mathbb{M}_{\mathbb{X}}\mathbb{U}(\beta_0).$$

3.4.2 The nonlinear case

Recall that the leading motivation for our methodology is global identification with nonlinear conditional models. In this subsection we provide further insights from a local perspective. The use of local linear approximations and matrix notation allow us to write $\hat{\theta}$ in a notation more familiar in econometrics, such as we did in the previous subsection. The objective function in matrix notation can be written as a sum of squared residuals as follows: $Q_n(\theta)$ $= n^{-1}\mathbb{U}\left(\theta\right)'\mathbb{U}\left(\theta\right)$, where $\mathbb{U}(\theta)$ is the $n \times 1$ vector whose $t-th$ element is $U_n\left(x_t, \theta\right)$, that is, the integrated generalized errors. In addition, let $\dot{\mathbb{U}}\left(\theta\right)$ be the $n \times k$ matrix whose t-th row is the $1 \times k$ vector $\dot{U}(x_t, \theta)$. Then,

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

Expression (19) 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" ($\mathbb{U}\left(\theta_0\right)$) onto the subspace generated by the "integrated derivatives" ($\dot{\mathbb{U}}\left(\theta_0\right)$). The difference with NLS is that in our case both the regressors and the errors are scaled integrated processes, see (9) and (10). Because the variables are integrated, orthogonality between the errors and regressors, $\mathbb{U}$ and $\dot{\mathbb{U}}$, respectively, is irrelevant for deriving the consistency of the CMM estimator, which just relies on (6). Note that this orthogonality is crucial in OLS and NLS for the consistency of the OLS and NLS estimators.

Further intuition on the test

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

can be gained by employing a straightforward asymptotic expansion of the objective function

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

Equation (20) 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

$$T_n = \mathbb{U}(\theta_0)'\,\mathbb{U}(\theta_0) - \mathbb{U}(\theta_0)'\,\dot{\mathbb{U}}(\theta_0)\left(\dot{\mathbb{U}}(\theta_0)'\,\dot{\mathbb{U}}(\theta_0)\right)^{-1}\dot{\mathbb{U}}(\theta_0)'\,\mathbb{U}(\theta_0) + o_p(1)\,. \tag{21}$$

Now define

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

where the projection matrix $\mathbb{M}(\theta_0)$ can be seen as the finite sample version of the L^2-projection that defined the process B_Φ in equation (17). Then, we can write (21) as

$$T_n = \mathbb{U}(\theta_0)'\,\mathbb{M}(\theta_0)\,\mathbb{U}(\theta_0) + o_p(1)\,, \tag{22}$$

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 free of nuisance parameters since the variables involved are not stationary but integrated.

Equation (21) provides a decomposition of T_n in terms of locally identifying and locally overidentifying restrictions analogous to that in the GMM approach, see Hall (2005), Sections 3.3 and 5.1.3. According to that decomposition, $\mathbb{P}(\theta_0)\,\mathbb{U}(\theta_0)$ are the k locally identifying restrictions whereas $\mathbb{M}(\theta_0)\,\mathbb{U}(\theta_0)$ are the $n-k$ locally overidentifying restrictions. This popular decomposition allows us to study the local behavior of both the estimator and the test. First, note that, since locally the model is linear, only k conditions are needed for local identification. Second, consider the local alternatives introduced above and notice that in that case, the identifying restrictions are $\mathbb{P}(\theta_0)\,(\mathbb{U}(\theta_0)+\mathbb{G})$ and the overidentifying restrictions are $\mathbb{M}(\theta_0)\,(\mathbb{U}(\theta_0)+\mathbb{G})$, where $\mathbb{G}$ is the $n-$th dimensional vector with t-th coordinate equal to $G(x_t)$. Following the GMM approach, the asymptotic distribution of the estimator would not be affected by local misspecification if $\mathbb{P}(\theta_0)\,\mathbb{G}=0$, and the test would be unable to detect the local alternative if $\mathbb{M}(\theta_0)\,\mathbb{G}=0$, i.e., if $\mathbb{G}$ is collinear with $\dot{\mathbb{U}}(\theta_0)$.

3.5 Bootstrap test

Since Φ depends on the DGP, the asymptotic distributions of both $U_n(v, \hat{\theta})$ and T_n also do. Hence, the theoretical results for the specification test cannot be automatically applied because there are not readily available 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, Florens and Renault (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 three motives. First, the main reason is the simplicity and widely applicability of the proposed bootstrap test. Contrary to the current applications of the bootstrap in specification testing, which are computationally complicated, the proposed bootstrap test is computationally very simple to implement because it just involves running linear regressions, see Remark 4 below. Second, the bootstrap test is valid under heteroskedasticity of any form and it is not a case specific procedure. Third, 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. From expression (22), we propose to estimate consistently the distribution of T_n with the distribution of T_n^* where

$$T_n^* = \mathbb{U}^*(\hat{\theta})' \mathbb{M}(\hat{\theta}) \mathbb{U}^*(\hat{\theta}), \tag{23}$$

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

$$U_n^*(x, \hat{\theta}) = \frac{1}{n} \sum_{t=1}^{n} u(y_t, \hat{\theta}) I(x_t \leq x) w_t, \tag{24}$$

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. In this article we follow Mammen (1993) and Stute, González-Manteiga and Presedo-Quindimil (1998) and the employed sequence $\{w_t\}$ 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 $P(W = 0.5(1 + \sqrt{5})) = 1 - (1 + \sqrt{5})/2\sqrt{5}.$ Notice that the third moment of W is equal to 1, and hence, this selection of $\{w_t\}$ guarantees that the original process U_n and the bootstrap process U_n^* possess the same asymmetry coefficient.

In detail, one would test the null hypothesis (15) as follows:

Step 1: Minimize the OF (8) 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 (9), and the integrated derivatives, $\dot{U}_n(x_t, \hat{\theta})$ from expression (10) 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 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 (24).

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 1. 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 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.

Remark 2. The proposed bootstrap procedure is consistent because given the data, $\mathcal{X}_n$, $\sqrt{n}U_n(x, \theta_0)$ and $\sqrt{n}U_n^*(x, \hat{\theta})$ share the same asymptotic limit distribution, namely B_Γ, i.e.,

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

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

$$P(\sqrt{n}U_n^*(x, \hat{\theta}) \le s \mid \mathcal{X}_n) \rightarrow_{a.s.} P(B_\Phi(x) \le s) \text{ as } n \rightarrow \infty$$

plus tightness a.s.. Then, since any projection is a linear continuous operator, any projection of $\sqrt{n}U_n(x, \theta_0)$ and $\sqrt{n}U_n^*(x, \hat{\theta})$ also enjoys the same limit distribution. Hence, consistency of bootstrap distribution trivially follows. Formally, we establish the next proposition.

Proposition 6: Under either H_0 and Assumption A in the Appendix or $H_{A,n}$ and Assumption C in the Appendix or under H_A and assumption D in the appendix,

$$T_n^* \Rightarrow_* \int (B_\Phi)^2 \, dP_{x_t} \ a.s.. \tag{25}$$

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

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 the simulation procedure described above.

Remark 3. 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 4. The wild bootstrap proposed in (23) 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 is asymptotically equivalent to resampling a process based on the decomposition

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

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 $U_n(x, \theta_0)$, 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 $U_n(x, \theta_0)$, 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 $U_n^*(x, \hat{\theta})$, which implies a great computational advantage.

Remark 5. Since under $H_{A,n}$ the bootstrap still fits the null distribution, the optimal original properties of the test still hold.

4 Simulations

In this section we illustrate the performance of the proposed methodology and compare it to GMM. We consider three models: a simplified version of the consumption-based asset pricing model, the linear model and the Box-Cox transformation model. It is well known that GMM faces severe estimation problems in the context of asset-pricing models, see Hall (2005) textbook and Stock and Wright (2000).

4.1 A simplified consumption-based asset pricing model

In our simplified version of the consumption-based asset pricing model we assume that the consumption ratio follows a Markov chain with four states and that there is only a risk free asset in the economy.[4] Hence, the risk free rate return is just the inverse of its price. This model is a generalization of the famous Mehra and Prescott (1985) model, which assumes that the consumption ratio follows a Markov chain with only two states. The Markovian assumption is a convenient way of reflecting that all past information can be summarized by the last value. In particular, we have calibrated the Markov chain using U.S. per capita real monthly consumption in nondurables and services data from January, 1959 to December, 2007. In order to define the states, we split the data into four groups according to the consumption ratio quartiles, and define the states as the median values for each group. That is, the four possible values for the consumption ratio (call them $\lambda_{(i)}$, i=1,2,3,4) are the 1/8, 3/8, 5/8 and 7/8 quantiles of the empirical distribution of the consumption ratio data. The i, j element from the transition matrix (call it p_{ij}) is the scaled number of times that we observe that the consumption ratio passes from group i to group j, for $i, j = 1, 2, 3, 4$. As a result, we set

$$\begin{pmatrix} \lambda_{(1)} \\ \lambda_{(2)} \\ \lambda_{(3)} \\ \lambda_{(4)} \end{pmatrix} = \begin{pmatrix} 0.999815 \\ 1.004133 \\ 1.007243 \\ 1.011996 \end{pmatrix}$$

and

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 0.20 & 0.28 & 0.31 & 0.21 \\ 0.21 & 0.22 & 0.28 & 0.29 \\ 0.26 & 0.29 & 0.21 & 0.24 \\ 0.33 & 0.20 & 0.20 & 0.27 \end{pmatrix}.$$

The model establishes that, when the system is at state i, the next period price of the risk free asset is given by

$$\beta_0 \sum_{j=1}^{4} \lambda_{(j)}^{-\alpha_0} p_{ij}.$$

Define

$$u_{t+1} = P_{t,t+1} - \beta_0 \lambda_{t+1}^{-\alpha_0}, \tag{26}$$

[4] We thank N. Kocherlakota for suggesting this design.

where $P_{t,t+1}$ denotes the price at time t of the riskfree bond maturing at time $t+1$ and λ_{t+1} is the consumption ratio at time $t+1$. Then, using the notation introduced in Section 2, data are generated so that the condition

$$E_t(u_{t+1}) = 0$$

is satisfied.

In the simulations we employ different combinations of (α_0, β_0) : β_0 takes two different values, 0.99 and 1.01, while α_0 takes five possible values, 0.5, 1, 2, 5 and 10. Note that we have considered a value for β_0 bigger than 1, since, as Kocherlakota (1996) points out, "the restriction that the discount factor is less than one emerges from introspection". In addition, a β_0 bigger than 1 solves the risk free rate puzzle and is not incompatible with economic theory, see Kocherlakota (1990). We follow Mehra and Prescott (1985) and set the maximum value for the risk-aversion parameter to be 10. We stress that this is a very simple model since it links consumption and equity return in a very tight manner so that consumption equals the dividend on aggregate equity. More complicated models that involve additional preference parameters could be entertained along the lines of Epstein and Zin (1989, 1991) "generalized expected utility", Constantinides (1990) "habit formation" or Abel (1990) "relative consumption". The considered model is just a simple example to examine how the CMM methodology performs in finite samples relative to the benchmark established by the efficient GMM methodology.

We compare the CMM estimator to the unfeasible efficient GMM estimator that employs as instruments

$$E_t(u_{t+1}^2)^{-1}E_t(\partial u_{t+1}/\partial\alpha_0)$$

and

$$E_t(u_{t+1}^2)^{-1}E_t(\partial u_{t+1}/\partial\beta_0)$$

where u_{t+1} is defined in (26) for the true α_0 and β_0. Notice two properties of these instruments. First, these instruments are unfeasible not only because they require the knowledge of the functional form of the involved conditional expectation functions, but also because these functions are evaluated at the true values of α_0 and β_0. Second, these instruments take into account the heteroskedasticity present in the model. We denote this unfeasible efficient estimator by EGMM.

We note that although we do not know with absolute certainty if for the considered values of the consumption ratios, $\lambda_{(j)}$, and the transition probabilities, p_{ij}, the two unfeasible efficient instruments identify globally the true parameter values, we believe they do. This belief is based on the plot of the two orthogonality conditions defined by the optimal instruments for some sensible values of β and for α in a sensible range $(0, 10)$. In these cases the two conditions do identify globally the true parameter values since the two orthogonality conditions were monotonically increasing in α in this range. However, we have also observed that by changing the values of the consumption ratios, $\lambda_{(j)}$, or the values of the transition probabilities, p_{ij}, the monotonicity is lost and globally identification can not be assured. Even if we believe that the EGMM orthogonality conditions identify the true parameters values, and hence, EGMM would be theoretically the best estimation procedure one could conceive, the message we are going to learn from the next simulation exercise is that EGMM can be very unreliable in practice, as opposed to the CMM estimator, which is more robust.

In Table I we report the bias and variance of both EGMM and CMM estimators for four sample sizes: 100, 500, 1000 and 5000, as well as the asymptotic values, which can be easily calculated given the simple Markov structure we are assuming. In Table I we consider $\beta_0 = 0.99$ and four possible values for α_0 : 0.5, 1, 2 and 5. Table II expands the results in Table I in two fronts. First, for the case $(\alpha_0, \beta_0) = (2, 0.99)$ we consider additional sample sizes: 10000, 20000 and 40000. Second, we also consider a case where β_0 is bigger than 1, and a case where α_0 takes a large value, $\alpha_0 = 10$. In both tables the number of replications is 5000.

Regarding these results, the first aspect we must mention is that the EGMM objective function is essentially flat in the dimension of α_0 for small and moderate sample sizes. This behavior is common for this example, see the figures in section 3.2 in Hall (2005). Hence, obtaining the EGMM estimates is not simple. Initially, we attempted the minimization of the EGMM objective function using the IMSL subroutines DUMPOL, DBCONF, DBCPOL and DUMINF. These subroutines differ on several aspects, such as the use of the gradient or the Hessian. However, all these subroutines were unable to deliver the EGMM estimates for a considerable number of replications, especially for $n = 100$ and 500. Hence, we computed a two-step Newton-Raphson procedure starting from the true parameter value. Since we still obtained very erratic EGMM estimates in some cases, we decided to report the EGMM results only for the cases where both EGMM estimates were positive and when the EGMM estimate for α_0 was below 45. Hence, the EGMM results in Tables I and II, consider only these cases. This truncation is crucial for the $n = 100$ case, somewhat important for $n = 500$, and marginal for $n = 1000$. We never needed to implement this truncation for sample sizes bigger than 1000. Contrary to EGMM, we stress that the CMM estimates are very simple to obtain due to their similarity with NLS.

Table I provides several initial messages:

a) In all situations the parameter β_0 is much better estimated than the parameter α_0.

b) Both estimators present larger bias and variance for larger values for α_0.

c) Regarding bias, EGMM typically presents a positive bias, whereas the CMM estimator bias is negative.

d) The variances of the CMM estimators do not decay monotonically to zero for small and moderate sample sizes. In fact, it appears that these variances are smaller in the $n = 100$ case than in the $n = 500$ case.

e) The variances of the EGMM estimators are much higher than those of the CMM estimators for sample sizes below $n = 5000$.

These points are not the most interesting messages from Table I, though. Notice that in Tables I and II we include the variances of the asymptotic distributions of these estimators in all cases. These asymptotic variances provide two very interesting messages. First, the asymptotic gain of the EGMM estimator with respect to the CMM is very small. Second, dividing these theoretical variances by the sample sizes and comparing these numbers to the finite sample variances, we can see that the CMM empirical finite sample variances become closer to the corresponding implied theoretical values much faster than EGMM. In fact, one can see that asymptotic theory provides a good approximation for sample sizes from $n = 1000$ for the CMM estimators, whereas for the EGMM estimator this does not happen for any sample size considered in Table I. This is the main reason for exploring in Table II the behavior of these estimators for additional sample sizes: 10000, 20000 and 40000.

Table II also expands the results of Table I by considering the case where $\beta_0 = 1.01$. The main messages from Table II are as follows. First, by comparing the cases $(\alpha_0, \beta_0) = (2, 0.99)$ with $(\alpha_0, \beta_0) = (2, 1.01)$ we observe that the fact that β_0 takes values bigger than 1 does not affect estimation in any significant aspect. Second, the EGMM estimator does not show its theoretical asymptotic advantage for the sample sizes considered here, although it is comparable to the CMM estimator for sample sizes as large as 20000. In fact, Table II indicates that asymptotic theory only provides a good approximation for the behavior of the EGMM for sample sizes as large as $n = 10000$. In order to illustrate better the relative performance of these estimators in Table III we report the variance ratios between these two estimators for the case $(\alpha_0, \beta_0) = (2, 0.99)$. Note that these ratios start to be close to their asymptotic values only from $n = 5000$, and, in addition, they converge very slowly to their asymptotic values.

Regarding the specification test, we do not report any results for this model because, given the poor performance of the estimators, very large sample sizes are needed to reasonably control the type I error.

4.2 The Linear and the Box-Cox Transformation Model

The Monte Carlo results in the previous subsection are interesting since they show that for a fundamental economic model realistically calibrated for the US economy we find out that the unfeasible efficient GMM estimator cannot reliably estimate the true parameter values for the considered sample sizes. In addition, the performance of the CMM estimator is reasonable, at least for large samples. However, these simulation results may wrongly convey the idea that extremely large sample sizes are needed for the proposed procedures to work. That was true for the previous example because of the ill-conditioned nature of the employed DGP, which led to essentially flat objective functions. In order to examine the finite sample behavior in more standard situations, in this subsection we consider two simple models: the linear and the Box-Cox transformation model.

4.2.1 The Linear Model

Although the results in this article refer to the general nonlinear model, it is of interest to examine the performance of the CMM estimator in the simple linear case that establishes that

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

We compare the CMM estimator with the efficient GMM estimator under homoskedasticity, which we call EGMM, which employs as instruments $\{1, x_t\}$. We set $\alpha_0 = 1$ and $\beta_0 = 2$, the x_t are iid draws from a $N(0,5)$, and we consider three cases for u_t : an iid sequence of N(0,1) variates, an iid sequence of standardized χ_1^2, and a heteroskedastic case where $u_t \mid x_t$ follows $\exp(0.25x_t)$ times the standardized χ_1^2. The sample sizes are 50, 100 and 200. The number of replications is 5000. This design has been partially motivated by Shin (2008).

In Table IV we have bias and MSE for the four estimators, $\widehat{\alpha}_{EGMM}$, $\widehat{\alpha}_{CMM}$, $\widehat{\beta}_{EGMM}$ and $\widehat{\beta}_{CMM}$. The first message from Table IV is the small magnitude of the bias for all estimators and any sample size. The behavior of the MSE is more revealing. Whereas for the two homoskedastic cases the EGMM estimators dominate the CMM estimators, in the heteroskedastic case the situation is reversed, so that for example, the MSE of the EGMM estimator of β more than doubles the MSE of the CMM estimator. Table IV reinforces the message of the robustness of the CMM estimators, which performs reasonably well in a variety of frameworks.

Next, we compare our proposed specification test to Stute's (1997). We are using Stute instead of the ORS test because the ORS test cannot be computed since we are in the exactly identified case. We report the results for the proposed bootstrap test explained in Section 3 and for the test based on the test statistic

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

where θ represents (α, β). This test statistic is motivated by Stute (1997), and extended by Delgado, Domínguez and Lavergne (2006) who also proposed bootstrap estimators for the critical values.

Table V reports the empirical rejection percentages 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. The number of bootstrap replications is 99, and we conducted 2000 experiments. Again, Table V 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 V shows, where we have added the results for two additional sample sizes, 500 and 1000, for the heteroskedastic case.

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

$$E\left(y_t \mid x_t\right) = \alpha_0 + \beta_0 x_t + \gamma_0 x_t^2 \tag{28}$$

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

$$E\left(y_t \mid x_t\right) = \begin{cases} \alpha_0 + \gamma_0 + \beta_0 x_t & \text{if} \quad -0.2 \leq x_t \leq 0.2, \\ \alpha_0 + \beta_0 x_t & \qquad \text{otherwise,} \end{cases} \tag{29}$$

with $\gamma_0 = 3.5$.

Table VI reports the percentage power for both tests for a 5% nominal level. For the quadratic alternative with homoskedastic noise, Table VI 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.

4.2.2 The Box-Cox Transformation Model

Recall 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

$$E\left(y_t^{(\lambda_0)} \mid x_t\right) = \alpha_0 + \beta_0 x_t. \tag{30}$$

Shin (2008) has also compared the finite sample behavior of the CMM estimator versus alternative estimators for a transformation related to the Box-Cox. We generate the data for the $\lambda_0 = 0$ case, which allow the implementation of the unfeasible efficient GMM estimator under homoskedasticity. This estimator, which we call EGMM, 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

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

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

Table VII reports the bias and mean squared error for the EGMM estimators and for the CMM estimators. Note that for all cases the biases are remarkably small and that the parameter λ_0 is the one estimated more accurately. In addition, the decay of all mean squared errors is roughly proportional to the sample size, indicating that the asymptotic approximation seems to work even for such small sample sizes. For the first two homoskedastic cases the EGMM estimator is theoretically the best estimator one can choose, and this fact is reflected in the tables. For the heteroskedastic case there is no such theoretical advantage, and in fact, the mean squared error of the CMM estimator of α_0 is lower than the one of EGMM estimator.

Next, we compare the specification tests. Again, we compare our proposed specification test, T_n, to Stute's (1997). Similar to the linear case, we are comparing our test against Stute instead of the ORS test for two reasons. First, the ORS test is inconsistent. Second, the

ORS test cannot be computed for the efficient GMM approach since we are just employing the three orthogonality conditions corresponding to the optimal instruments. Table VIII presents the empirical rejection percentages of the proposed specification test T_n and of the $T_n^{(2)}$ test for testing the null hypothesis that there exists some α_0, β_0 and λ_0 such that model (30) is correct, that is,

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

We consider three nominal levels: 0.10, 0.05 and 0.01, the same three sample sizes and the same number of replications as above. These tables show 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 EGMM estimator presents size distortions that seems to increase as the sample size increases. Since the EGMM estimator does not show any bias, this size distortion has to be a finite sample issue. In order to confirm it, in Table VIII we consider three additional sample sizes, $n = 600$, 1500 and 3000 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. Table VIII illustrates again that the methodology proposed in this article is more robust than the GMM methodology.

The final Monte Carlo results in Table IX refer to a power comparison between both tests for two cases motivated by (28) and (29). Specifically, we consider that

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

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,

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

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 IX shows that for 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 VIII showed that the $T_n^{(2)}$ presents in all cases higher rejection probabilities under the null.

5 Conclusions

In this article we have introduced a simple unified methodology for performing consistent statistical inference for CMR models. Consistency derives from the use of a compatibility index that takes into account an infinite number of UMR, which fully impose the definition of the model. The methodology is very simple to implement and in simulations it appears to be more robust than the GMM methodology. In addition, our unified approach allows the proposal of a new extremely simple bootstrap procedure for estimating the critical values. Despite the considered models are defined through conditional and nonlinear equations, the proposed wild bootstrap is extraordinarily simple to implement since it just requires computing two integrated series (the residuals and their derivatives) and running linear regressions.

Regarding consistent specification testing, our approach highlights the importance of the estimation stage in the model checking stage, an issue that has been overlooked in the previous literature. In particular, the practice of plugging in efficient, but possibly inconsistent, estimators does not present any theoretical advantage. In case the employed estimator is inconsistent, the test cannot control the type I error, and in case the estimator is consistent and efficient, the test does not necessarily presents more power (the reason is that the estimator is efficient under the assumption of correct specification, whereas the power of a test refers to situations where the model is not properly specified).

Finally, it could be objected that the proposed MD methodology does not weight efficiently the infinite UMR's. However, since constructing an efficient compatibility index involves computationally costly, and many times arbitrary, nonparametric estimation techniques, it is sensible to test first whether the model is properly specified. Hence, a sensible strategy is to carry out the specification test using the CMM methodology, and pursue an efficient MD methodology only in case the test indicates that the model is correctly specified.

APPENDIX: Assumptions

This appendix collects the assumptions for the propositions. These sets of assumptions do not pretend to be minimal.

Assumption A (see DL):

a) *$u(y,\cdot)$ is continuous in Θ for each y in 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. if and only if $\theta = \theta_0$.*

b) *Z_t is ergodic and strictly stationary.*

c) *$\Theta \subset R^m$ is compact.*

d) *$u(y,\cdot)$ is once continuously differentiable in a neighborhood of θ_0 and satisfies that $E\left[\sup_{\theta \in \aleph_0} \left|\dot{u}(Y_t,\theta)\right|\right] < \infty$ where $\aleph_0$ 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) $E\left[u^4(y_t,\theta_0)\left\|X_t\right\|^{1+\delta}\right] < \infty$.

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

Assumption B:

a) $\hat{\theta} \rightarrow_{a.s.} \theta_1$.

b) $u(y,\cdot)$ is continuous at θ_1 for each y and either b1) or b2) below holds: $\forall\theta \in \aleph_1$ neighbohood 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.*

Assumption C:

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

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

Assumption D:

a) Assumption B.

b) Assumption A, replacing θ_0 by θ_1.

Acknowledgement

This article is based on The Mario Henrique Simonsen Lecture at the 2007 Latin American Meeting of the Econometric Society. We thank Marine Carrasco, Miguel Delgado, Juan Carlos Escanciano and Narayana Kocherlakota for useful conversations and seminar participants at CIREQ-Time Series Conference, Indiana University and Yale University for useful comments. 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. Lobato acknowledges financial support from the Mexican CONACYT, reference number 59028, and from Asociación Mexicana de Cultura. This article was also supported by Spanish Ministry of Education, Plan Nacional I+D+i, SEJ2007-62908.

References

Abel, A.B. (1990), "Asset Prices under Habit Formation and Catching up with the Joneses", American Economic Review, Vol. 80, No. 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, pp.38-42.

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.

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

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.

Constantinides, G.M. (1990), "Habit Formation: A Resolution of the Equity Premium Puzzle", Journal of Political Economy, 98, 519-543

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

Davidson, R. and J.G. MacKinnon (2006), "Bootstrap Methods in Econometrics", Palgrave Handbooks of Econometrics: Vol. 1 Econometric Theory edited by Kerry Patterson and Terence C. Mills, MacMillan.

Delgado M., M. Domínguez, and 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.

_____ , and I.N. Lobato (2003), "Testing the martingale difference hypothesis", *Econometric Reviews*, 22, 351-377.

_____ (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions", Econometrica, 1601-1615.

Epstein, L.G. and S.E. Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework ", Econometrica, 57, 937-969.

_____ (1991), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis", Journal of Political Economy, 99, 263-286.

Escanciano, J.C. (2006), "Goodness-of-fit Tests for Linear and Nonlinear Time Series Models", Journal of the American Statistical Association, 101, 474, 531-541.

Gregory, A., J.F. Lamarche and G. Smith (2002),. "Information-Theoretic Estimation of Preference Parameters: Macroeconomic Applications and Simulation Evidence", Journal of Econometrics, 107, 213-233.

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

Hall, P., and Heyde, C. (1980), Martingale Limit Theory and Its Application. Academic Press, New York.

Hansen, L.P. (1982), "Large-Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-1054.

_____, and K. Singleton (1982), "Generalized instrumental variables estimation of non-linear rational expectations models", Econometrica 50, pp. 1269–1286

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

Kocherlakota, N. (1990) "On the 'Discount' Factor in Growth Economies," Journal of Monetary Economics, 25, 43-47.

_____ (1996), "The Equity Premium: It's Still A Puzzle," Journal of Economic Literature, 34, 42-71.

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.

Lavergne, P. and V. Patilea (2008), "Smooth Minimum Distance Estimation and Testing in Conditional Moment Restrictions Models: Uniform in Bandwidth Theory", preprint.

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

Mehra, R. y E. Prescott (1985). "The Equity Premium: A Puzzle?" Journal of Monetary Economics, vol. 15, no. 2, pp. 145-161.

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

Shin, Y. (2008), "Semiparametric Estimation of the BoxCox Transformation Model", Econometrics Journal, 11, 517-537.

Stinchcombe, M., White, H. (1998), "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative", Econometric Theory 14 295 - 324.

Stock, J. and J. Wright (2000) "GMM with Weak Identification", Econometrica, 68, 1055-1096

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.

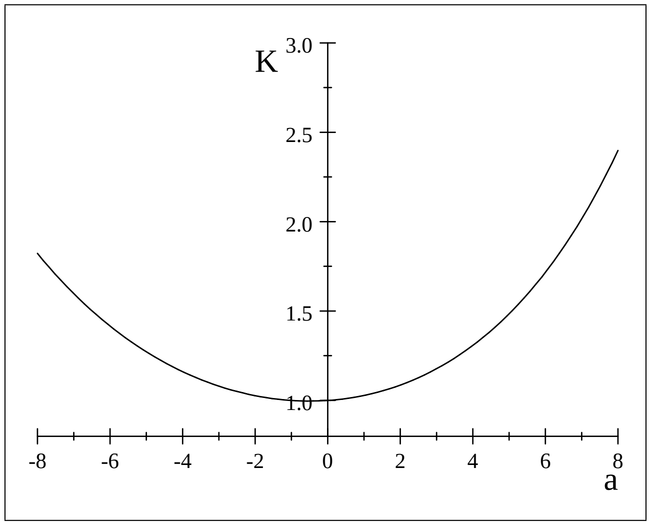

Figure 1: The function

$$0.33(0.8^{-a} + 1 + 1.2^{-a}) = K$$

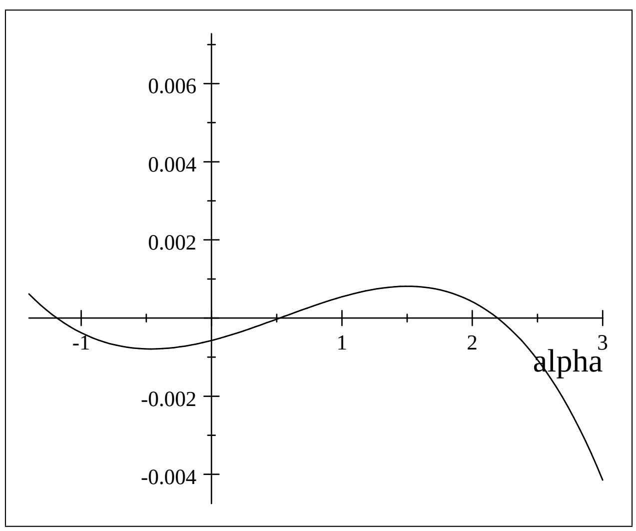

Figure 2: The UMR imposed by the optimal GMM instrument.

Table I

Bias and Variance (Var) of EGMM and CMM estimators: the asset pricing model.

	$\widehat{\alpha}_{EGMM}$		$\widehat{\alpha}_{CMM}$		$\widehat{\beta}_{EGMM}$		$\widehat{\beta}_{CMM}$	
n	Bias	Var	Bias	Var	Bias	Var	Bias	Var
				$(\alpha_0, \beta_0) = (0.5, 0.99)$				
100	.3520	4.654	-.3985	.0361	.0021	$17.03*10^{-5}$	-.0023	$119*10^{-8}$
500	.3056	2.340	-.1908	.0537	.0017	$7.798*10^{-5}$	-.0011	$177*10^{-8}$
1000	.1644	.9633	-.0949	.0383	.0009	$3.333*10^{-5}$	-.0005	$128*10^{-8}$
5000	.0163	.0138	-.0165	.0077	.0001	$45.1*10^{-8}$	-.0001	$25.4*10^{-8}$
asymptotic	0	38.76	0	38.88	0	$1.282*10^{-3}$	0	$1.286*10^{-3}$
				$(\alpha_0, \beta_0) = (1, 0.99)$				
100	.5796	10.24	-.7883	.1661	.0036	$88.2*10^{-5}$	-.0045	$54*10^{-7}$
500	.5993	7.091	-.3736	.3099	.0035	$24.7*10^{-5}$	-.0021	$104*10^{-7}$
1000	.2705	1.254	-.1858	.1585	.0016	$418*10^{-7}$	-.0011	$52*10^{-7}$
5000	.0321	.0402	-.0306	.0325	.0002	$13.2*10^{-7}$	-.0002	$10.1*10^{-7}$
asymptotic	0	154.6	0	155.1	0	$5.096*10^{-3}$	0	$5.114*10^{-3}$
				$(\alpha_0, \beta_0) = (2, 0.99)$				
100	.8233	20.18	-1.587	.5509	.0045	$79.7*10^{-5}$	-.0090	$180*10^{-7}$
500	.9954	13.08	-.7567	1.004	.0054	$53.1*10^{-5}$	-.0043	$326*10^{-7}$
1000	.4796	4.141	-.3813	.6359	.0028	$79.7*10^{-5}$	-.0022	$207*10^{-7}$
5000	.0061	.1609	-.0643	.1282	.0003	$52.7*10^{-7}$	-.0003	$40.3*10^{-7}$
asymptotic	0	615.2	0	617.1	0	$2.014*10^{-2}$	0	$2.021*10^{-2}$
				$(\alpha_0, \beta_0) = (5, 0.99)$				
100	.7107	37.17	-4.017	3.982	.0043	$151*10^{-5}$	-.0225	$124*10^{-6}$
500	1.549	31.72	-1.883	6.401	.0090	$107*10^{-5}$	-.0105	$205*10^{-6}$
1000	1.033	15.56	-.9444	3.695	.0059	$550*10^{-6}$	-.0053	$118*10^{-6}$
5000	.1678	1.026	-.1500	.7541	.0010	$33.0*10^{-6}$	-.0008	$25.2*10^{-6}$
asymptotic	0	3783	0	3795	0	0.1214	0	0.1218

Table II

Bias and Variance (Var) of EGMM and CMM estimators: the asset pricing model.

	$\widehat{\alpha}_{EGMM}$		$\widehat{\alpha}_{CMM}$		$\widehat{\beta}_{EGMM}$		$\widehat{\beta}_{CMM}$	
n	Bias	Var	Bias	Var	Bias	Var	Bias	Var
				$(\alpha_{0,}\ \beta_0) = (2, 0.99)$				
100	.8233	20.18	-1.587	.5509	.0045	$79.7*10^{-5}$	-.0090	$180*10^{-7}$
500	.9954	13.08	-.7567	1.004	.0054	$53.1*10^{-5}$	-.0043	$326*10^{-7}$
1000	.4796	4.141	-.3813	.6359	.0028	$79.7*10^{-5}$	-.0022	$207*10^{-7}$
5000	.0606	.1609	-.0643	.1282	.0003	$52.7*10^{-7}$	-.0003	$40.3*10^{-7}$
10000	.0276	.0684	-.0330	.0609	.0002	$22.4*10^{-7}$	-.0002	$19.9*10^{-7}$
20000	.0134	.0331	-.0165	.0312	.0001	$10.8*10^{-7}$	-.0001	$10.2*10^{-7}$
40000	.0061	.0162	-.0088	.0157	.0000	$5.30*10^{-7}$	-.0000	$5.16*10^{-7}$
asymptotic	0	615.2	0	617.1	0	$2.014*10^{-2}$	0	$2.021*10^{-2}$
				$(\alpha_0,\ \beta_0) = (2, 1.01)$				
100	.7351	17.70	-1.591	.6035	.0042	$653*10^{-6}$	-.0092	$204*10^{-7}$
500	.8026	9.769	-.7447	1.028	.0049	$444*10^{-6}$	-.0043	$350*10^{-7}$
1000	.4944	5.415	-.3709	.6183	.0028	$224*10^{-6}$	-.0021	$211*10^{-7}$
5000	.0699	.1568	-.0623	.1260	.0004	$53.5*10^{-7}$	-.0004	$40.7*10^{-7}$
asymptotic	0	615.2	0	617.1	0	$2.096*10^{-2}$	0	$2.104*10^{-2}$
				$(\alpha_0,\ \beta_0) = (10, 0.99)$				
100	-.6211	55.03	-7.939	15.76	-.0038	.00230	-.0434	.00046
500	1.507	47.35	-3.644	21.57	.0103	.00288	-.0199	.00066
1000	1.509	33.12	-1.835	15.60	.0086	.00106	-.0100	.00048
5000	.3040	4.108	-.3024	3.067	.0017	.00013	-.0017	.00010
asymptotic	0	14723	0	14772	0	0.4566	0	0.4583

Table III

Variance comparison between EGMM and CMM estimators: the asset pricing model.

n	$\frac{var(\widehat{\alpha}_{EGMM})}{var(\widehat{\alpha}_{CMM})}$	$\frac{var(\widehat{\beta}_{EGMM})}{var(\widehat{\beta}_{CMM})}$
100	36.63	44.28
500	13.03	16.29
1000	6.512	38.50
5000	1.255	1.308
10000	1.123	1.126
20000	1.061	1.059
40000	1.028	1.028
asymptotic	0.997	0.997

Table IV

Bias and Mean Squared Error (MSE) of EGMM and CMM estimators: the Linear model.

n	$\widehat{\alpha}_{EGMM}$		$\widehat{\alpha}_{CMM}$		$\widehat{\beta}_{EGMM}$		$\widehat{\beta}_{CMM}$	
	bias	MSE	bias	MSE	bias	MSE	bias	MSE
Homoskedastic Normal Noise								
50	.0019	.0020	.0017	.0022	-.0002	.0042	-.0005	.0058
100	-.0016	.0101	-.0010	.0111	-.0004	.0020	.0003	.0030
200	-.0013	.0050	-.0016	.0055	-.0005	.0010	-.0010	.0014
Homoskedastic Chi-Square Noise								
50	-.0034	.0207	-.0012	.0227	-.0006	.0040	.0013	.0061
100	-.0006	.0010	-.0010	.0011	.0009	.0021	.0005	.0030
200	.0022	.0050	.0020	.0054	.0001	.0011	-.0001	.0015
Heteroskedastic Noise								
50	.0001	.0379	-.0002	.0313	-.0004	.0154	-.0005	.0070
100	.0006	.0186	.0006	.0153	.0018	.0082	.0004	.0033
200	-.0005	.0092	-.0004	.0073	-.0005	.0041	-.0006	.0016

Table V

Size of the proposed T_n and Stute's test that employs the EGMM estimator ($T_n^{(2)}$). 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

Table VI

Percentage power of the proposed T_n and Stute's test that employs the EGMM estimator ($T_n^{(2)}$). Linear Model. Nominal Level is 5%.

	$T_n^{(2)}$			T_n		
n	50	100	200	50	100	200
quadratic						
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						
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

Table VII

Bias and Mean Squared Error (MSE) of EGMM and CMM estimators: the Box-Cox transformation model.

n	$\widehat{\alpha}_{EGMM}$	$\widehat{\alpha}_{CMM}$	$\widehat{\beta}_{EGMM}$	$\widehat{\beta}_{CMM}$	$\widehat{\lambda}_{EGMM}$	$\widehat{\lambda}_{CMM}$
			bias			
			Homoskedastic Normal Noise			
50	-.0032	-.0044	-.0000	-.0001	-.0003	-.0008
100	.0003	-.0007	-.0003	-.0009	.0000	-.0002
200	.0008	.0009	.0000	-.0001	.0001	.0000
			Homoskedastic Chi-Square Noise			
50	.0031	-.0038	-.0006	-.0040	-.0000	-.0012
100	-.0003	-.0003	-.0009	-.0023	-.0001	-.0006
200	.0017	.0000	-.0008	-.0007	.0001	-.0002
			Heteroskedastic Noise			
50	.0356	-.0199	-.0045	-.0242	.0054	-.0005
100	.0202	-.0098	-.0020	-.0131	.0029	-.0005
200	.0121	-.0048	-.0036	-.0065	.0016	-.0003
			MSE			
			Homoskedastic Normal Noise			
50	.0351	.0423	.0053	.0092	$150*10^{-6}$	$271*10^{-6}$
100	.0172	.0212	.0024	.0043	$62*10^{-6}$	$124*10^{-6}$
200	.0083	.0102	.0012	.0023	$28*10^{-6}$	$61*10^{-6}$
			Homoskedastic Chi-Square Noise			
50	.0360	.0403	.0054	.0082	$153*10^{-6}$	$241*10^{-6}$
100	.0172	.0208	.0025	.0043	$60*10^{-6}$	$116*10^{-6}$
200	.0083	.0102	.0011	.0023	$28*10^{-6}$	$58*10^{-6}$
			Heteroskedastic Noise			
50	.0674	.0462	.0079	.0081	$440*10^{-6}$	$511*10^{-6}$
100	.0302	.0259	.0036	.0043	$242*10^{-6}$	$300*10^{-6}$
200	.0162	.0137	.0018	.0023	$134*10^{-6}$	$160*10^{-6}$

Table VIII

Size of the proposed T_n and Stute's test that employs the EGMM estimator ($T_n^{(2)}$).

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

Table IX

Percentage power of the proposed T_n and Stute's test that employs the EGMM estimator ($T_n^{(2)}$). Box-Cox Model. Nominal Level is 5%.

	$T_n^{(2)}$			T_n		
n	50	100	200	50	100	200
			quadratic			
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			
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