UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ECONOMIA

DOCUMENTO DE TRABALHO Nº **2003/09**

Small Sample Bias of Alternative Estimation Methods for Moment Condition Models: Monte Carlo Evidence for Covariance Structures and Instrumental variables

Joaquim J. S. Ramalho
Universidade de Évora, Departamento de Economia

UNIVERSIDADE DE ÉVORA
DEPARTAMENTO DE ECONOMIA
Largo dos Colegiais, 2 – 7000-803 Évora – Portugal
Tel.: +351 266 740 894 Fax: +351 266 742 494
www.decon.uevora.pt   wp.economia@uevora.pt

Resumo/Abstract

It is now widely recognized that the most commonly used efficient two-step GMM estimator may have large bias in small samples. This problem has motivated the search for alternative estimators with better finite sample properties. Two classes of alternatives are considered in this paper. The first includes estimators which are asymptotically first-order equivalent to the GMM estimator, namely the continuous-updating, exponential tilting, and empirical likelihood estimators. Analytical and bootstrap bias-adjusted GMM estimators form the second class of alternatives. Two extensive Monte Carlo simulation studies are conducted in this paper for covariance structure and instrumental variable models. We conclude that all alternative estimators offer much reduced bias as compared to the GMM estimator, particularly the empirical likelihood and some of the bias-corrected GMM estimators analyzed.

Palavras-chave/Keyword: GMM, Continuous Updating, Empirical Likelihood, Exponential Tilting, Analytical and Bootstrap Bias-Adjusted Estimators, Covariance Structure Models, Instrumental Variables, Monte Carlo Simulation.

Classificação JEL/JEL Classification: C13, C14.

Small Sample Bias of Alternative Estimation Methods for Moment Condition Models: Monte Carlo Evidence for Covariance Structures and Instrumental Variables

Joaquim J. S. Ramalho*
Universidade de Évora and CEMAPRE

First Draft: March 2001
This revision: May 2003

Abstract

It is now widely recognized that the most commonly used efficient two-step GMM estimator may have large bias in small samples. This problem has motivated the search for alternative estimators with better finite sample properties. Two classes of alternatives are considered in this paper. The first includes estimators which are asymptotically first-order equivalent to the GMM estimator, namely the continuous-updating, exponential tilting, and empirical likelihood estimators. Analytical and bootstrap bias-adjusted GMM estimators form the second class of alternatives. Two extensive Monte Carlo simulation studies are conducted in this paper for covariance structure and instrumental variable models. We conclude that all alternative estimators offer much reduced bias as compared to the GMM estimator, particularly the empirical likelihood and some of the bias-corrected GMM estimators analyzed.

JEL Classification: C13, C14.

Keywords: GMM, Continuous Updating, Empirical Likelihood, Exponential Tilting, Analytical and Bootstrap Bias-Adjusted Estimators, Covariance Structure Models, Instrumental Variables, Monte Carlo Simulation.

*The author gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER. The work for this paper was partially completed while the author was a Ph.D. student at the University of Bristol. Helpful comments and suggestions were given by Richard Smith, Esmeralda Ramalho, and seminar participants at the University of Bristol, University of York and the 56th European Meeting of the Econometric Society, 2001, Lausanne. Address for correspondence: Joaquim J. S. Ramalho, Universidade de Évora, Departmento de Economia, 7000-803 ÉVORA, Portugal. E-mail: jsr@uevora.pt.

1 Introduction

Many econometric models are specified solely in terms of moment conditions. The efficient two-step generalized method of moments (GMM) due to Hansen (1982) is the most common approach to estimation and inference in such models. Despite its popularity, GMM suffers from some important drawbacks, the principal of them being its finite sample behaviour. In fact, it has been recognized for several years now that the first-order asymptotic distribution of the GMM estimator provides a poor approximation to its small sample distribution. There is increasing Monte Carlo simulation evidence indicating that in finite samples GMM estimators may be badly biased and the associated tests may have actual sizes substantially different from the nominal ones; see, for example, the July 1996 special issue of the *Journal of Business & Economic Statistics*.

The poor performance of the GMM estimator for the sample sizes typically encountered in economic applications has motivated the search for alternative estimators. In this paper we analyze the small sample bias of two classes of alternatives. The first contains alternative procedures which are asymptotically first-order equivalent to efficient GMM estimation, such as continuous-updating (CU) [Hansen, Heaton and Yaron (1996)], exponential tilting (ET) [Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)] and empirical likelihood (EL) [Qin and Lawless (1994) and Imbens (1997)]. The last two are the main particular cases of the minimum discrepancy (MD) estimators discussed by Corcoran (1998) and of the generalized empirical likelihood (GEL) estimators considered by Smith (1997). Analytical and bootstrap bias-adjusted GMM estimators form the second class of alternative estimators that we examine in this paper. The former were developed by Newey and Smith (2001), while alternative bootstrap methods applicable in the moment condition framework were suggested by Hall and Horowitz (1996) and Brown and Newey (2002).

Unlike the case of the GMM estimator, little is known about the finite sample bias of its alternatives. Indeed, to the best of our knowledge, although this issue has been analyzed *inter alia* by Hansen, Heaton and Yaron (1996), Horowitz (1998), Stock and Wright (2000), and Imbens (2002), all of them limited their studies to the Monte Carlo comparison between the GMM estimator and one or two particular alternatives. Thus, despite promising results reported in all cases, further investigation is still needed in order to assess those alternative estimators in the same framework. This is precisely the main aim of this paper, where we undertake two simulation studies examining the small sample behaviour of several estimators in two different settings for which there is previous evidence of the inadequate performance of the GMM estimator. In each case we investigate and compare the finite sample properties of GMM, CU, EL, ET, and various analytical and bootstrap bias-corrected versions of the GMM estimator.

This paper is organized as follows. Section 2 introduces some notation and provides a brief review of the main characteristics of GMM estimation. Section 3 discusses the two classes of alternative estimators examined in this paper. Section 4 reports the main results obtained in two Monte Carlo studies, the first dedicated to covariance structure models, the other to instrumental variable models. Section 5 concludes.

2 GMM estimation

Let y_i, $i = 1, ..., n$, be independent and identically distributed observations on a data vector y, θ a k-dimensional vector of parameters of interest and $g(y, \theta)$ an s-dimensional vector of functions of the observed variables and parameters of interest, with $s \geq k$. It is assumed that the true parameter vector θ_0 uniquely satisfies the moment conditions

$$E_F [g(y, \theta_0)] = 0, \tag{1}$$

where $E_F [\cdot]$ denotes expectation taken with respect to the unknown distribution function $F(y)$.

Define $g_i(\theta) \equiv g(y_i, \theta)$, $i = 1, ..., n$, and $g_n(\theta) \equiv n^{-1} \sum_{i=1}^n g_i(\theta)$. It is assumed that the normalized sample counterparts of the moment conditions (1), $g_n(\theta)$ and $\sqrt{n} g_n(\theta_0)$, obey, respectively, a uniform (in θ) weak law of large numbers, $g_n(\theta) \xrightarrow{p} E_F [g(y, \theta)]$, and a central limit theorem, $\sqrt{n} g_n(\theta_0) \xrightarrow{d} N(0, V)$, where the asymptotic variance matrix $V \equiv E_F \left[g_i(\theta_0) g_i(\theta_0)' \right]$ is positive definite and $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and convergence in distribution, respectively.

The efficient two-step GMM estimator $\hat{\theta}$ is obtained from minimization of the optimal quadratic form of the sample moment indicators,

$$\hat{\theta} \equiv \arg \min_{\theta} g_n(\theta) \left[V_n\left(\tilde{\theta}\right) \right]^{-1} g_n(\theta), \tag{2}$$

where $\tilde{\theta}$ is a preliminary consistent estimator for θ_0 and $\tilde{V}_n \equiv V_n\left(\tilde{\theta}\right)$ is a positive semi-definite consistent estimator for the limiting covariance matrix V, for example $\tilde{V}_n = n^{-1} \sum_{i=1}^n g_i\left(\tilde{\theta}\right) g_i\left(\tilde{\theta}\right)'$. Thus, $\hat{\theta}$ satisfies the system of first-order conditions

$$\hat{G}_n' \tilde{V}_n^{-1} \hat{g}_n = 0, \tag{3}$$

where $\hat{g}_n \equiv g_n\left(\hat{\theta}\right)$ and $\hat{G}_n \equiv \frac{\partial \hat{g}_n}{\partial \theta'}$ is a $(s \times k)$ matrix that converges almost surely and uniformly in θ to $G \equiv E_F \left[\frac{\partial g_i(\theta_0)}{\partial \theta'} \right]$, which is assumed to be full column rank. Under suitable regularity conditions, see Hansen (1982) and Newey and McFadden (1994), $\hat{\theta}$ is a consistent asymptotically normal estimator of θ_0,

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left[0, \left(G' V^{-1} G\right)^{-1}\right], \tag{4}$$

and, as shown by Chamberlain (1987), it attains the semiparametric efficiency bound in the model where all that is known are the moment conditions (1), that is, the GMM estimator is asymptotically efficient among all estimators based on (1).

Despite its asymptotic attractiveness, the GMM estimator displays some bias in small samples; see *inter alia* Tauchen (1986), Kocherlakota (1990), Ferson and Foerster (1994) and several papers published in the July 1996 special issue of the *Journal of Business & Economic Statistics* such as Altonji and Segal (1996), Andersen and Sorensen (1996), and Hansen, Heaton and Yaron (1996). Therefore, we discuss two classes of alternative estimators for moment condition models in the next section and examine their finite sample properties in section 4.

3 Alternative estimators

3.1 First-order equivalent estimators

As the significant small sample bias of the GMM estimator seems to arise from the necessity of utilizing a consistent estimate of V in an initial step, one-step estimators for moment condition models have recently been suggested. In this sub-section we describe the most popular estimators that can be included in this category. All of them are asymptotically first-order equivalent to GMM but possess different higher-order asymptotic properties. Furthermore, while the GMM estimator is not invariant to linear transformations of the original moment conditions, all one-step estimators are insensitive to how the moment restrictions are scaled. Conversely, their computation is more complicated and time-consuming.

3.1.1 Continuous-updating estimator

Similarly to GMM, the CU estimator, proposed by Hansen, Heaton and Yaron (1996), is obtained from minimization of a quadratic form of the sample moment indicators,

$$\hat{\theta} \equiv \arg \min_{\theta} g_n(\theta)' [V_n(\theta)]^{-1} g_n(\theta), \tag{5}$$

but the optimization is now performed simultaneously over the θ in the weighting matrix as well as the θ in the average sample moments.

Both Donald and Newey (2000) and Newey and Smith (2001) argue that the CU estimator should have smaller bias in finite samples than the GMM estimator. The former authors gave a jackknife interpretation of the CU estimator, demonstrating that, in the first-order conditions arising from (5), own observation terms are automatically deleted, which eliminates one known important source of bias for GMM estimators. On the other hand, Newey and Smith (2001) derived stochastic expansions for both estimators, providing asymptotic expressions for their biases. Let $g_i \equiv g_i(\theta_0)$, $G_i \equiv \frac{\partial g_i}{\partial \theta'}$, $H \equiv (G'V^{-1}G)^{-1} G'V^{-1}$, $P \equiv V^{-1} - V^{-1}G(G'V^1G)^{-1} G'V^{-1}$, $\bar{V}_{\theta_j} \equiv E\left[\frac{\partial g_i g_i'}{\partial \theta}\right]$, W denote the weight matrix used in the first-step, $H_W \equiv (G'WG)^{-1} G'W$, e_j be an k-vector whose j-element is one and the others are zero and a be an s-vector such that $a_j \equiv \frac{1}{2} tr\left\{(G'V^1G)^{-1} E_F\left[\frac{\partial^2 g_{ij}(\theta_0)}{\partial\theta\partial\theta'}\right]\right\}$, with g_{ij} denoting the jth element of $g_i(\theta)$, $j = 1, ..., s$. The asymptotic bias of the GMM estimator is given by

$$b_{gmm} = B_I + B_G + B_V + B_W, \tag{6}$$

where $B_I = \frac{1}{n} H[-a + E_F(G_i H g_i)]$, $B_G = -\frac{1}{n}(G'V^{-1}G)^{-1} E_F(G_i' P g_i)$, $B_V = \frac{1}{n} H E_F(g_i g_i' P g_i)$, and $B_W = H \sum_{j=1}^{k} \bar{V}_{\theta_j} (H_W - H)' e_j$. Each of the four terms of (7) has its own interpretation. Following Newey and Smith (2001), the first term is the asymptotic bias for the (infeasible) optimal GMM estimator based on the first-order conditions $G'V^{-1}\hat{g}_n$, where the optimal linear combination matrix $G'V^{-1}$ does not need to be estimated. The second and third terms are due to the necessity of estimating G and V in that optimal linear combination of moments, respectively. The last term arises from the choice of the first-step estimator, being zero if W is a scalar multiple of V^{-1}. On the other hand, the asymptotic bias of the CU estimator is given only by

$$b_{CU} = B_I + B_V. \tag{7}$$

Thus, the CU is affected by two less sources of bias than the GMM estimator since the terms B_G and B_W drop out. See Newey and Smith (2001) for details concerning these derivations.

There is relatively little Monte Carlo evidence on the small sample bias of the CU estimator. Indeed, to the best of our knowledge, only Hansen, Heaton and Yaron (1996) and Stock and Wright (2000) have undertaken simulation studies involving this estimator. They obtained similar conclusions, which indicate that the CU estimator is effectively approximately median unbiased but has a finite sample distribution with very fat tails, exhibiting sometimes extreme outlier behaviour. We investigate this question further in section 4.

3.1.2 Empirical likelihood and exponential tilting estimators

Using either of the two previous methods, only k linear combinations of the s sample moment conditions are in fact set equal to zero. However, it is possible to find a weighting scheme such that all moment conditions are satisfied in the sample. Consider again the moment conditions given in (1), $E_F[g(y, \theta_0)] = 0$, where the distribution $F(y)$ is unknown. Implicitly, by giving the same weight (n^{-1}) to each observation, GMM uses the empirical distribution function $F_n(y) \equiv n^{-1} \sum_{i=1}^{n} 1(y_i \le y)$ as estimate for $F(y)$, where the indicator function $1(y_i \le y)$ is equal to 1 if $y_i \le y$ and 0 otherwise. The distribution $F_n(y)$ is the nonparametric maximum likelihood estimator of $F(y)$, being the best estimator when no information about the population of interest is available. However, because the moment conditions (1) are assumed to be satisfied in the population, this information can be exploited in order to obtain a more efficient estimator of $F(y)$. Thus, we may select, as suggested firstly by Back and Brown (1993), the estimator $\hat{\theta}$ that minimizes the distance, relatively to some metric, between $F_n(y)$ and a distribution function $F_{MD}(y)$ satisfying the moment conditions (1). The distribution $F_{MD}(y)$ is, hence, the member of the class $\mathcal{F}(\theta)$ of all distribution functions that satisfy (1), $\mathcal{F}(\theta) \equiv \{F_{MD} : E_{F_{MD}}[g(y, \theta_0)] = 0\}$, that is closest to F_n.

In the selection of a particular probability measure in $\mathcal{F}(\theta)$, different metrics for the closeness between $F_{MD}(y)$ and $F_n(y)$ may be used. Let $\mathcal{M}(F_n, F_{MD})$ be the distance metric utilized. Then, any MD estimator $\hat{\theta}$ can be described as the solution to the program

$$\hat{\theta} \equiv \arg\min_{\theta} \mathcal{M}(F_n, F_{MD}), \text{ subject to } p_i^{MD} \geq 0, \sum_{i=1}^{n} p_i^{MD} = 1 \text{ and } \sum_{i=1}^{n} p_i^{MD} g_i(\theta) = 0, \quad (8)$$

where $p_i^{MD} \equiv dF_{MD}(y)$, $i = 1, ..., n$, and the last restriction is an empirical measure counterpart to the moment conditions (1), imposing them numerically in the sample. Several estimation methods based on (8), differing only in the choice of metric $\mathcal{M}(\cdot)$, have been proposed. The most common choices for $\mathcal{M}(\cdot)$ are particular cases of the Cressie-Read power-divergence statistic [Cressie and Read (1984)].[1] In this case, the computationally complicated MD optimization (8) can be replaced by a simpler one. Indeed, Newey and Smith (2001) showed that for any MD estimator based on the Cressie-Read statistic there is a dual GEL estimator.[2] GEL estimators are obtained as solution to the saddle point problem

$$\hat{\theta} \equiv \arg\min_{\theta} \sup_{\phi} \sum_{i=1}^{n} \rho\left[\phi' g_i(\theta)\right], \quad (9)$$

where $\rho(\cdot)$ is a carrier function that conveys the information provided by the moment conditions (1) and ϕ is an s-vector of auxiliary parameters that can be interpreted as Lagrange multipliers associated to the last restriction of (8). Here, we focus on the most well known special cases of GEL (and MD) estimators: ET, where $\rho(\cdot) = -e^{\phi' g_i(\theta)}$, and EL, for which $\rho(\cdot) = \ln\left[1 + \phi' g_i(\theta)\right]$.[3]

After estimating θ and ϕ in (9), the implied probabilities referred to in Back and Brown (1993), previously denoted by p_i^{MD} and from now on by p_i^{GEL}, $i = 1, ..., n$, may be estimated by calculating

[1] For a more general specification of $\mathcal{M}(\cdot)$, which includes the Cressie-Read family as a particular case, see Corcoran (1998).

[2] Newey and Smith (2001) emphasize that outside the Cressie-Read family an explicit dual relationship between MD and GEL estimators is not likely to exist. See Smith (1997) for a detailed description of GEL estimators.

[3] Note that these and other expressions presented below are slightly different from those appearing in Newey and Smith (2001) due to the normalizations imposed by them on the function $\rho(\cdot)$. However, that does not affect the GEL estimators of θ.

the ratios

$$\hat{p}_i^{GEL} \equiv p_i^{GEL}\left(\hat{\theta},\hat{\phi}\right) = \frac{\rho_1\left[\hat{\phi}' g_i\left(\hat{\theta}\right)\right]}{\sum_{i=1}^{n}\rho_1\left[\hat{\phi}' g_i\left(\hat{\theta}\right)\right]}, \tag{10}$$

where $\rho_j(v) \equiv \frac{\partial^j \rho(v)}{\partial v^j}$ $(j = 1, 2, ...)$. These $\hat{p}_i^{GEL}$ sum to one by construction and, as implied by (8), the sample moment conditions $\sum_{i=1}^{n}\hat{p}_i^{GEL} g_i\left(\hat{\theta}\right) = 0$ are numerically imposed. They are also positive when $\hat{\phi}' g_i\left(\hat{\theta}\right)$ is small uniformly in i. Thus, an efficient estimator of $F(y)$ in (1) can be obtained by calculating the so-called GEL distribution

$$\hat{F}_{GEL}(y) = \sum_{i=1}^{n}\hat{p}_i^{GEL} 1\left(y_i \leq y\right), \tag{11}$$

which means that an efficient estimator of $E_F\left[a\left(y,\theta_0\right)\right]$, for any function $a(\cdot)$, is given by $\sum_{i=1}^{n}\hat{p}_i^{GEL}$ $a\left(y,\hat{\theta}\right)$. Some of the bias-corrected estimators discussed in section 3.2 are based on a variant of these GEL implied probabilities. See Ramalho and Smith (2002a,b) for other interesting applications of the weights $\hat{p}_i^{GEL}$.

Similarly to the GMM and CU estimators, Newey and Smith (2001) derived asymptotic expressions for the bias of GEL estimators,

$$b_{GEL} = B_I + (1-\eta) B_V, \tag{12}$$

where $\eta = \frac{\rho_1(0)\rho_3(0)}{2\left[\rho_2(0)\right]^2}$ is a scalar. This expression is very similar to that presented for the CU estimator in (7), apart from the weight $(1-\eta)$.[4] Hence, like the CU estimator, GEL estimators have two less sources of bias than the GMM estimator. Furthermore, for the EL estimator the last term of (12) disappears, as $\eta = 1$.[5] Thus, the EL estimator removes the bias due to estimation of the weighting matrix in the optimal linear combination of moments. Its bias is then the same as for the (infeasible) GMM estimator based on the optimal linear combination of moment conditions. With regard to the ET estimator, $\eta = \frac{1}{2}$, so the bias term B_V is halved relatively to the CU estimator.

Although the GEL formulation is simpler than the MD one, the computation of GEL estimators is still not straightforward since $\rho(\cdot)$ in (9) is a saddle function. This seems to be the main reason why so little evidence about the finite sample performance of EL and ET estimators has been reported so far. To the best of our knowledge, only Imbens (2002) has examined this issue. Furthermore, they have not been used in empirical work yet. Recently, Imbens (2002) has suggested three alternative procedures that simplify the computation of GEL estimators. In the

[4] In fact, Newey and Smith (2001) show that the CU estimator can also be interpreted as a member of the class of GEL estimators.

[5] Note that $\rho_1(0) = 1$, $\rho_2(0) = -1$ and $\rho_3(0) = 2$ for the EL estimator and $\rho_1(0) = \rho_2(0) = \rho_3(0) = -1$ in the ET case.

Monte Carlo experiments undertaken in this paper we follow his penalty approach. Although very time-consuming, the estimates obtained appear to be very reliable.

3.2 Bias-corrected GMM estimators

Bias-corrected GMM estimators constitute the second class of alternatives to standard GMM estimation of moment condition models that we consider in this paper. The bias of the GMM estimator $\hat{\theta}_{GMM}$ may be defined as

$$b(\theta_0) = E_F\left(\hat{\theta}_{GMM} - \theta_0\right). \tag{13}$$

If we are able to estimate $b(\theta_0)$, we can obtain a bias-corrected GMM estimator $\hat{\theta}_{BCGMM}$ by calculating

$$\hat{\theta}_{BCGMM} = \hat{\theta}_{GMM} - \hat{b}, \tag{14}$$

where $\hat{b}$ denotes the estimated bias. There are several approaches to bias correction. In this section we analyze the ability of both analytical and bootstrap methods to estimate the bias (13) and obtain bias-corrected GMM estimators with attractive finite sample properties.

3.2.1 Analytical bias-corrected GMM estimators

Since an asymptotic bias formula for the GMM estimator is already available, see (6), the utilization of analytical methods is computationally the simplest way of obtaining bias-adjusted GMM estimators. Indeed, all we need to do is evaluate (6) at the GMM estimator, using a consistent estimator of the distribution function $F(y)$ to estimate the expectations present in that expression, and then calculate (14).

In this paper we consider two alternative estimators for $F(y)$. One is the empirical distribution $F_n(y)$ which weights equally all functions of each observation i by n^{-1}, $i = 1, ..., n$. The other is a variant of the GEL distribution function given in (11), namely that suggested by Brown and Newey (2002). Such distribution, which we call here first-stage GEL (FSGEL) distribution, is obtained as follows. First, the objective GEL function $\sum_{i=1}^{n} \rho\left[\phi' g_i(\theta)\right]$, see (9), is optimized in order only to ϕ, keeping $\theta = \hat{\theta}_{GMM}$. Then, the resulting estimators, $\hat{\phi}_{FSGEL}$, are used to obtain the FSGEL distribution

$$F_{FSGEL}(y) = \sum_{i=1}^{n} \hat{p}_i^{FSGEL} 1(y_i \leq y), \tag{15}$$

where the probabilities $\hat{p}_i^{FSGEL} \equiv p_i\left(\hat{\theta}_{GMM}, \hat{\phi}_{FSGEL}\right)$ are calculated as in (10). We denote by ABCa the analytical bias-corrected GMM estimator based on $F_n(y)$ and by ABCb the one based on $F_{FSGEL}(y)$.

To the best of our knowledge, no evidence about the ability of both approaches to reduce the bias of the GMM estimator has been provided so far.

3.2.2 Bootstrap bias-corrected GMM estimators

Alternatively, the bias of the GMM estimator can be estimated using the bootstrap. Assume that a random sample S of size n is collected from a population whose (unknown) cumulative distribution function is $F(y)$. Bootstrap samples are generated by randomly sampling the original data with replacement. This resampling is based on a certain cumulative distribution function, $F^*(y)$, which assigns each observation a given probability of being sampled. The bias (13) can be estimated as follows:

1. Compute $\hat{\theta}_{GMM}$ accordingly to (2) using the original data;
2. Generate B bootstrap samples S_j^*, $j = 1, ..., B$, of size n by sampling the original data randomly with replacement accordingly with the chosen distribution function $F^*(y)$:
$$S_j^* = \left\{ y_{j1}^*, ..., y_{jn}^* \right\},$$
where y_{ji}^*, $i = 1, ..., n$, denotes the observations included in the bootstrap sample S_j^*;
3. For each bootstrap sample calculate the GMM estimator $\hat{\theta}_j^*$:
$$\hat{\theta}_j^* \equiv \arg\min_{\theta} g_{jn}^*(\theta)\, \tilde{V}_{jn}^{*-1} g_{jn}^*(\theta)\,,\ j = 1, ...B,$$
where $g_{jn}^*(\theta) = n^{-1}\sum_{i=1}^{n} g\left(y_{ji}^*, \theta\right)$ and $\tilde{V}_{jn}^{*-1}$ is evaluated at $\tilde{\theta}_j^*$, a preliminary consistent estimator for θ_0 based on the bootstrap sample S_j^*;
4. Average the B GMM estimators calculated in the preceding step:
$$\bar{\theta}^* = \frac{1}{B}\sum_{j=1}^{B} \hat{\theta}_j^*;$$
5. Estimate the bias of the GMM estimator $\hat{\theta}$ by calculating:
$$\hat{b} = \bar{\theta}^* - \hat{\theta}_{GMM}. \tag{16}$$

Subtracting the bias (16) from the GMM estimator $\hat{\theta}_{GMM}$, it is then possible to obtain the bias-corrected GMM estimator defined in (14):

$$\hat{\theta}_{BCGMM} = 2\hat{\theta}_{GMM} - \bar{\theta}^*. \tag{17}$$

This general procedure to obtain bootstrap estimators may be implemented in several distinct forms. In this paper we consider three alternatives. The first is the so-called nonparametric (NP) bootstrap, where the resampling is based on the empirical distribution function, $F^*(y) = F_n(y)$, so each observation has equal probability n^{-1} of being drawn. Although this is the most commonly applied bootstrap technique in econometrics, its direct application in the GMM framework seems to be unsatisfactory in many cases. Indeed, when the model is overidentified, while the population moment conditions $E_F[g(y,\theta)] = 0$ are satisfied at $\theta = \theta_0$, the estimated sample moments are typically non-zero, that is, there is no θ such that $E_{F_n}[g(y,\theta)] = 0$ is met, except in very special cases. Therefore, $F_n(y)$ may be a poor approximation to the true underlying distribution of the data and, hence, the NP bootstrap may not yield a substantial improvement over first-order asymptotic theory in standard applications of GMM.[6]

As the key factor to successful application of bootstrap techniques in the GMM context seems to require the satisfaction of a bootstrap version of the population moment conditions, Brown and Newey (2002) suggested looking for a different resampling distribution, say $F_1(y)$, such that $E_{F_1}[g(y,\theta)] = 0$ for $\theta = \hat{\theta}_{GMM}$. Namely, instead of using the empirical distribution to resample the original data, Brown and Newey (2002) proposed the employment of the FSGEL distribution discussed in the previous sub-section, which assigns each observation a different probability of being drawn. In fact, since $\sum_{i=1}^{n} \hat{p}_i^{FSGEL} g_i\left(\hat{\theta}_{GMM}\right) = 0$ is the first-order condition of the FSGEL optimization problem, see (9) and (10), this FSGEL bootstrap imposes the moment conditions, evaluated at the GMM estimator $\hat{\theta}_{GMM}$, on the sample: $E_{FSGEL}\left[g\left(y, \hat{\theta}_{GMM}\right)\right] = 0$. Furthermore, Brown and Newey (2002) proved that the FSGEL bootstrap is asymptotically efficient relative to any bootstrap based on the empirical distribution function.

The other bootstrap method developed specifically for moment condition models was proposed by Hall and Horowitz (1996). These authors suggested keeping $F_n(y)$ as resampling distribution and, instead, replacing the moment indicators $g(y,\theta)$ used in the GMM estimation criterion (2) by the recentered moment indicators:

$$g^c\left(y_j^*, \theta\right) = g\left(y_j^*, \theta\right) - \frac{1}{n}\sum_{i=1}^{n} g_i\left(\hat{\theta}_{GMM}\right), \; j = 1, ..., B. \tag{18}$$

Clearly, as $n^{-1}\sum_{i=1}^{n} g_i\left(\hat{\theta}_{GMM}\right) = E_{F_n}\left[g\left(y, \hat{\theta}_{GMM}\right)\right]$, this recentering guarantees that the expectation of the modified moment indicators $g^c(\cdot)$ with respect to the empirical distribution is zero,

[6] However, Hahn (1996) demonstrated theoretically that the NP bootstrap distribution of any GMM estimator converges weakly to the limit distribution of the estimator. According to this author, the arguments against the use of the NP bootstrap in the moment condition context apply to Hansen's (1982) J test of overidentifying moment conditions, not to the GMM estimator itself. Hence, we decided to include the analysis of the bias of the NP bootstrap GMM estimator in the two Monte Carlo experiments that we conduct in section 4, investigating whether or not it behaves better than simple GMM estimators in finite samples and how it performs comparatively with the more refined bootstrap methods discussed below.

$E_{F_n}\left[g^c\left(y_j^*,\theta\right)\right]=0$. To implement this recentered NP (RNP) bootstrap method some adaptations must be made to the general procedures described earlier. Namely, in step 1 we have to calculate also $g_n\left(\hat{\theta}_{GMM}\right)$ and in step 3 GMM estimation is now based on the recentered moment indicators (18), including that used to obtain the preliminary estimator $\tilde{\theta}_j^*$, and $\tilde{V}_{jn}^*=n^{-1}\sum_{i=1}^n g^c\left(y_{ji}^*,\tilde{\theta}_j^*\right)g^c\left(y_{ji}^*,\tilde{\theta}_j^*\right)'$, $j=1,...,B$. A Monte Carlo study by Horowitz (1998) showed that the RNP bootstrap is able to reduce significantly the bias of the GMM estimator in some cases. To the best of our knowledge, no simulation experiments examining the ability of the other bootstrap methods to estimate the bias of the GMM estimator have been realized.

4 Monte Carlo simulation

In this section we conduct two Monte Carlo simulation studies, the first concerning models of covariance structures, the other dedicated to instrumental variable models. In both cases our main aim is the analysis of the small sample bias of all the alternative estimators for moment condition models discussed throughout this paper. Therefore, for each estimator we report the estimated mean and median bias, standard error (SE), root mean squared error (RMSE), median absolute error (MAE) and the 0.05 and 0.95 quantiles of its Monte Carlo distribution. In each experiment, 1000 Monte Carlo replications of samples of both 100 and 500 observations were generated. All bootstrap estimators were based on 100 bootstrap samples in each Monte Carlo replication. The estimators based on the FSGEL distribution were implemented using the EL criterion function.

4.1 Covariance structure models

Covariance structure models are important in the analysis of a variety of economic processes. Basically, they are employed to model the serial correlation structure of one economic variable in longitudinal data or the relation between movements in different economic variables (such as earnings and hours changes) over time. For applications involving these models see, for example, Abowd and Card (1987, 1989), Behrman, Rozenzweig and Taubman (1994), Griliches (1979) and Hall and Mishkin (1982). Altonji and Segal (1996) carried out an extensive Monte Carlo analysis of the finite sample properties of the efficient GMM estimator in this framework and found that this estimator is severely downward biased in small samples for most data distributions and in relatively large samples for 'badly behaved' distributions. They explain this poor performance as due to the correlation between the estimated second moments used to construct the moment indicators and the sampling optimal weighting matrix. Indeed, as they argued, moment conditions consisting of second moments are likely to be highly correlated with their covariance matrix "because individual observations that increase the sample estimate of a variance will also tend to increase the sample estimate of the variance of the variance" [Altonji and Segal (1996), p. 356]. In a similar study, Horowitz (1998) showed that for $n = 500$ the RNP bootstrap GMM estimator, although also biased in some cases, offers much reduced bias as compared to the standard GMM estimator.

4.1.1 Experimental design

Our first simulation study is based on one of the experimental designs analyzed by Altonji and Segal (1996). We consider a setting where the objective is the estimation of a common population variance θ_0 for a scalar random variable y_t, $t = 1, ..., T$, from observations on a balanced panel of individuals covering $T = 10$ time periods. We assume a common population mean $E(y_t) = 0$ and that y_{ti} is independent over $t = 1, ..., T$ and $i = 1, ..., n$. Thus, for each time period, the variance of the observations can be computed using the standard unbiased estimator

$$m_t(y_t) = \frac{1}{n-1}\sum_{i=1}^{n} y_{ti}^2,\ t = 1, ..., 10. \tag{19}$$

These estimates of the second moments are stacked into a 10-dimensional vector, $m(y) = [m_1(y_1),$ $..., m_{10}(y_{10})]'$, where $y = (y_1, ..., y_{10})'$, and are related to the population variance θ_0 through the 10-vector of moment conditions

$$E[g(y, \theta_0)] = E[m(y) - \iota\theta_0] = 0, \tag{20}$$

where ι is a 10-vector of ones.

In this Monte Carlo study, all samples are generated in a way that ensures that the data are independent across t and i. Five different distributions for y_t, scaled to have mean 0 and variance $\theta_0 = 1$, are considered. Although the elements of $m(y)$ are independent, this information is ignored in the estimation of the covariance matrix $\tilde{V}_n = \sum_{i=1}^{n} g_i\left(\tilde{\theta}\right) g_i\left(\tilde{\theta}\right)'$, where

$$g_i(\theta) = m_i(y_i) - \iota\theta = \frac{n}{n-1} y_i^2 - \iota\theta. \tag{21}$$

In this framework, it is straightforward to show that the GMM estimator is given by

$$\hat{\theta}_{GMM} = w\sum_{i=1}^{n}\frac{1}{n} m_i(y_i), \tag{22}$$

where $w = \left(\iota'\tilde{V}_n^{-1}\iota\right)^{-1}\iota'\tilde{V}_n^{-1}$, while GEL estimators may be expressed as

$$\hat{\theta}_{GEL} = \frac{\iota'}{10}\sum_{i=1}^{n}\hat{p}_i^{GEL} m_i(y_i). \tag{23}$$

Thus, over i the GMM estimator ascribes equal weights whereas GEL applies the GEL implied probabilities. Over t, GMM assigns distinct weights, given by the vector w, while for GEL each time period receives an equal weight. Note that in (22) we have evaluated $\tilde{V}_n$ at a non-efficient GMM estimator $\tilde{\theta}$ using the identity as weighting matrix, in which case identical weights were assigned over both i and t:

$$\tilde{\theta} = \frac{\iota'}{10 * n} \sum_{i=1}^{n} m_i (y_i) . \tag{24}$$

4.1.2 Results

Tables 1 and 2 report the results obtained for GMM and its three asymptotically first-order equivalent alternatives for $n = 100$ and 500, respectively. The results obtained for the GMM estimator are very similar to those presented by Altonji and Segal (1996). As in their study, this estimator is clearly downward biased. This distortion is particularly marked for 'badly-behaved' distributions, namely thicker-tailed symmetric (t_5) and long-tailed skewed (lognormal and exponential) distributions. The worst case is given by the lognormal distribution, where the mean and median biases for $n = 100$ are, respectively, 41.9% and 43.5% and the empirical 90% confidence interval does not cover the true value $\theta_0 = 1$. Increasing the sample size to 500 significantly improves the estimation but, for the aforementioned distributions, the GMM estimator still displays substantial bias.

Table 1 about here

Table 2 about here

While all estimators exhibit similar SE, the improvement for ET and EL in terms of both mean and median bias, RMSE and MAE is clear, mainly in the latter case. Indeed, relative to the GMM estimator, for $n = 100$ the mean bias of the EL estimator is less between 20.3% (lognormal) and 84.6% (normal), the median bias between 19.5% (lognormal) and 87.5% (normal), the RMSE between 9.3% (normal) and 32.2% (exponential) and the MAE between 10.5% (normal) and 43.5% (exponential). For the ET estimator, the improvements over GMM are much more modest. On the other hand, the results for the CU estimator are worse than those for the GMM estimator for $n = 100$ and very similar for $n = 500$. Thus, whichever data distribution is considered, the best is the EL estimator, followed by ET, GMM and, finally, CU, as is also clearly visible in Figure 1, which shows the sampling cumulative density functions for the four estimators for the $n = 100$ case. A theoretical explanation for this small sample behaviour arises from the bias functions derived by Newey and Smith (2001). In fact, since $G = -\iota$ in this example, from (6), (7) and (12) it follows that $b_{GMM} = b_{CU} = 2b_{ET} = B_V$ and $b_{EL} = 0$. Although these bias values do not correspond exactly to the Monte Carlo results we achieved, they reflect the hierarchy we found for the four estimators.

Figure 1 about here

Tables 3 and 4 report the results obtained for bias-corrected GMM estimators (see also Figure 2). In general, all analytical and bootstrap GMM estimators substantially reduce the bias of the GMM estimator at the expense of a rather modest increase in their SE. Indeed, the gain from bias reduction outweighs the increased contribution of SE to RMSE in almost all cases. The behaviour of these estimators is not uniform, however. The improvements are much less significative for the NP bootstrap, as expected. The RNP and FSGEL bootstrap methods are the best in terms of bias but the former estimator exhibits too much variability in the lognormal and, only for $n = 100$, t_5 cases. On the other hand, ABCb tends to dominate the other analytical bias-corrected GMM estimator according to all criteria. Relative to the RNP and FSGEL bootstrap techniques, the bias performance of the ABCb estimator is slightly inferior but its RMSE is the best in many cases.

Table 3 about here

Table 4 about here

Figure 2 about here

Comparing the results obtained for the two classes of estimators, we see that EL is the only serious competitor for the best bias-corrected GMM estimators analyzed, namely for $n = 500$, where its bias is similar and its RMSE is less. For $n = 100$, the bias of EL is in general larger but its RMSE is similar.

4.2 Instrumental variable models

In this second Monte Carlo investigation we consider instrumental variable models, one of the most widely spread applications of GMM. There are numerous studies showing that, in small samples, GMM estimators are not unbiased, especially when the number of instruments is large [e.g. Tauchen (1986), Kocherlakota (1990), and Anderson and Sorenson (1996)] or the correlation between regressors and instruments is weak [e.g. Nelson and Startz (1990) and Bound, Jaeger and Baker (1995)]. In this section we present additional evidence confirming those results and examine how the other alternative estimation methods under analysis perform in this framework.

4.2.1 Data generating process

Consider the linear model described by equation

$$y = X\theta_0 + u, \tag{25}$$

where y and X are n-vectors of observations on a dependent and an explanatory variable, respectively, and u is a n-vector of normal errors with mean zero and variance one. Analogously to Nelson and Startz (1990), we generate the regressor X and the s instruments Z_j, $j = 1, ..., s$, that constitute the matrix of instruments Z from

$$X = \lambda u + \epsilon \tag{26}$$

and

$$Z_j = \gamma_j \epsilon + v_j,\ j = 1, ..., s, \tag{27}$$

where ϵ and v_j are random disturbances independently generated from a $N(0, I)$ distribution and λ and γ_j are fixed parameters that allow the correlations ρ_{xu} between X and u and ρ_{xz_j} between X and the instrument Z_j to be controlled according to equations

$$\lambda = \frac{\rho_{xu}}{\sqrt{1-\rho_{xu}^2}} \tag{28}$$

and

$$\gamma_j = \rho_{xz_j}\sqrt{\frac{1+\lambda^2}{1-\left(1+\lambda^2\right)\rho_{xz_j}^2}}. \tag{29}$$

Five different experiments were performed, as described in Table 5. In the two first experiments only one of the instruments utilized in estimation (Z_1) is not worthless. However, while in experiment 1 there is a single overidentifying moment condition, in the second case (and all the others) the number of instruments is large relative to the number of regressors. Experiment 3 investigates the effects of increasing the correlation between the explanatory variable and the instrument Z_1. Experiment 4 examines the consequences of lower feedbacks from u to X in equation (25), an effect which is not usually analyzed [the only exception seems to be Blomquist and Dahlberg (1999)] but, as Nelson and Startz (1990) implicitly acknowledge, the correlation ρ_{xu} is one of the most important determinants of the accuracy with which an instrumental variable model may be estimated, because high feedbacks from u to X make the model poorly identified even when the correlation between regressors and instruments is relatively important. Finally, in experiment 5, we repeat experiment 2 but now the additional nine instruments utilized convey information about X. In all experiments the parameter θ_0 was fixed at 1 and in the construction of the bootstrap samples we resampled with replacement from the original (y, X, Z) sample.

Table 5 about here

In each experiment, as we ignore the homoskedasticity assumption, the GMM estimator is given by

$$\hat{\theta}_{GMM} = \left(X'Z\tilde{V}_n^{-1}Z'X\right)^{-1} X'Z\tilde{V}_n^{-1}Z'y,$$

while GEL estimators can be expressed as

$$\hat{\theta}_{GEL} = \left(X'\hat{\Pi}Z\tilde{V}_n^{-1}Z'\hat{\Pi}X \right)^{-1} X'\hat{\Pi}Z\tilde{V}_n^{-1}Z'\hat{\Pi}y,$$

where $\hat{\Pi}$ is a $(n \times n)$ diagonal matrix with typical element $\hat{p}_i^{gel}$, $i = 1, ..., n$. Comparing the two expressions, we see that the difference between these estimators results from the weights applied to the $Z'X$ and $Z'y$ matrices: the GMM estimator applies unit weights whereas GEL estimators weight each component of those matrices using the GEL implied probabilities.

4.2.2 Results

Table 6 reports the results obtained for GMM, CU, ET and EL estimators for $n = 100$. In Figure 3 we show also their cumulative distribution functions. Similarly to the results widely reported by other simulation studies, the GMM estimator is significantly biased in all experiments. Its best (least bad) performance in terms of bias occurs when only two instruments are used (experiment 1), precisely the case where it exhibits more dispersion, which reflects the traditional trade-off between bias and efficiency that usually happens when the number of moment conditions is increased and the GMM estimator is employed. Note that this effect occurs not only when the nine instruments added are useless (experiments 2-4) but, surprisingly, also in experiment 5, where each one of the new instruments has the same correlation with X as the instrument Z_1 in experiment 1. Notice also that the decrease in the dispersion of the GMM estimator when new instruments are added is such that its RMSE is substantially lower in experiments 2-5.

Table 6 about here

Figure 3 about here

The bias of GMM is particularly significant in experiment 2, where this method clearly overestimates the parameter θ_0, producing estimates greater than 1, the true value of θ_0, in 95.5% of the replications realized. In experiment 3 the GMM estimator still presents a substantial bias but, due to the higher correlation between Z_1 and X, there is an important improvement in its small sample properties. In fact, although 10 instruments are still worthless, the mean bias of the GMM estimator is reduced by 74.4% and its standard error by 36.8% by merely increasing ρ_{xz_1} from 0.3 to 0.7. With regard to the feedback from u to X in equation (25), analyzed in experiment 4, its decrease seems to have two distinct consequences for the GMM estimator. On the one hand, its bias diminishes considerably, which was expected because, although the correlation between Z_1 and X is still 0.7, the component of the regressor not correlated with the error term now has a higher influence over the behaviour of the dependent variable.[7] On the other hand, there is an

[7] See also the bias expressions presented in footnote 8.

increase in its dispersion, probably due to the higher variability of y, which in turn results directly from X and u being less dependent. Finally, the results obtained in experiment 5, although, as expected, better than those achieved for experiment 2, are worse than those of experiment 3, which emphasizes the importance of high correlations between instruments and regressors in this framework. Indeed, despite the existence of 10 useless instruments in experiment 3 and only 1 in experiment 5, the presence of a single good instrument in the former case is sufficient for better results than those obtained when 10 reasonable instruments are used in the latter.

Unlike the previous simulation study, the CU, ET and EL estimators now exhibit a very similar behaviour in most experiments (the exception is the second one), as can be immediately seen from Figure 3, where their sampling cumulative density functions are almost indistinguishable. This happens because, in this case of moments consisting of products of instruments with a Gaussian residual, the third moments of g_i are zero, so the bias term B_V of (7) and (12) disappears and, hence, the asymptotic biases of the three estimators become equal.[8] Furthermore, the three estimators are nearly median unbiased in all the cases considered (again, the exception is experiment 2, where the CU estimator displays some bias). However, for the poorest identified models (experiments 1 and 2), the Monte Carlo distributions of their estimators are quite disperse, having very heavy left tails. These results conform with those obtained by Hansen, Heaton and Yaron (1996), which showed that the criterion function for the CU estimator can sometimes lead to extreme outliers for $\hat{\theta}$ but that, in general, this estimator will be median unbiased [see also the results reported by Stock and Wright (2000)].[9] By increasing the correlation between instruments and regressors, much more concentrated sampling distributions for these three estimators are obtained, without extreme values. For this reason, only small mean biases are present in experiments 3-5, substantially less than that of the GMM estimator. However, even in these cases, the GMM estimator possesses the least RMSE.

Table 7 presents the results for $n = 500$. There is a significant improvement in the properties of all estimation methods but various points should be noted. First, even for this sample size, the GMM estimator exhibits important biases, particularly in experiment 2. Thus, it seems that it would be necessary to dramatically increase the number of observations to avoid this. Second, the CU, ET and EL estimators appear even more similar. Their variability is much less for this

[8] In fact, in this case $b_{CU} = b_{ET} = b_{EL} = -\frac{1}{n}\frac{\sigma_{xu}}{\sigma_u^2}\left(G'V^{-1}G\right)^{-1}$, while $b_{GMM} = \frac{1}{n}(s-2)\frac{\sigma_{xu}}{\sigma_u^2}\left(G'V^{-1}G\right)^{-1}$, where $\sigma_{xu} = E\left(Xu|\, Z\right)$ and $\sigma_u^2 = E\left(u^2\middle|\, Z\right)$. Note that while the bias of the latter estimator increases linearly with the number of moment conditions, the biases of the others do not depend on it, as our simulation results confirm.

[9] Note that the median bias is more appropriate than the mean bias to assess the performance of the CU estimator because, in this example, it coincides with the limited information maximum likelihood estimator, which is known to have no finite moments [see *inter alia* Mariano (1982)].

sample size, so they are now also approximately mean unbiased in all cases and present in general less RMSE than the GMM estimator. Comparing the results obtained for experiments 1 and 2, we can confirm that these methods are relatively indifferent to the addition of worthless instruments, unlike the GMM estimator that continues to present the habitual trade-off between bias and efficiency.

Table 7 about here

With regard to the bias-corrected GMM estimators, Figure 4 clearly shows their relatively uniform performance, see also Tables 8 and 9.[10] In the first two experiments, which concern the poorest identified models, the behaviour of all bias-corrected estimators was not particularly promising. In the first case, they produced similar biases to the GMM estimator itself and the sampling distributions of the bootstrap estimators are much more variable. In the second case, although they cut the median bias of the GMM estimator by about 30-40%, the bias is still very high (around 16-20%). However, their behaviour improves substantially in the remaining experiments and for $n = 500$, where all analytical and bootstrap bias-corrected estimators yielded encouraging results. Apart from experiment 1, the analytical methods performed slightly better in terms of bias and slightly worse in terms of dispersion than bootstrap estimators. Relative to the other class of alternative estimators, the bias-corrected methods are less efficient in the remotion of the bias of the GMM estimator but, due to their lower variability, exhibit less RMSE in most cases.

Table 8 about here

Table 9 about here

Figure 4 about here

5 Conclusion

In this paper we investigated through some Monte Carlo experiments the finite sample properties of various methods which are theoretically appropriate for the estimation of moment condition models. Two different settings, where GMM is known to produce biased estimators, were considered. Clearly, our results showed that, in general, all the alternatives analyzed are better than GMM to estimate both covariance structure (the exception is CU) and instrumental variable models. Although no estimator seems to fully dominate the others, we found that, overall, in the first

[10] In the calculation of the analytical bias-adjusted GMM estimators we estimated also the bias term B_V, since the information that $B_V = 0$ is usually unknown in empirical work.

class of alternatives EL seems to be the best, while in the second the NP bootstrap produced results clearly inferior to the other bias-corrected estimators.

In covariance structure models, the FSGEL bootstrap produced the best results, leading to the least biased estimators in most cases and sharing with the EL method the best performance according to the RMSE criterion. For instrumental variable models, ET and EL estimators appeared to be nearly median unbiased in all cases and also mean unbiased for larger sample sizes. However, in poorly identified models, they exhibited great variability which suggests that some care must be taken in their application in small samples and when there are doubts about the quality of the instruments. In those cases, any of the bias-corrected GMM estimators simulated is an interesting alternative, since their RMSE in much less.

A natural extension of the investigation undertaken in this paper is the study of the finite sample properties of analytical bias-corrected CU, ET and EL estimators, which can be based on the bias expressions deduced by Newey and Smith (2001). Another potential avenue for future research is the analysis of the ability of the bootstrap to eliminate the bias of those three estimators, which is, however, a formidable task, requiring a great deal of computing time. Most of all, we hope that the results found in this paper help to motivate the utilization of these alternative estimation methods in applied work, since the increased computational burden is largely compensated by the achievement of estimators with better finite sample properties.

References

Abowd, J. M. and Card, D. (1987), "Intertemporal labor supply and long-term employment contracts", *American Economic Review*, 77, 50-68.

Abowd, J. M. and Card, D. (1989), "On the covariance structure of earnings and hours changes", *Econometrica*, 57, 411-445.

Altonji, J. G. and Segal, L. M. (1996), "Small-sample bias in GMM estimation of covariance structures", *Journal of Business & Economic Statistics*, 14(3), 353-365.

Andersen, T. G. and Sorensen, B. E. (1996), "GMM estimation of a stochastic volatility model: a Monte Carlo study", *Journal of Business & Economic Statistics*, 14(3), 328-352.

Back, K. and Brown, D. P. (1993), "Implied probabilities in GMM estimators", *Econometrica*, 61(4), 971-975.

Behrman, J., Rosenzweig, M. and Taubman, P. (1994), "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment", *Journal of Political Economy*, 102, 1131-1174.

Blomquist, S. and Dahlberg, M. (1999), "Small sample properties of LIML and Jackknife IV estimators: experiments with weak instruments", *Journal of Applied Econometrics*, 14, 69-88.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistical Association*, 90(430), 443-450.

Brown, B. W. and Newey, W. K. (2002), "Generalized method of moments, efficient bootstrapping, and improved inference", *Journal of Business & Economic Statistics*, 20(4), 507-517.

Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions", *Journal of Econometrics*, 34, 305-334.

Corcoran, S. A. (1998), "Bartlett adjustment of empirical discrepancy statistics", *Biometrika*, 85(4), 967-.

Cressie, N. and Read, T. R. C. (1984), "Multinomial goodness-of-fit tests", *Journal of the Royal Statistical Society, Series B*, 46(3), 440-464.

Donald, S. G. and Newey, W. K. (2000), "A Jackknife interpretation of the continuous updating estimator", *Economics Letters*, 67, 239-243.

Ferson, W. E. and Foerster, S. R. (1994), "Finite sample properties of the generalised method of moments in tests of conditional asset pricing models", *Journal of Financial Economics*, 36(1), 29-55.

Griliches, Z. (1979), "Sibling models and data in economics: beginnings of a survey", *Journal* *of Political Economy*, 87, S37-S64.

Hahn, J. (1996), "A note on bootstrapping generalised method of moments estimators", *Econometric Theory*, 12(1), 187-197.

Hall, P. and Horowitz, J. L. (1996), "Bootstrap critical values for tests based on generalised-method-of-moments estimators", *Econometrica*, 64(4), 891-916.

Hall, R. E. and Mishkin, F. (1982), "The sensitivity of consumption to transitive income: estimates from panel data on households", *Econometrica*, 50, 461-481.

Hansen, L. P. (1982), "Large sample properties of generalised method of moments estimators", *Econometrica*, 50(4), 1029-1054.

Hansen, L. P., Heaton, J. and Yaron, A. (1996), "Finite-sample properties of some alternative GMM estimators", *Journal of Business & Economic Statistics*, 14(3), 262-280.

Horowitz, J. L. (1998), "Bootstrap methods for covariance structures", *The Journal of Human Resources*, 33(1), 39-61.

Imbens, G. W. (1997), "Generalized method of moments and empirical likelihood", *Journal of Business & Economic Statistics*, 20(4), 493-506.

Imbens, G. W. (2002), "One-step estimators for over-identified generalised method of moments models", *Review of Economic Studies*, 64, 359-383.

Imbens, G. W., Spady, R. H. and Johnson, P. (1998), "Information theoretic approaches to inference in moment condition models", *Econometrica*, 66(2), 333-357.

Kitamura, Y. and Stutzer, M. (1997), "An information-theoretic alternative to generalised method of moments estimation", *Econometrica*, 65(4), 861-874.

Kocherlakota, N. R. (1990), "On tests of representative consumer asset pricing models", *Journal of Monetary Economics*, 26, 285-304.

Mariano, R. S. (1982), "Analytical small-sample distribution theory in econometrics: the simultaneous-equations case", *International Economic Review*, 23(3), 503-533.

Nelson, C. R. and Startz, R. (1990), "The distribution of the instrumental variables estimator and its t-statistic when the instrument is a poor one", *Journal of Business*, 63(1), S125-S140.

Newey, W. K. and McFadden, D. (1994), "Large sample estimation and hypothesis testing", in Engle, R. F. and McFadden, D. L. (eds.), *Handbook of Econometrics, Vol. 4*, Elsevier Science Publishers, 2111-2245.

Newey, W. K. and Smith, R. J. (2001), "Higher order properties of GMM and generalized empirical likelihood estimators", mimeo.

Qin, J. and Lawless, J. (1994), "Empirical likelihood and general estimating equations", *Annals of Statistics*, 22(1), 300-325.

Ramalho, J. J. S. and Smith, R. J. (2002a), "Generalized empirical likelihood non-nested tests", *Journal of Econometrics*, 107, 99-125.

Ramalho, J. J. S. and Smith, R. J. (2002b), "Generalized empirical likelihood Pearson-type specification tests", mimeo.

Smith, R. J. (1997), "Alternative semi-parametric likelihood approaches to generalised method of moments estimation", *Economic Journal*, 107(441), 503-519.

Stock, J. H. and Wright, J. H. (2000), "GMM with weak identification", *Econometrica*, 68(5), 1055-1096.

Tauchen, G. (1986), "Statistical properties of generalised method of moments estimators of structural parameters obtained from financial market data", *Journal of Business & Economic Statistics*, 4(4), 397-425.

Table 1: Covariance structure models: first-order equivalent estimators (n=100)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
GMM	-.103	-.107	0.795	1.009	.065	.122	.107
CU	-.117	-.120	0.773	0.996	.070	.136	.120
ET	-.085	-.090	0.810	1.030	.067	.109	.091
EL	-.056	-.059	0.841	1.065	.068	.088	.068
			t_{10}				
GMM	-.049	-.050	0.865	1.038	.053	.073	.054
CU	-.056	-.057	0.854	1.033	.056	.079	.059
ET	-.036	-.039	0.877	1.053	.055	.066	.046
EL	-.018	-.021	0.897	1.074	.056	.058	.041
			Normal				
GMM	-.026	-.024	0.898	1.050	.047	.054	.038
CU	-.030	-.028	0.890	1.048	.049	.057	.040
ET	-.016	-.014	0.906	1.061	.048	.051	.036
EL	-.004	-.003	0.916	1.074	.048	.049	.034
			Lognormal				
GMM	-.419	-.435	0.436	0.770	.108	.433	.435
CU	-.486	-.496	0.330	0.722	.122	.501	.496
ET	-.419	-.432	0.400	0.789	.131	.439	.432
EL	-.334	-.350	0.481	0.892	.132	.359	.350
			Exponential				
GMM	-.148	-.154	0.716	0.999	.087	.171	.154
CU	-.171	-.174	0.667	0.987	.098	.197	.174
ET	-.123	-.125	0.728	1.030	.091	.153	.126
EL	-.074	-.078	0.783	1.084	.090	.116	.087

Table 2: Covariance structure models: first-order equivalent estimators (n=500)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
GMM	-.039	-.039	0.906	1.015	.034	.052	.040
CU	-.040	-.040	0.905	1.014	.034	.053	.041
ET	-.027	-.027	0.919	1.026	.033	.043	.030
EL	-.014	-.014	0.931	1.041	.034	.037	.025
			t_{10}				
GMM	-.014	-.014	0.947	1.026	.025	.028	.020
CU	-.014	-.014	0.947	1.026	.025	.029	.020
ET	-.008	-.009	0.954	1.032	.025	.026	.018
EL	-.002	-.003	0.959	1.038	.025	.025	.017
			Normal				
GMM	-.006	-.006	0.961	1.029	.021	.021	.014
CU	-.006	-.006	0.960	1.029	.021	.021	.014
ET	-.003	-.003	0.964	1.032	.021	.021	.013
EL	.001	.001	0.967	1.036	.021	.021	.013
			Lognormal				
GMM	-.227	-.230	0.647	0.916	.083	.242	.230
CU	-.233	-.235	0.632	0.910	.085	.248	.235
ET	-.181	-.186	0.701	0.962	.080	.198	.186
EL	-.122	-.129	0.753	1.032	.081	.147	.130
			Exponential				
GMM	-.043	-.044	0.891	1.028	.041	.059	.046
CU	-.044	-.045	0.888	1.028	.041	.060	.047
ET	-.026	-.028	0.911	1.043	.039	.047	.033
EL	-.009	-.009	0.928	1.057	.039	.040	.029

Table 3: Covariance structure models: bias-corrected GMM estimators (n=100)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
GMM	-.103	-.107	0.795	1.009	.065	.122	.107
NPB	-.071	-.075	0.809	1.059	.076	.104	.082
RNPB	-.035	-.044	0.843	1.101	.108	.113	.065
FSGELB	-.043	-.049	0.841	1.089	.075	.087	.066
ABCa	-.056	-.061	0.831	1.068	.072	.092	.070
ABCb	-.046	-.052	0.843	1.083	.073	.086	.065
			t_{10}				
GMM	-.049	-.050	0.865	1.038	.053	.073	.054
NPB	-.023	-.026	0.883	1.073	.060	.064	.045
RNPB	-.010	-.012	0.898	1.084	.058	.059	.042
FSGELB	-.010	-.012	0.897	1.085	.058	.059	.039
ABCa	-.015	-.017	0.894	1.079	.057	.059	.042
ABCb	-.011	-.013	0.897	1.081	.057	.058	.041
			Normal				
GMM	-.026	-.024	0.898	1.050	.047	.054	.038
NPB	-.006	-.004	0.912	1.074	.050	.050	.035
RNPB	-.000	.000	0.921	1.078	.049	.049	.034
FSGELB	-.000	.000	0.922	1.079	.049	.049	.034
ABCa	-.001	-.000	0.920	1.078	.049	.049	.034
ABCb	-.000	-.000	0.921	1.077	.049	.049	.033
			Lognormal				
GMM	-.419	-.435	0.436	0.770	.108	.433	.435
NPB	-.386	-.411	0.428	0.878	.144	.412	.411
RNPB	-39.898	-.286	0.455	1.834	1243.082	1243.722	.326
FSGELB	-.276	-.305	0.518	1.011	.159	.319	.307
ABCa	-.357	-.378	0.464	0.884	.136	.382	.378
ABCb	-.298	-.321	0.509	0.965	.147	.332	.322
			Exponential				
GMM	-.148	-.154	0.716	0.999	.087	.171	.154
NPB	-.096	-.104	0.731	1.093	.112	.147	.113
RNPB	-.057	-.066	0.768	1.132	.109	.123	.086
FSGELB	-.050	-.058	0.782	1.135	.107	.118	.083
ABCa	-.084	-.090	0.754	1.090	.103	.132	.102
ABCb	-.064	-.070	0.776	1.115	.101	.120	.088

Table 4: Covariance structure models: bias-corrected GMM estimators (n=500)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
GMM	-.039	-.039	0.906	1.015	.034	.052	.040
NPB	-.020	-.020	0.921	1.040	.038	.042	.029
RNPB	-.012	-.014	0.927	1.052	.039	.040	.028
FSGELB	-.014	-.015	0.927	1.048	.038	.040	.028
ABCa	-.017	-.018	0.926	1.044	.037	.040	.029
ABCb	-.015	-.016	0.927	1.046	.037	.040	.028
			t_{10}				
GMM	-.014	-.014	0.947	1.026	.025	.028	.020
NPB	-.003	-.003	0.956	1.039	.026	.026	.018
RNPB	-.001	-.002	0.957	1.041	.026	.026	.018
FSGELB	-.001	-.001	0.958	1.040	.026	.026	.018
ABCa	-.002	-.002	0.958	1.040	.026	.026	.018
ABCb	-.001	-.002	0.958	1.040	.026	.026	.018
			Normal				
GMM	-.006	-.006	0.961	1.029	.021	.021	.014
NPB	.001	.001	0.965	1.035	.021	.021	.013
RNPB	.001	.001	0.966	1.036	.021	.021	.013
FSGELB	.001	.001	0.967	1.036	.021	.021	.013
ABCa	.001	.001	0.967	1.036	.021	.021	.013
ABCb	.001	.001	0.967	1.036	.021	.021	.013
			Lognormal				
GMM	-.227	-.230	0.647	0.916	.083	.242	.230
NPB	-.162	-.167	0.674	1.028	.110	.196	.169
RNPB	-.074	-.113	0.719	1.136	.742	.746	.130
FSGELB	-.123	-.128	0.716	1.067	.108	.163	.134
ABCa	-.163	-.167	0.687	1.006	.099	.190	.169
ABCb	-.134	-.140	0.720	1.037	.098	.166	.143
			Exponential				
GMM	-.043	-.044	0.891	1.028	.040	.059	.046
NPB	-.013	-.014	0.913	1.067	.045	.047	.032
RNPB	-.008	-.010	0.919	1.068	.044	.045	.031
FSGELB	-.008	-.010	0.920	1.069	.044	.045	.031
ABCa	-.012	-.013	0.916	1.063	.044	.045	.031
ABCb	-.010	-.011	0.919	1.063	.043	.044	.030

Table 5: Instrumental variable models: experimental designs

Experiment	s	ρ_{xu}	ρ_{xz_1}	ρ_{xz_2}	$\rho_{xz_3} = ... = \rho_{xz_{11}}$
1	2	0.7	0.3	0	-
2	11	0.7	0.3	0	0
3	11	0.7	0.7	0	0
4	11	0.3	0.7	0	0
5	11	0.7	0.3	0	0.3

Table 6: Instrumental variable models: first-order equivalent estimators (n = 100)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			Model 1				
GMM	.020	.066	0.590	1.366	.324	.325	.162
CU	.077	.004	0.218	1.330	6.141	6.141	.172
ET	-.111	.004	0.206	1.320	.698	.707	.173
EL	-.124	.007	0.178	1.328	.866	.874	.172
			Model 2				
GMM	.270	.272	1.012	1.507	.152	.310	.273
CU	.468	.043	-0.059	1.598	12.949	12.957	.217
ET	-.233	.005	-0.269	1.407	1.610	1.627	.205
EL	-.201	.004	-0.228	1.406	1.313	1.329	.202
			Model 3				
GMM	.069	.082	0.897	1.209	.096	.118	.095
CU	-.019	.005	0.740	1.162	.134	.136	.081
ET	-.016	.003	0.751	1.157	.123	.124	.077
EL	-.015	.002	0.759	1.153	.121	.122	.074
			Model 4				
GMM	.040	.051	0.802	1.260	.140	.146	.105
CU	-.018	-.005	0.680	1.253	.183	.184	.114
ET	-.016	-.005	0.695	1.236	.169	.170	.109
EL	-.016	-.003	0.698	1.237	.169	.170	.107
			Model 5				
GMM	.100	.110	0.905	1.253	.107	.146	.118
CU	-.034	-.001	0.634	1.192	.186	.189	.103
ET	-.029	-.003	0.669	1.181	.160	.163	.096
EL	-.028	.000	0.676	1.185	.158	.161	.093

Table 7: Instrumental variable models: first-order equivalent estimators (n = 500)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			Model 1				
GMM	.001	.014	0.797	1.157	.113	.113	.072
CU	-.011	.004	0.776	1.155	.118	.118	.073
ET	-.011	.004	0.776	1.155	.118	.118	.073
EL	-.011	.004	0.776	1.155	.118	.118	.073
			Model 2				
GMM	.088	.091	0.942	1.223	.089	.125	.097
CU	-.008	.005	0.783	1.162	.123	.124	.078
ET	-.008	.005	0.787	1.160	.122	.123	.078
EL	-.008	.006	0.786	1.160	.122	.123	.078
			Model 3				
GMM	.018	.018	0.946	1.085	.044	.047	.033
CU	-.001	-.000	0.918	1.068	.046	.046	.029
ET	-.001	-.001	0.921	1.068	.046	.046	.029
EL	-.001	-.001	0.921	1.068	.046	.046	.029
			Model 4				
GMM	.012	.014	0.912	1.106	.060	.061	.040
CU	.001	.003	0.899	1.098	.061	.061	.039
ET	.001	.002	0.898	1.096	.061	.061	.039
EL	.001	.002	0.900	1.098	.061	.061	.039
			Model 5				
GMM	.025	.028	0.940	1.104	.051	.057	.039
CU	-.002	.001	0.910	1.085	.055	.055	.036
ET	-.002	.001	0.908	1.083	.055	.055	.035
EL	-.002	.002	0.909	1.084	.055	.055	.035

Table 8: Instrumental variable models: bias-corrected GMM estimators (n = 100)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			Model 1				
GMM	.020	.066	0.590	1.366	.324	.325	.162
NPB	-.027	.061	0.387	1.371	.525	.526	.164
RNPB	-.028	.060	0.383	1.370	.524	.524	.163
FSGELB	-.020	.060	0.442	1.371	.520	.520	.169
ABCa	.020	.067	0.592	1.366	.325	.325	.161
ABCb	.020	.067	0.592	1.366	.325	.325	.161
			Model 2				
GMM	.270	.272	1.012	1.507	.152	.310	.273
NPB	.187	.199	0.825	1.508	.217	.287	.224
RNPB	.187	.199	0.823	1.506	.212	.283	.221
FSGELB	.191	.200	0.851	1.500	.203	.279	.217
ABCa	.163	.188	0.749	1.496	.243	.293	.222
ABCb	.182	.196	0.804	1.499	.218	.284	.217
			Model 3				
GMM	.069	.082	0.897	1.209	.096	.118	.095
NPB	.031	.047	0.825	1.188	.111	.116	.087
RNPB	.018	.033	0.811	1.178	.110	.112	.080
FSGELB	.014	.031	0.813	1.171	.111	.112	.079
ABCa	.008	.026	0.795	1.169	.115	.115	.080
ABCb	.011	.028	0.805	1.170	.113	.113	.080
			Model 4				
GMM	.040	.051	0.802	1.260	.140	.146	.105
NPB	.012	.021	0.745	1.249	.157	.158	.106
RNPB	.007	.019	0.743	1.243	.153	.153	.100
FSGELB	.004	.017	0.745	1.244	.154	.154	.105
ABCa	.005	.019	0.740	1.244	.155	.155	.104
ABCb	.006	.021	0.746	1.240	.152	.153	.102
			Model 5				
GMM	.100	.110	0.905	1.253	.107	.146	.118
NPB	.043	.057	0.802	1.226	.134	.140	.099
RNPB	.032	.044	0.790	1.214	.132	.136	.094
FSGELB	.031	.045	0.796	1.213	.130	.134	.093
ABCa	.022	.038	0.780	1.210	.138	.140	.093
ABCb	.029	.045	0.798	1.208	.133	.136	.092

Table 9: Instrumental variable models: bias-corrected GMM estimators (n = 500)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			Model 1				
GMM	.001	.014	0.797	1.157	.113	.113	.072
NPB	.001	.015	0.793	1.156	.114	.114	.073
RNPB	.001	.015	0.794	1.158	.114	.114	.073
FSGELB	.001	.014	0.787	1.157	.114	.114	.073
ABCa	.001	.014	0.797	1.157	.113	.113	.072
ABCb	.001	.014	0.797	1.157	.113	.113	.072
			Model 2				
GMM	.088	.091	0.942	1.223	.089	.125	.097
NPB	.027	.036	0.838	1.192	.111	.114	.081
RNPB	.026	.034	0.835	1.189	.111	.114	.080
FSGELB	.029	.039	0.849	1.192	.108	.112	.080
ABCa	.018	.030	0.826	1.186	.116	.117	.080
ABCb	.023	.034	0.835	1.187	.112	.114	.080
			Model 3				
GMM	.018	.018	0.946	1.085	.044	.047	.033
NPB	.003	.005	0.926	1.074	.046	.046	.029
RNPB	.002	.004	0.925	1.071	.046	.046	.029
FSGELB	.001	.004	0.925	1.072	.046	.046	.030
ABCa	.001	.002	0.924	1.072	.046	.046	.029
ABCb	.001	.002	0.924	1.071	.046	.046	.030
			Model 4				
GMM	.012	.014	0.912	1.106	.060	.061	.040
NPB	.003	.006	0.903	1.099	.061	.061	.041
RNPB	.003	.005	0.902	1.098	.061	.061	.041
FSGELB	.002	.005	0.898	1.098	.061	.061	.041
ABCa	.002	.005	0.901	1.099	.061	.061	.040
ABCb	.002	.004	0.901	1.099	.061	.061	.040
			Model 5				
GMM	.025	.028	0.940	1.104	.051	.057	.039
NPB	.004	.008	0.911	1.091	.054	.055	.036
RNPB	.003	.006	0.910	1.090	.054	.055	.036
FSGELB	.003	.006	0.912	1.090	.054	.055	.037
ABCa	.002	.005	0.910	1.088	.055	.055	.036
ABCb	.003	.005	0.911	1.089	.054	.054	.036

Figure 1: Covariance structure models: sampling cumulative density functions (n=100; 1000 replications)

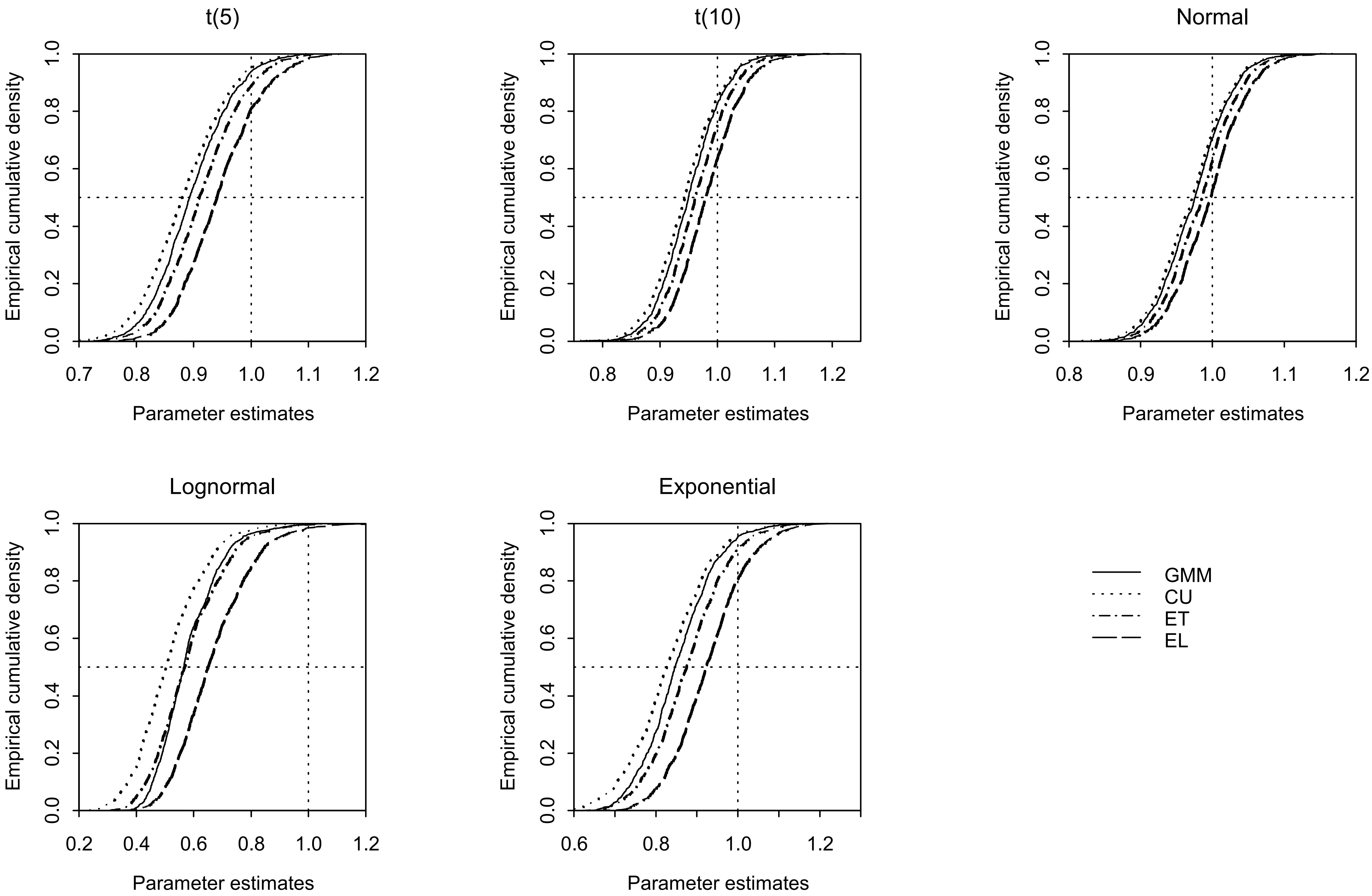

Figure 2: Covariance structure models: sampling cumulative density functions for bias-corrected GMM estimators (n=100; 1000 replications)

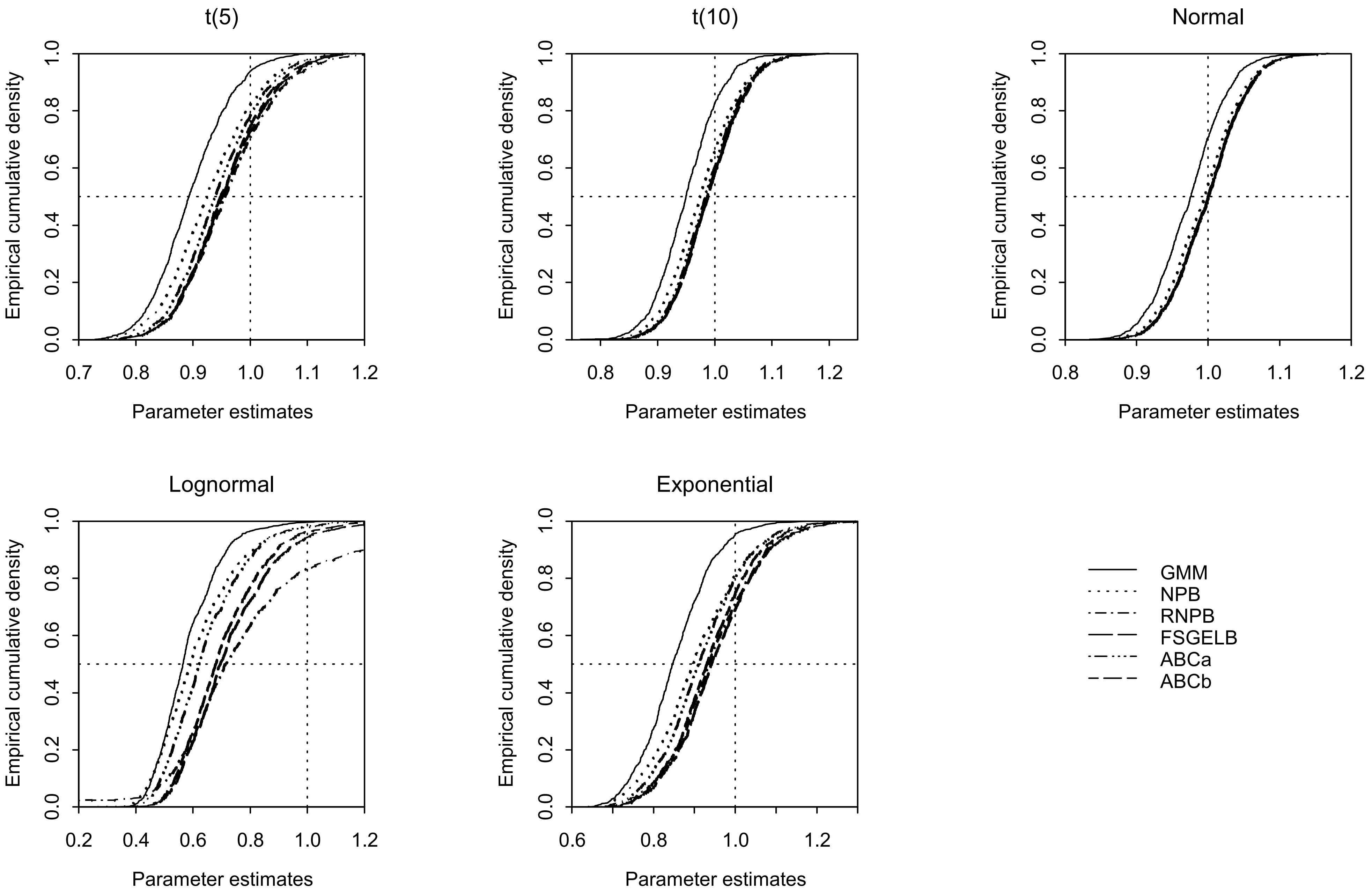

Figure 3: Instrumental variable models: sampling cumulative density functions (n=100; 1000 replications)

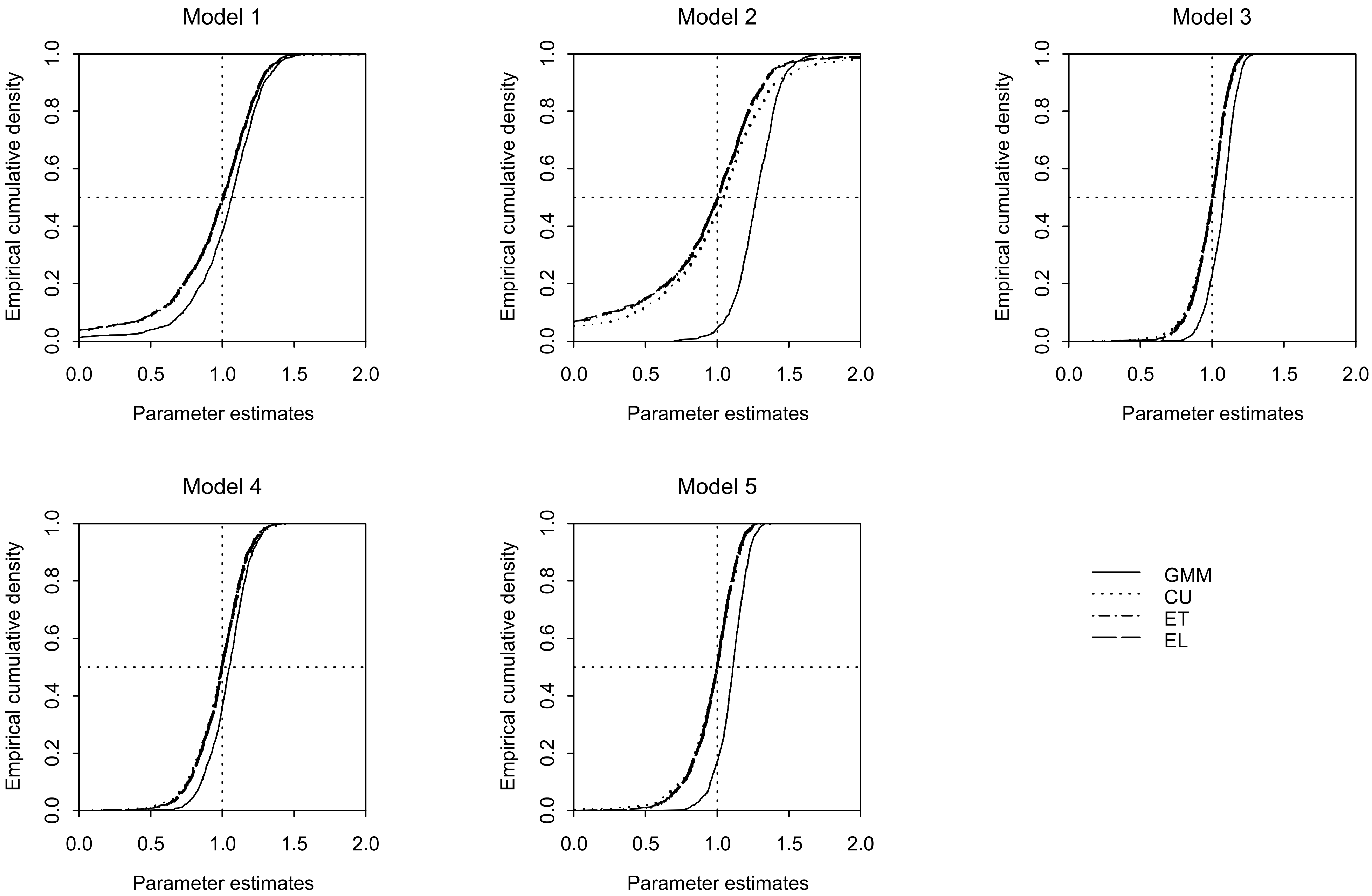

Figure 4: Instrumental variable models: sampling cumulative density functions for bias-corrected GMM estimators (n=100; 1000 replications)