UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ECONOMIA

DOCUMENTO DE TRABALHO Nº **2005/10**
June

Bootstrap bias-adjusted GMM estimators

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

* The author gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER. Address for correspondence: Joaquim J. S. Ramalho, Universidade de Évora, Departmento de Economia, 7000-803 ÉVORA, Portugal. Tel.: +351-266740894. Fax: +351-266740809.
E-mail: jsr@uevora.pt.

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

Abstract:

The ability of six alternative bootstrap methods to reduce the bias of GMM parameter estimates is examined in an instrumental variable framework using Monte Carlo analysis.
Promising results were found for the two bootstrap estimators suggested in the paper.

Palavras-chave/Keywords: GMM, Bootstrap, Empirical Likelihood, Instrumental Variables, Monte Carlo

Classificação JEL/JEL Classification: C13, C14

1 Introduction

It is now widely recognized that the efficient two-step generalized method of moments (GMM) estimator may have large biases for the sample sizes typically encountered in economic applications; see, for example, the several Monte Carlo studies that appeared in the July 1996 special issue of the *Journal of Business & Economic Statistics*. In this paper we analyze the ability of six alternative bootstrap procedures to reduce the finite sample bias of GMM parameter estimates. Three of those alternatives were already proposed by other authors: the standard, nonparametric (NP) bootstrap; Hall and Horowitz's (1996) recentered nonparametric (RNP) bootstrap; and Brown and Newey's (2002) constrained empirical likelihood (CEL) bootstrap. Monte Carlo evidence by Horowitz (1998) and Ramalho (2005) shows that application of these bootstrap methods reduces the bias of the GMM estimator but does not completely eliminate it. Therefore, in this paper we suggest two alternative bootstrap techniques, both of which use the empirical likelihood (EL) distribution function (see Qin and Lawless, 1994) to generate the bootstrap samples. The finite sample bias of all the corresponding bootstrap bias-corrected GMM estimators are examined in an instrumental variable framework through a Monte Carlo analysis.

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. Throughout, we assume that $s > k$ and 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 $g_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} g_i(\theta)$. Regularity conditions are assumed such that $g_n(\theta) \xrightarrow{p} E_F[g(y, \theta)]$ and $\sqrt{n} g_n(\theta_0) \xrightarrow{d} N(0, V)$, 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 GMM estimator $\hat{\theta}_{GMM}$ is obtained from minimization of the optimal quadratic form of the sample moment indicators

$$Q_n = g_n(\theta)' \left[V_n\left(\tilde{\theta}\right)\right]^{-1} g_n(\theta), \tag{2}$$

where $\tilde{\theta}$ is a preliminary consistent estimator for θ_0 and $\tilde{V}_n \equiv V_n\left(\tilde{\theta}\right)$ is a consistent estimator for V. Let $G \equiv E_F\left[\frac{\partial g_i(\theta_0)}{\partial \theta'}\right]$. Under suitable regularity conditions, see Newey and McFadden (1994), $\hat{\theta}_{GMM}$ is a consistent, asymptotically normal estimator of θ_0, $\sqrt{n}\left(\hat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left[0, \left(G'V^{-1}G\right)^{-1}\right]$, and is asymptotically efficient among all estimators based only on (1).

3 The empirical likelihood distribution function

Consider again the moment conditions (1). Implicitly, by giving the same weight (n^{-1}) to each observation, GMM uses the empirical distribution function $F_n(y) \equiv n^{-1}\sum_{i=1}^{n} 1(y_i \leq y)$ as estimate for $F(y)$, where the indicator function $1(y_i \leq y)$ is equal to 1 if $y_i \leq y$ and 0 otherwise. However, since 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 $F(y)$. Actually, we may obtain an alternative estimator for θ in (1) by choosing the estimator $\hat{\theta}$ that minimizes the distance, relatively to some metric, between $F_n(y)$ and a distribution function $F_p(y)$ satisfying the moment conditions (1). The distribution $F_p(y)$ is, hence, the member of the class $\mathcal{F}(\theta)$ of all distribution functions that satisfy (1), $\mathcal{F}(\theta) \equiv \left\{F_p : E_{F_p}\left[g(y, \theta_0)\right] = 0\right\}$, that is closest to F_n.

In the selection of a particular probability measure in $\mathcal{F}(\theta)$, different metrics for the closeness between $F_p(y)$ and $F_n(y)$ may be used. The most common choices for the metric $\mathcal{M}(F_n, F_p)$ are particular cases of the Cressie-Read (1984) power-divergence statistic, namely $\mathcal{M}(\cdot) = \sum dF_n(y) \ln\left[dFn(y)/dF_p(y)\right]$ which produces the so-called EL estimator. Thus, the EL estimator $\hat{\theta}_{EL}$ can be described as the solution to the program

$$\max_{\theta} \sum_{i=1}^{n} \ln p_i^{EL} \text{ subject to } p_i^{EL} \geq 0, \ \sum_{i=1}^{n} p_i^{EL} = 1 \text{ and } \sum_{i=1}^{n} p_i^{EL} g(y_i, \theta) = 0, \tag{3}$$

where $p_i^{EL} \equiv dF_p(y_i)$, $i = 1, ..., n$, and the last restriction is an empirical measure counterpart to the moment conditions (1), imposing them numerically in the sample; for an alternative motivation of EL estimators, see Newey and Smith (2004).

From optimization of (3), it is straightforward to show that

$$\hat{p}_i^{EL} \equiv p_i^{EL}\left(\hat{\theta}_{EL}, \hat{\lambda}_{EL}\right) = \frac{1}{n\left[1 + \hat{\lambda}_{EL}' g\left(y_i, \hat{\theta}_{EL}\right)\right]}, \ i = 1, ..., n, \tag{4}$$

where $\hat{\theta}_{EL}$ and $\hat{\lambda}_{EL}$, the s-vector of Lagrange multipliers associated to the last restriction of (3), result from unconstrained optimization of the saddle function $n^{-1}\sum_{i=1}^{n} \ln\left[1 + \lambda' g_i(\theta)\right]$. Thus, the EL distribution function $F_p(y)$ is given by

$$F_p(y) = \sum_{i=1}^{n} \hat{p}_i^{EL} 1(y_i \leq y). \quad (5)$$

See Qin and Lawless (1994) for details.

4 Alternative bootstrapping for the GMM estimator

Assume that a random sample S of size n is collected from a population whose (unknown) distribution function is $F(y)$. Bootstrap samples are generated by randomly sampling S with replacement. This resampling is based on a certain distribution function, $F^*(y)$, which assigns each observation a given probability of being sampled. In general, using the bootstrap, the bias of the GMM estimator $\hat{\theta}_{GMM}$ can be estimated as follows: 1) compute $\hat{\theta}_{GMM}$ by minimizing (2) based on S; 2) generate B bootstrap samples S_j^*, $j = 1, ..., B$, of size n accordingly with the chosen $F^*(y)$: $S_j^* = \left\{y_{j1}^*, ..., y_{jn}^*\right\}$, 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 $\hat{\theta}_j^* \equiv \arg\min_{\theta} g_{jn}^*(\theta) \hat{V}_{jn}^{*-1} g_{jn}^*(\theta)$, $j = 1, ..., B$, where $g_{jn}^*(\theta) = n^{-1} \sum_{i=1}^{n} g\left(y_{ji}^*, \theta\right)$ and $\hat{V}_{jn}^{*-1}$ uses 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:

$$\hat{b} = \bar{\theta}^* - \hat{\theta}_{GMM}. \quad (6)$$

Subtracting the bias (6) from the GMM estimator $\hat{\theta}_{GMM}$, it is then possible to obtain the bias-corrected GMM estimator

$$\hat{\theta}_{BCGMM} = 2\hat{\theta}_{GMM} - \bar{\theta}^*. \quad (7)$$

As discussed next, these general procedures may be used to reduce the finite sample bias of GMM parameter estimates in several distinct forms.

4.1 Nonparametric bootstrap

The NP bootstrap is probably the most commonly applied bootstrap technique in econometrics. In this case, the bootstrap samples are generated using the empirical distribution function $F_n(y)$, so each observation has equal probability n^{-1} of being drawn. However, direct application of the NP bootstrap 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.

4.2 Recentered nonparametric bootstrap

In order to guarantee that the moment conditions exploited by GMM estimators hold exactly in each replication of the bootstrap, Hall and Horowitz (1996) suggested using the recentered moment indicators

$$g^c\left(y_j^*,\theta\right) = g\left(y_j^*,\theta\right) - \frac{1}{n}\sum_{i=1}^{n} g\left(y_i,\hat{\theta}_{GMM}\right), \tag{8}$$

since $E_{F_n}\left[g^c\left(y_j^*,\theta\right)\right] = 0$. To implement this 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 (8), with the weight matrix specified accordingly.

4.3 Constrained empirical likelihood bootstrap

Instead of recentering the moment conditions and keeping $F_n(y)$ as resampling distribution, Brown and Newey (2002) suggested generating the bootstrap samples using a different distribution, say $F_1(y)$, such that $E_{F_1}\left[g\left(y,\hat{\theta}_{GMM}\right)\right] = 0$. Namely, they proposed the employment of a constrained version of the EL distribution function (5), which is given by

$$F_p^c(y) = \sum_{i=1}^{n} \hat{p}_i^{CEL} 1\left(y_i \leq y\right), \tag{9}$$

where

$$\hat{p}_i^{CEL} = \frac{1}{n\left[1 + \hat{\lambda}_{CEL}' g\left(y_i,\hat{\theta}_{GMM}\right)\right]},\ i = 1,...,n, \tag{10}$$

and $\hat{\lambda}_{CEL}$ results from maximization of $n^{-1}\sum_{i=1}^{n} \ln\left[1 + \lambda' g_i\left(\hat{\theta}_{GMM}\right)\right]$; in other words, $F_p^c(y)$ results from solving the program (3) conditional on $\theta = \hat{\theta}_{GMM}$. Since $\sum_{i=1}^{n} \hat{p}_i^{CEL} g_i\left(\hat{\theta}_{GMM}\right) = 0$ is the first-order condition characterizing $\hat{\lambda}_{CEL}$, this CEL bootstrap imposes, in effect, the moment conditions, evaluated at $\hat{\theta}_{GMM}$, on the sample: $E_{F_{CEL}}\left[g\left(y,\hat{\theta}_{GMM}\right)\right] = 0$.

Brown and Newey (2002) proved that the CEL bootstrap is asymptotically efficient relative to the NP and RNP methods, since $F_p^c(y)$ is a more efficient estimator of $F(y)$ than $F_n(y)$.

4.4 Recentered empirical likelihood bootstrap

The two bootstrap methods that we propose in this paper are based on the EL distribution $F_p(y)$ given in (5). Although they are not expected to be more efficient than the CEL bootstrap, the fact that $F_p(y)$ is used instead of $F_p^c(y)$ may lead to better results in finite samples for two reasons: first, the former distribution do not result from an optimization conditional on $\theta = \hat{\theta}_{GMM}$ as the latter; second, there are some Monte Carlo evidence suggesting that $\hat{\theta}_{EL}$ displays less bias than $\hat{\theta}_{GMM}$ in small samples; see *inter alia* Ramalho (2005).

As before, some correction seems to be necessary to apply this EL bootstrap to the GMM estimator, since $\sum_{i=1}^{n} \hat{p}_i^{EL} g\left(y_i, \hat{\theta}_{GMM}\right) \neq 0$ in general. Analogously to Hall and Horowitz (1996), we suggest using the recentered moment indicators

$$g^c\left(y_j^*, \theta\right) = g\left(y_j^*, \theta\right) - \sum_{i=1}^{n} \hat{p}_i^{EL} g\left(y_i, \hat{\theta}_{GMM}\right), \tag{11}$$

since $E_{F_{EL}}\left[g^c\left(y_j^*, \theta\right)\right] = 0$. This recentered EL (REL) bootstrap can be implemented applying similar procedures to those described for the RNP method, with only two (obvious) alterations: $F_p(y)$ is used instead of $F_n(y)$ and (11) instead of (8).

4.5 Post-hoc empirical likelihood bootstrap

The expected failure of the EL bootstrap in providing significantly less biased GMM estimators can be also explained as follows. Let $p^{EL} = \left(\hat{p}_1^{EL}, ..., \hat{p}_n^{EL}\right)$ be the n-dimensional resampling vector that assigns each observation a given probability of being sampled in the EL bootstrap. By using this resampling vector and estimating the bias utilizing the formula given in (6), we are not adequately estimating the bias of the GMM estimator that we intended to correct. Actually, in the calculation of (6), we are comparing GMM estimators that can be based on quite distinct samples: while $\hat{\theta}_{GMM}$ results from the minimization of the quadratic form (2), $\bar{\theta}^*$ is the average of the standard GMM estimators $\hat{\theta}_j$, $j = 1, ..., B$, each of which, due to the way the bootstrap samples are constructed, can be interpreted as minimizing also (2) but with $g_n(\theta)$ replaced by $g_p(\theta) \equiv \sum_{i=1}^{n} \hat{p}_i^{EL} g(y_i, \theta)$, which, in small samples, can be rather different. Based on these arguments, we suggest below the post-hoc EL (PHEL) bootstrap, which uses a post-sampling adjustment to the EL bootstrap GMM estimator.[1]

Define $p_j^a \equiv \left(p_{j1}^a, ..., p_{jn}^a\right)$ as the actual or post-resampling vector calculated from the bootstrap sample S_j^*, that is $p_{ji}^a = \#\left\{y_{ji}^* = y_i\right\}/n$ is the proportion of times that the i-th original data

[1] For other applications of post-sampling adjustments, see Efron (1990).

point appeared in the bootstrap sample S_j^*. Define also the average post-resampling vector $\bar{p}^a \equiv (\bar{p}_1^a, ..., \bar{p}_n^a) = B^{-1} \sum_{j=1}^{B} p_j^a$. In this framework, the j-th bootstrap estimator $\bar{\theta}_j^*$ can be expressed as a function of the j-th post-resampling vector: $\bar{\theta}_j^* = \theta\left(p_j^a\right)$. Similarly, we have for the original GMM estimator $\hat{\theta}_{GMM} = \theta\left(p^0\right)$, where $p^0 = \left(n^{-1}, ..., n^{-1}\right)$. Define also $\hat{\theta}^a = \theta\left(\bar{p}^a\right)$ as the GMM estimator resultant from the application of the average post-sampling probabilities $\bar{p}^a$, i.e. based on $\bar{g}^a(\theta) = \sum_{i=1}^{n} \bar{p}_i^a g(y_i, \theta)$.

Instead of using $\hat{b} = \bar{\theta}^* - \theta\left(p^0\right)$, we propose the calculation of the bias of the GMM estimator as:

$$\bar{b} = \bar{\theta}^* - \theta\left(\bar{p}^a\right). \tag{12}$$

The intuition behind this is the following. Although the theoretical expectation of the resampling vector p^{EL} is p^0, its actual average is $\bar{p}^a$. Thus, using $\theta\left(\bar{p}^a\right)$ instead of $\theta\left(p^0\right)$ in the estimation of the bias, we might be able to correct for this discrepancy. In fact, in (12), we are effectively comparing GMM estimators based on similar samples, in opposition to what was happening before. The bias-corrected GMM estimator is then found by calculating:

$$\hat{\theta}_{BCGMM} = \hat{\theta}_{GMM} - \bar{\theta}^* + \hat{\theta}^a. \tag{13}$$

When both n and B go to infinity, $\hat{\theta}^a$ will converge to $\hat{\theta}_{GMM}$, so asymptotically this method will produce the same results as the other bootstrap techniques discussed in the previous sections. Note that we could have also opted for estimating the bias by $\bar{b} = \bar{\theta}^* - \theta\left(\hat{p}^{EL}\right)$, since $\bar{p}^a \simeq \hat{p}^{EL}$. However, the utilization of the post-resampling probabilities are expected to provide a slight further improvement.

In terms of procedures, the algorithm presented earlier must be modified as follows. In step 3, for each bootstrap sample, in addition to the GMM estimator $\bar{\theta}_j^*$, we calculate also p_j^a. In step 4, the average post-resampling vector $\bar{p}^a$ needs also to be calculated. In the final step, we need to obtain $\hat{\theta}^a$ and, instead of (6), the bias is calculated according to (12).

5 Monte Carlo simulation

Consider the linear instrumental variable model described by equations

$$Y_i = \theta_0 \cdot X_i + \epsilon_i,$$

$$X_i = \sum_{j=1}^{s} \pi \cdot Z_{ij} + u_i,$$

where Y_i and X_i denote the dependent variable and an exogenous regressor, respectively. All the instruments Z_{ij} are i.i.d. $\mathcal{N}(0, I_s)$ variables, while $(\epsilon_i, u_i)'$ is $\mathcal{N}(0, \Omega)$, where Ω is a (2×2)-matrix with diagonal and off diagonal elements 1 and ρ, respectively. We considered three different degrees of non-orthogonality between X_i and ϵ_i, $\rho = (0.25, 0.50, 0.75)$. Let $R_f^2 = s\pi^2/(s\pi^2 + 1)$ be the theoretical R^2 of the first stage regression, which measures the overall fit of the instruments to the endogenous regressor X_i. We fix $s = 10$ and set the value of π in such a way that $R_f^2 = (0.15, 0.30)$ in all the experiments. The value of θ_0 was fixed in order to keep constant the overall fit of Y_i to X_i in the structural equation ($R^2 = 0.5$). For each one of the 6 parameter combinations of s, R_f^2, and ρ we generated 5000 Monte Carlo samples of size $n = 200$.

In Table 1 we report for each estimator the mean and median bias, the median absolute error (MAE), and the standard error (SE) across replications. As expected, the bias of the GMM estimator increases with the endogeneity of the model and decreases with the strength of the instruments. The same pattern can be observed for all the bootstrap GMM estimators. The utilization of any one of the bootstrap methods allows the bias of the GMM estimator to be substantially reduced, although at the expense of an increment in its dispersion. Clearly, the two estimators suggested in this paper display the best performances in terms of mean and median bias, particularly the PHEL bootstrap, which produces the only estimator which is approximately mean unbiased in all cases. Conversely, the Monte Carlo distribution of this estimator is slightly more disperse. Overall, these results suggest that the estimators developed in this paper will be useful, at least, in settings similar to those replicated in this Monte Carlo study.

References

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

Brown, B.W. and W.K. Newey, 2002, Generalised Method of Moments, Efficient Bootstrapping, and Improved Inference, Journal of Business and Economic Statistics 20, 507-517.

Efron, B., 1990, More efficient bootstrap computations, Journal of the American Statistical Association 85(409), 79-89.

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

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

Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. 4 (North-Holland, Amsterdam) 2111-2245.

Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica 72(1), 219-255.

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

Ramalho, J.J.S., 2005, Small sample bias of alternative estimation methods for moment condition models: Monte Carlo evidence for covariance structures, Studies in Nonlinear Dynamics and Econometrics, forthcoming.

Table 1: Monte Carlo results (5000 replications; n = 200; s = 10)

Estimator	Bias		MAE	SE
	mean	median		
$\rho = 0.25,\ R_f^2 = 0.15$				
GMM	.050	.054	.112	.158
NP	.019	.027	.123	.189
RNP	.019	.027	.121	.187
CEL	.019	.026	.122	.186
REL	.016	.023	.122	.191
PHEL	.004	.017	.127	.214
$\rho = 0.25,\ R_f^2 = 0.3$				
GMM	.023	.026	.074	.108
NP	.006	.009	.078	.119
RNP	.006	.009	.077	.117
CEL	.006	.009	.077	.117
REL	.005	.008	.078	.118
PHEL	.004	.007	.078	.119
$\rho = 0.5,\ R_f^2 = 0.15$				
GMM	.099	.106	.131	.153
NP	.036	.049	.127	.187
RNP	.036	.050	.125	.185
CEL	.037	.049	.124	.183
REL	.026	.041	.126	.191
PHEL	.004	.027	.129	.214
$\rho = 0.5,\ R_f^2 = 0.3$				
GMM	.045	.049	.080	.106
NP	.011	.018	.078	.119
RNP	.011	.017	.078	.118
CEL	.010	.016	.077	.117
REL	.008	.014	.078	.119
PHEL	.005	.012	.078	.120
$\rho = 0.75,\ R_f^2 = 0.15$				
GMM	.147	.157	.165	.142
NP	.052	.070	.132	.184
RNP	.053	.072	.131	.181
CEL	.057	.076	.131	.177
REL	.033	.054	.129	.190
PHEL	.001	.031	.132	.214
$\rho = 0.75,\ R_f^2 = 0.3$				
GMM	.067	.073	.090	.102
NP	.016	.025	.081	.119
RNP	.016	.026	.079	.118
CEL	.016	.024	.079	.117
REL	.009	.018	.079	.120
PHEL	.005	.015	.080	.122