Studies in Nonlinear Dynamics and Econometrics

**Quarterly Journal**
**July 1997, Volume 2, Number 2**
**The MIT Press**

*Studies in Nonlinear Dynamics and Econometrics* (ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; tel.: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $130.00. Canadians add additional 7% GST. Prices subject to change without notice.

# Finite Sample Properties of the Efficient Method of Moments

**Rómulo A. Chumacero**
**Department of Economics**
**University of Chile**
***rchumace@decon.facea.uchile.cl***

***Abstract.*** *Gallant and Tauchen (1996) describe an estimation technique, known as Efficient Method of Moments (EMM), that uses numerical methods to estimate parameters of a structural model. The technique uses as matching conditions (or moments, in the GMM jargon) the gradients of an auxiliary model that fits a subset of variables that may be simulated from the structural model.*

*This paper presents three Monte Carlo experiments to assess the finite sample properties of EMM. The first one compares it with a fully efficient procedure (Maximum Likelihood) by estimating an invertible moving-average (MA) process. The second and third experiments compare the finite sample properties of the EMM estimators with those of GMM by using stochastic volatility models and consumption-based asset-pricing models. The experiments show that the gains in efficiency are impressive; however, given that both EMM and GMM share the same type of objective function, finite sample inference based on asymptotic theory continues to lead, in some cases, to "over rejections," even though they are not as significant as in GMM.*

***Keywords.*** Monte Carlo, efficient method of moments, maximum likelihood, generalized method of moments, stochastic volatility, asset pricing

***Acknowledgments.*** This paper is based on Chapter III of my doctoral thesis at Duke University. I would like to thank the members of my committee, Profs. Ron Gallant, Kent Kimbrough, Ravi Bansal, Pietro Peretto, and especially George Tauchen for their comments. I would also like to thank the participants of the Macro Workshop at Duke and the Economics Workshop at the University of Chile, and to Bruce Mizrach and an anonymous referee for their suggestions. The usual disclaimer applies.

## 1 Introduction

Almost 15 years have passed since Hansen (1982) introduced the Generalized Method of Moments (GMM) estimation technique. Even though the asymptotic properties of GMM are well known, its application to small samples presents several problems that are well documented in the literature.

Beginning with the work by Tauchen (1986), several Monte Carlo experiments were conducted to assess the small sample properties of GMM in different setups. The general consensus of these exercises appears to be that:[1]

- Regardless of the method used for choosing the weighting matrix, inference based on the GMM criterion function leads to excessive rejections of the true hypothesis ("over rejection").

[1] See, e.g., Kocherlakota (1990), Andersen and Sorensen (1994), Burnside and Eichenbaum (1994), Christiano and den Haan (1994), and Hansen, Heaton, and Yaron (1995).

 

- Depending on the experiment, important biases on the estimates of several parameters have been detected (particularly in consumption-based asset-pricing models).
- Results tend to improve very slowly with increments in the sample size.

These troublesome results were found despite the fact that most of the Monte Carlo environments in which the experiments were conducted consisted of simple models of endowment economies (as in Tauchen [1986], Kocherlakota [1990], and Hansen, Heaton, and Yaron [1995]) or simple univariate processes (as in Christiano and den Haan [1994] or Burnside and Eichenbaum [1994]).

This paper presents three Monte Carlo experiments applied to the estimation method developed by Gallant and Tauchen (1996), known as Efficient Method of Moments (EMM), and compares their results with available alternative estimation procedures. The experiments are characterized by varying degrees of complexity, so that a comparison of EMM performance can be assessed in several dimensions. The first experiment compares it with a fully efficient procedure (Maximum Likelihood) by fitting an invertible MA process with an autoregressive (AR) representation. The second and third experiments compare the finite sample properties of EMM estimators with those of GMM by using stochastic volatility models and consumption-based asset-pricing models.

The paper is organized as follows. Section 2 presents a brief introduction to the EMM estimation technique. Section 3 relates the results of a simple Monte Carlo experiment in which EMM is compared to Maximum Likelihood (ML). Section 4 compares the small-sample properties of EMM and GMM in a stochastic volatility model. Section 5 reports the results of EMM and GMM when applied to a simple consumption-based asset-pricing model. Finally, Section 6 summarizes the main findings.

## 2 The Efficient Method of Moments

The purpose of this section is to present a brief introduction to the type of EMM estimators that will be used in the following sections. The interested reader is referred to Gallant and Tauchen (1996) for a formal treatment of this and other setups in which EMM can be applied.

Consider a stationary stochastic process $p(y_t \mid x_t, \rho)$ that describes $y_t$ in terms of the exogenous variables $x_t$ and structural parameters $\rho$ that the researcher is interested in estimating. Consider an auxiliary model $f(y_t \mid x_t, \theta)$ that has an analytical expression, whereas the density $p(y_t \mid x_t, \rho)$ does not. Gallant and Tauchen (1996) propose to use the scores of the auxiliary model:

$$(\partial/\partial\theta) \ln f(y_t \mid x_t, \hat{\theta}_T) \tag{1}$$

in order to generate the GMM moment conditions

$$m_T(\rho, \hat{\theta}_t) = \int\int (\partial/\partial\theta) \ln f(y \mid x, \hat{\theta}_T)\, p(y \mid x, \rho)\, dy p(x \mid \rho)\, dx \tag{2}$$

where $\hat{\theta}_T$ is defined as the maximum likelihood of $f(\cdot)$ for a sample of size $T$. That is:

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\text{argmax}} \frac{1}{T}\sum_{t=1}^{T} \ln f(y_t \mid x_t, \theta) \tag{3}$$

For cases in which analytical expressions for Equation (2) are not available, simulations may be required to compute them; in that situation, we define the moments as:

$$m_T(\rho, \hat{\theta}_T) = \frac{1}{N}\sum_{n=1}^{N} (\partial/\partial\theta) \ln f(\tilde{y}_n \mid \tilde{x}_n, \hat{\theta}_T) \tag{4}$$

where $N$ is the sample size of the Monte Carlo integral approximation drawn from one sample of $y, x$ generated for a given value of $\rho$ in the structural model.

The GMM estimator of $\rho$ with an efficient weighting matrix is given by:

$$\hat{\rho}_N = \underset{\rho \in R}{\operatorname{argmin}}\, m_N'(\rho, \hat{\theta}_T)(\hat{I}_T)^{-1} m_N(\rho, \hat{\theta}_T) \tag{5}$$

If the auxiliary model constitutes a good statistical description of the data-generating process of $y$, the outer product of the gradients (OPG) can be used in the weighting matrix; that is:

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^{T} [(\partial/\partial\theta) \ln f(y_t \mid x_t, \hat{\theta}_T)][(\partial/\partial\theta) \ln f(y_t \mid x_t, \hat{\theta}_T)]' \tag{6}$$

Gallant and Tauchen (1996) demonstrate the strong convergence and asymptotic normality of the estimator presented in Equation (5), as well as the asymptotic distribution of the objective function that $\hat{\rho}_N$ minimizes. That is, let $k$ be the dimension of $\rho$ and $q$ the dimension of $\theta$, then:

$$TJ_T = Tm_N'(\hat{\rho}_N, \hat{\theta}_T)(\hat{I}_T)^{-1} m_N(\hat{\rho}_N, \hat{\theta}_T) \xrightarrow{d} \chi^2_{q-k} \tag{7}$$

which corresponds to the familiar overidentifying restrictions test described by Hansen (1982). As in GMM, identification requires that $q \geq k$. Thus, statistical inference may be carried out in identical fashion as in GMM. However, depending on the complexity of the auxiliary model, Wald-type tests based on the variance-covariance matrix obtained by differentiating the moments may be difficult to construct.

One major advantage of EMM is that the econometrician does not need to observe directly all the variables in the structural model to compute $\rho$. This feature is extremely attractive, because in many cases the poor small-sample performance of GMM estimators is due to the limited amount of observations that the econometrician has to estimate the structural model. In other cases, the poor small-sample performance of GMM may be explained by the arbitrariness of the choice of moments in the estimation. EMM seeks to improve efficiency on the estimation of the structural parameters by choosing an auxiliary model that presents a good statistical description of the data. Intuitively, EMM would lead to full efficiency (that is comparable to the ML estimator) if the auxiliary model is judiciously chosen.

Even though Gallant and Tauchen (1996) derived the asymptotic properties of EMM, nothing is known with respect to its performance in small samples. The rest of the paper presents different environments in which this is assessed.

In all of the experiments that will be discussed in the following sections, we focus our attention on four statistics that were suggested by Hansen, Heaton, and Yaron (1995) and that are frequently used to conduct inference with GMM:

- The first corresponds to the familiar test of overidentifying restrictions already introduced in Equation (7) that will be called $TJ_T$.
- The second corresponds to the difference between the value of the objective function evaluated at the true parameter ($\rho_0$) and $TJ_T$, which will be called $TJ_T^a$. That is:

$$TJ_T^a = Tm_N'(\rho_0, \hat{\theta}_T)(\hat{I}_T)^{-1} m_N(\rho_0, \hat{\theta}_T) - TJ_T \xrightarrow{d} \chi^2_k \tag{8}$$

- The third corresponds to the difference between the value of the objective when a subset of $\rho$ (of dimension $k_1$) is constrained to its true value and $TJ_T$, which will be called $TJ_T^b$. This statistic converges to a $\chi^2$ distribution with $k$–$k_1$ degrees of freedom.
- The last statistic that will be computed is the standard Wald statistic, which is used for constructing confidence intervals and testing hypotheses. A special case of interest that will be tested in the Monte Carlo experiment corresponds to that in which we want to conduct inference about $\rho$ when it is a scalar. In that case, we have:

$$(\hat{\rho}_T - \rho_0)/\sigma_{\hat{\rho}_T} \xrightarrow{d} N(0, 1) \tag{9}$$

**Table 1**
Properties of Estimators: MA(1) Model

| Estimator | $\alpha = 0.5$ | | | $\sigma = 1$ | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Standard Deviation** | **RMSE** | **Mean** | **Standard Deviation** | **RMSE** |
| ML | 0.498 | 0.056 | 0.056 | 0.994 | 0.045 | 0.045 |
| EMM2a | 0.491 | 0.077 | 0.077 | 0.985 | 0.049 | 0.051 |
| EMM2b | 0.492 | 0.077 | 0.077 | 0.985 | 0.049 | 0.051 |
| EMM3a | 0.484 | 0.066 | 0.068 | 0.979 | 0.048 | 0.053 |
| EMM3b | 0.485 | 0.067 | 0.069 | 0.979 | 0.049 | 0.053 |

Notes: The results were obtained by estimating 1,000 samples, each of size 250. RMSE = Root Mean Square Error. ML = results obtained with the Conditional Maximum Likelihood estimator. EMM2a = Results from EMM with an AR(2) auxiliary model and the OPG as weighting matrix. EMM2b = Results from EMM with an AR(2) auxiliary model and the HAC weighting matrix. EMM3a = Results from EMM with an AR(3) auxiliary model and the OPG as weighting matrix. EMM3b = Results from EMM with an AR(3) auxiliary model and the HAC weighting matrix.

## 3 A Simple Example

In this section, we consider the estimation of the parameters of an invertible MA(1) process using EMM. This experiment was proposed by Gourieroux and Monfort (1993) when they assessed the small sample properties of an estimation technique called Indirect Inference (II), which is similar to EMM. This environment is useful to compare the performance of EMM vis-à-vis a fully efficient method such as ML.

### *3.1 The environment*

The process that we are interested in estimating is:

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1} \qquad \varepsilon_t \approx N(0, \sigma^2) \tag{10}$$

where the true value of $\alpha$ is 0.5, and the true value of $\sigma^2$ is 1. The aim is to compare the properties of the exact ML estimators of $\alpha$ and $\sigma$ with those of several EMM estimators based on AR(j) auxiliary models. Thus, in this case, our auxiliary model would be:

$$y_t = \sum_{i=1}^{j} \beta_i y_{t-i} + u_t \qquad u_t \approx N(0, \sigma_u^2) \tag{11}$$

To make the correspondence with the notation introduced in the previous section, the structural parameters of this experiment are $\rho = (\alpha, \sigma)$, and the parameters of the auxiliary model are $\theta = (\beta_1, \beta_2, \ldots, \beta_j, \sigma_u^2)$.

The experiment was conducted by estimating 1,000 samples, each of size 250, using both EMM and ML.

### *3.2 The results*

Given that the value of $\alpha$ is relatively far away from the unit circle, results obtained with the exact ML estimator and the conditional ML estimator were almost identical, and we will only report the estimates of the latter.

In the case of EMM, we report the results of four estimators that differ on the choice of the auxiliary model and the choice of weighting matrix. In particular, EMMja denotes the EMM estimator with an AR(j) auxiliary model ($j = 2, 3$) and the OPG as weighting matrix, while EMMjb denotes the EMM estimator with an AR(j) auxiliary model ($j = 2, 3$) and the heteroskedasticity and autocorrelation consistent (HAC) weighting matrix estimated using the quadratic spectrum kernel and the VAR(1) prewhitening method suggested by Andrews and Monahan (1992).

Table 1 and Figure 1 present the basic descriptive statistics and estimated probability density functions (pdfs) of the ML and EMM estimators. For all the methods, the bias is small while the differences in RMSE among the different estimators do not depend on the choice of the weighting matrix. This result arises from the fact that $\alpha$ is relatively far away from the unit circle, and there is not much autocorrelation left, even with

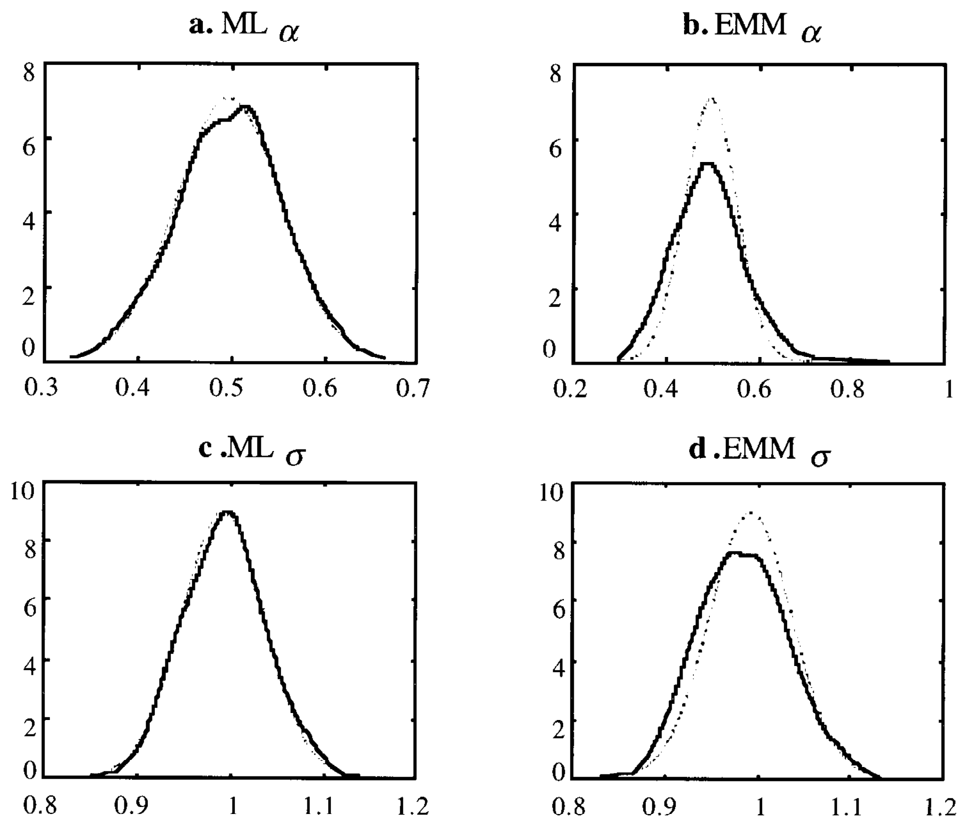


**Figure 1**
Estimated Probability Density Functions for the MA(1) Model
The continuous lines correspond to the Epanechnikov kernel estimators of the pdfs using the bandwidth proposed by Silverman (1986). The dotted lines correspond to the normal pdf with the same mean and variance of the ML estimator. **a**, Density of the ML estimator of $\alpha$. **b**, Density of the EMM estimator of $\alpha$ with an AR(2) auxiliary model and the OPG as the weighting matrix. **c**, Density of the ML estimator of $\sigma$. **d**, Density of the EMM estimator of $\sigma$ with an AR(2) auxiliary model and the OPG as the weighting matrix.

the AR(2) approximation. Results with increasing sample sizes show a steady convergence toward the ML RMSE, even though there is still an important loss in efficiency (particularly in the case of $\alpha$).

Figure 2 shows the results of the four statistics described in the previous section. Whenever possible, we compare these statistics with the Likelihood Ratio Test (LRT) that is obtained from the ML estimator. In reporting our results, we use the graphical method advocated by Davidson and McKinnon (1994) and Hansen, Heaton, and Yaron (1995). These plots are referred to as "$p$-value" plots, and are presented for the intervals [0, 0.5]. Briefly, the $p$-value plots are constructed as follows:

- For each probability value (depicted in the $x$-axis), a corresponding $\chi^2$ critical value is computed.
- Compute the fraction of actual statistics in the Monte Carlo samples that are above that critical value (depicted in the $y$-axis).
- The 45-degree line is thus the appropriate reference for assessing the quality of the limiting distribution. A band about the 45-degree line corresponding to the 90% confidence region (based on the Kolmogorov-Smirnov test) is also plotted.

Given that the results obtained using the OPG and HAC weighting matrices are almost identical, Figure 2 reports the statistics only with OPG. Notice that in all cases except for the test in Figure 2b, the small sample

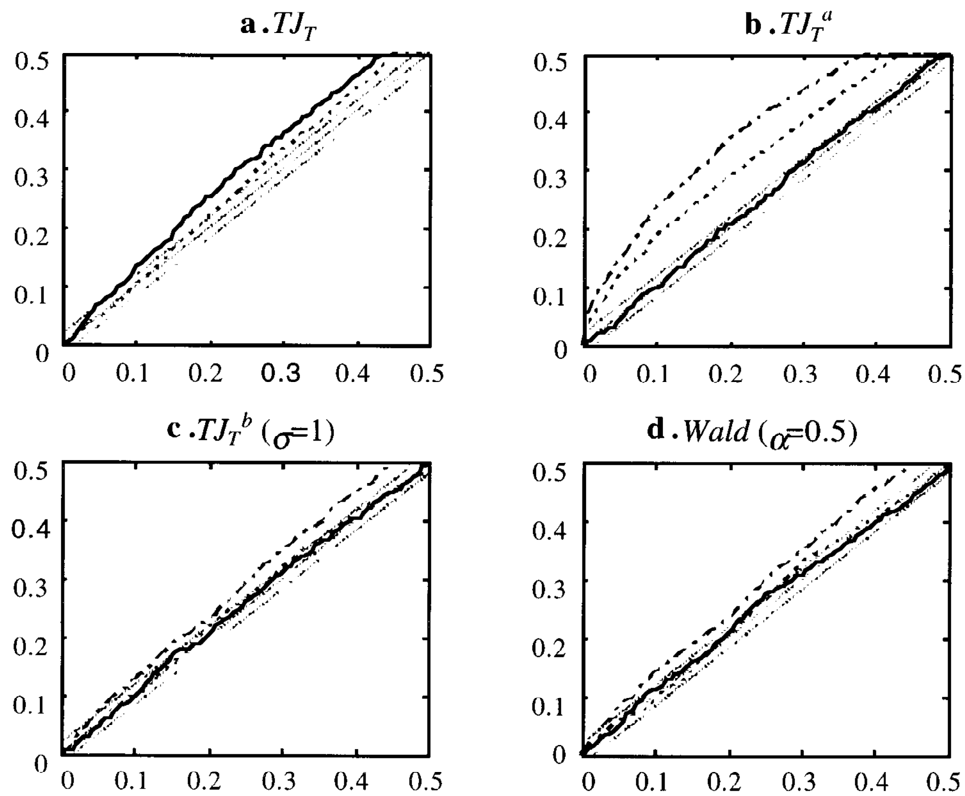


**Figure 2**
Empirical and Theoretical Distributions of Tests: MA(1) Model
The $x$-axis corresponds to the theoretical $p$-value. The $y$-axis corresponds to the empirical $p$-value. A band about the 45-degree line corresponding to the 90% confidence region based on the Kolmogorov-Smirnov test is plotted using dotted lines. **a**, The continuous line corresponds to the empirical distribution of the EMM estimator with an AR(2) auxiliary model and the OPG as weighting matrix with a $\chi^2(1)$ limiting distribution. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(3) auxiliary model and the OPG as weighting matrix with a $\chi^2(2)$ limiting distribution. **b**, The continuous line corresponds to a likelihood ratio test of the ML estimator for the same null. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(2) auxiliary model and the OPG as weighting matrix. The dash-dot line corresponds to the empirical distribution of the EMM estimator with an AR(3) auxiliary model and the OPG as weighting matrix. All have a $\chi^2(2)$ limiting distribution. **c**, The continuous line corresponds to a likelihood ratio test of the ML estimator for the same null. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(2) auxiliary model and the OPG as weighting matrix. The dash-dot line corresponds to the empirical distribution of the EMM estimator with an AR(3) auxiliary model and the OPG as weighting matrix. All have a $\chi^2(1)$ limiting distribution. **d**, The continuous line corresponds to the empirical distribution of the ML estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(2) auxiliary model and the OPG as weighting matrix. The dash-dot line corresponds to the empirical distribution of the EMM estimator with an AR(3) auxiliary model and the OPG as weighting matrix. All have a N(0, 1) limiting distribution.

distributions are not statistically different from their theoretical distributions for conventional levels of significance. The sizes of tests of hypotheses are comparable to the LRT.

These results show that, at least in the case of very simple stochastic processes, estimation and inference based on EMM produce similar results to ML estimation.

## 4 The Stochastic Volatility Model

The stochastic volatility model is frequently used for a statistical description of financial market data because of its ability to account for several features common on this type of series, such as leptokurtosis and persistent volatility. As Gallant, Hsieh, and Tauchen (forthcoming) put it, “estimation of the stochastic volatility model presents intriguing challenges, and a variety of procedures have been proposed for fitting the model," including Bayesian inference (Jacquier, Polson, and Rossi 1994), simulated likelihood (Danielson 1994), GMM (Melino and Turnbull 1990), and EMM (Gallant, Hsieh, and Tauchen forthcoming). Given its increasing popularity and widespread use, we consider a simple stochastic volatility to assess the virtues and shortcomings of EMM in a more complex environment than the one previously described.

### *4.1 The environment*

Andersen and Sorensen (1994, 1996) studied the small sample properties of GMM estimators of the stochastic volatility model, and showed their very poor performance in this environment. We use their same specifications to compare the properties of EMM with those of GMM reported there.

We consider a simple version of the lognormal stochastic volatility model of the form:

$$\begin{aligned} y_t &= h_t \varepsilon_t, \qquad \varepsilon_t \approx N(0,1) \\ \ln h_t^2 &= \omega + \delta \ln h_{t-1}^2 + \sigma u_t, \qquad u_t \approx N(0,1) \end{aligned} \tag{12}$$

The structural parameters for this exercise are given by $\rho = (\omega, \delta, \sigma)$. With the additional restrictions that $0 < \delta < 1$ and $\sigma > 0$, $y_t$ is strictly stationary and ergodic.

Following Andersen and Sorensen (1994), two specifications were chosen:

$$\begin{aligned} (\omega, \delta, \sigma) &= (-0.368, 0.95, 0.260) \\ (\omega, \delta, \sigma) &= (-0.147, 0.98, 0.166) \end{aligned} \tag{13}$$

For each sample size ($T = 500$, 1,000, 2,000), 1,000 samples were estimated.[2]

As discussed previously, to estimate this type of model with EMM, we need to select an auxiliary model that provides a good approximation to the DGP of the series. The model chosen in this case is a modified version of the Exponential ARCH (EARCH) model. To take into account the "fat tails" associated with Equation (12), the error was assumed to follow a $t$ distribution, given that a normal distribution would not be able to identify the structural parameters.

Thus, the log likelihood is:

$$\begin{aligned} \sum_{t=1}^{T} \ln f(y_t \mid x_t, \theta) \;=\; & T \ln \left\{ \frac{\Gamma[(v+1)/2]}{\pi^{1/2}\Gamma(v/2)} (v-2)^{-1/2} \right\} - (1/2) \sum_{t=1}^{T} \ln w_t \\ & - [(v+1)/2] \sum_{t=1}^{T} \ln \left[ 1 + \frac{(y_t - x_t\varphi)^2}{w_t(v-2)} \right] \end{aligned} \tag{14}$$

where $x_t$ denotes the appropriate lags of $y_t$ concatenated with a column vector of ones, $\Gamma(\cdot)$ is the gamma function, $v$ is the parameter that represents the degrees of freedom, and

$$\ln w_t = \zeta + \alpha_1 (y_{t-1} - x'_{t-1}\varphi)^2 + \cdots + \alpha_m (y_{t-m} - x'_{t-m}\varphi)^2 \tag{15}$$

The function presented in Equation (14) is numerically maximized in order to find the parameters corresponding to the AR part ($\varphi$), the exponential ARCH part described in Equation (15), and the degrees of freedom ($v$) subject to the constraint $v > 2$.[3] Thus, for this example we have that the parameters of the

[2]The settings chosen in Andersen and Sorensen also include a third specification, where the value of the autoregressive component of Equation (12) is further away from the unit circle. We did not use this third specification, for two reasons. First, the performance of GMM in Andersen and Sorensen (1994) deteriorates rapidly when the autoregressive component approaches unity, thus to compare both methods it is better to choose the most demanding settings. The second reason is entirely practical, given that each experiment takes approximately three weeks to run.

[3]See Hamilton (1994) for details.

auxiliary model are $\theta = (\zeta, \alpha, \varphi, \upsilon)$. The moment conditions described in Equation (4) were obtained by finding the numerical derivatives of Equations (14) and (15). For the experiments outlined above, an AR(1), ARCH(1) model was selected.[4]

Before we introduce the results of the Monte Carlo experiment, it is useful to describe in a more systematic fashion the steps followed in this exercise.

#### 4.1.1 Steps followed in the Monte Carlo Experiment

1. For each sample $i = 1, \ldots, 1{,}000$, generate two samples of iid N(0, 1) variables denoted $\varepsilon_t$, $u_t$, for $t = 1, \ldots, T$; where $T = 500, 1{,}000, 2{,}000$.
2. For each specification presented in Equation (13), use the realizations of $\varepsilon$ and $u$ to obtain a series for $y_t$, for $t = 1, \ldots, T$; $T = 500, 1{,}000, 2{,}000$.
3. Obtain the estimates of the auxiliary model for $y_t$ by maximizing the log likelihood described in Equations (14) and (15).
4. After the auxiliary parameters are obtained, the EMM procedure begins by generating two samples of iid N(0, 1) random numbers of sample size 10,000 (denote them $Z_j^1$, $u_j^1$; $j = 1, \ldots, 10{,}000$). These samples should not change along the estimation of the structural parameters for the sample. With an initial guess of the structural parameters, use $Z^1$ and $u^1$ to generate a realization of a pseudo-realization of $y$.
5. Evaluate the scores of the log likelihood of the auxiliary parameters with the pseudo-realization of $y$ obtained in 4).
6. Build the moment conditions described in Equation (4) with the gradients obtained in 5) and obtain the value of the objective function of Equation (5) with Equation (6) for the values of the structural parameters chosen.
7. Repeat steps 4) through 6) until convergence is achieved.
8. Return to step 1) to repeat the procedure for another sample.

### *4.2 The results*

As could be expected, if the auxiliary model presents a reasonably good approximation to the DGP of the data, Equation (6) should in fact be a good choice of the weighting matrix, because in order to obtain identification of the structural model, any pattern of heteroskedasticity or autocorrelation that could be inferred from the data should have been taken into account in the auxiliary model. For this reason, the estimation of an HAC weighting matrix such as the one proposed in Andrews (1991) is not necessary.[5]

One major advantage using EMM with respect to GMM was the numerical stability of the procedure. Andersen and Sorensen (1994) report that regardless of the parameter setting, sample size, and number of moments to be matched, the algorithm diverged a significant number of times, particularly when the autoregressive parameter approached the unit circle.[6] This was not the case with EMM, where there was no single case in which the procedure diverged, thus providing a clear advantage in terms of numerical stability.

The results of the Monte Carlo study for each specification presented in Equation (13) are displayed in Tables 2 and 3 and in Figures 3 and 4. They show that EMM unequivocally outperforms GMM in several

[4]Even though the structural model does not present autocorrelation, we include an AR(1) component in the auxiliary model to identify the structural parameters. This auxiliary model, or one in which the AR(1) component is not included (while including a constant in the level), generate almost identical results.

[5]HAC weighting matrices were also used with almost identical results.

[6]There were cases where in order to obtain 1,000 usable estimates, they had to generate in excess of 2,000 artificial samples.

**Table 2**
Properties of Estimators: Stochastic Volatility Model

| Estimator | $\omega = -0.368$ | | | $\delta = 0.95$ | | | $\sigma = 0.260$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Standard Deviation | RMSE | Mean | Standard Deviation | RMSE | Mean | Standard Deviation | RMSE |
| ***T* = 500** | | | | | | | | | |
| GMM | −0.635 | 1.139 | 1.170 | 0.919 | 0.106 | 0.110 | 0.218 | 0.186 | 0.191 |
| EMM | −0.648 | 0.389 | 0.476 | 0.913 | 0.052 | 0.063 | 0.298 | 0.156 | 0.161 |
| ***T* = 1000** | | | | | | | | | |
| GMM | −0.304 | 0.247 | 0.255 | 0.959 | 0.033 | 0.034 | 0.179 | 0.079 | 0.113 |
| EMM | −0.378 | 0.215 | 0.215 | 0.938 | 0.024 | 0.027 | 0.274 | 0.102 | 0.103 |
| ***T* = 2000** | | | | | | | | | |
| GMM | −0.286 | 0.179 | 0.197 | 0.961 | 0.025 | 0.027 | 0.190 | 0.070 | 0.099 |
| EMM | −0.371 | 0.095 | 0.095 | 0.951 | 0.017 | 0.017 | 0.262 | 0.046 | 0.046 |

Notes: The results were obtained by estimating 1,000 samples, each of size $T$. RMSE = Root Mean Square Error. GMM = results obtained with the GMM reported on Tables 15 and 16 of Andersen and Sorensen (1994). EMM = Results from EMM with an AR(1), ARCH(1), $t$-type auxiliary model, and the OPG weighting matrix.

**Table 3**
Properties of Estimators: Stochastic Volatility Model

| Estimator | $\omega = -0.147$ | | | $\delta = 0.98$ | | | $\sigma = 0.166$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Standard Deviation | RMSE | Mean | Standard Deviation | RMSE | Mean | Standard Deviation | RMSE |
| ***T* = 500** | | | | | | | | | |
| GMM | NR | NR | NR | NR | NR | NR | NR | NR | NR |
| EMM | −0.420 | 0.287 | 0.395 | 0.943 | 0.039 | 0.053 | 0.262 | 0.096 | 0.135 |
| ***T* = 1000** | | | | | | | | | |
| GMM | NR | NR | NR | NR | NR | NR | NR | NR | NR |
| EMM | −0.224 | 0.130 | 0.151 | 0.968 | 0.012 | 0.017 | 0.193 | 0.103 | 0.106 |
| ***T* = 2000** | | | | | | | | | |
| GMM | −0.140 | 0.112 | 0.112 | 0.981 | 0.015 | 0.015 | 0.125 | 0.054 | 0.068 |
| EMM | −0.167 | 0.069 | 0.072 | 0.979 | 0.009 | 0.009 | 0.163 | 0.035 | 0.035 |

Notes: The results were obtained by estimating 1,000 samples, each of size $T$. RMSE = Root Mean Square Error. GMM = results obtained with the GMM reported on Table 17 of Andersen and Sorensen (1994). NR = Values not reported by Andersen and Sorensen. EMM = Results from EMM with an AR(1), ARCH(1), $t$-type auxiliary model, and the OPG weighting matrix.

dimensions. We enumerate some of them:

- EMM provides spectacular gains in efficiency as compared to ad-hoc GMM. The idea of selecting an auxiliary model that captures the basic features of the DGP of a variable and using its gradient as the moments to match indicate a very significant improvement in the efficiency of the estimation as measured by the RMSE.
- In both specifications, the efficiency gained is remarkable. Even in the case where the autoregressive coefficient is very near the unit circle, the ratio among RMSE is $1.6:1$. Even greater improvements are observed when the data shows less persistence. Depending on the configuration of the parameters and the sample size, the RMSE of GMM can be more than two times greater than the RMSE of EMM. It is also important to notice that EMM is able to provide unbiased and more efficient estimators for all sample sizes.
- The RMSE declines exponentially with the sample size, in sharp contrast to GMM.
- There is a practical reason why EMM may be preferable to GMM, namely, its numerical stability: there was no case in which convergence was not achieved. Of course, there is a trade-off with computer time used.
- The unconditional distributions of the parameters estimated by EMM are well approximated by normal densities (except for the case of $\sigma$).

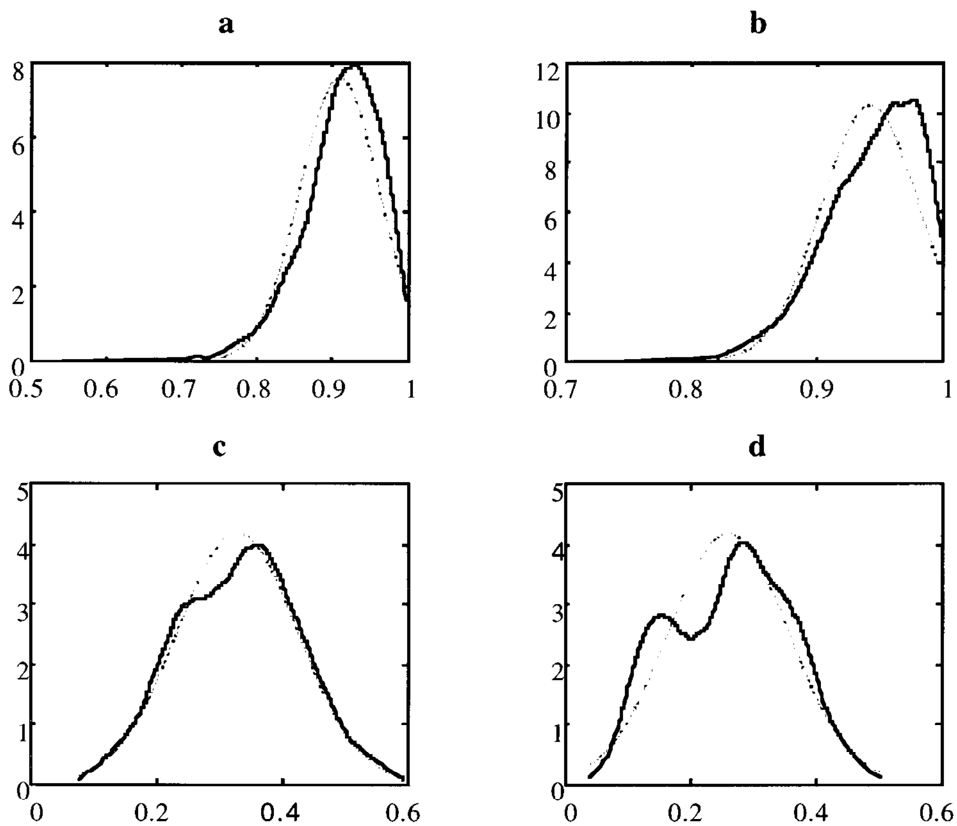


**Figure 3**
Estimated Probability Density Functions for the Stochastic Volatility Model ($T = 500$)
The continuous lines correspond to the Epanechnikov kernel estimator of the pdfs using the bandwidth proposed by Silverman (1986). The dotted lines correspond to the normal pdf with the same mean and variance of the EMM estimators. **a**, EMM estimator of $\delta$ in the first experiment. **b**, EMM estimator of $\delta$ in the second experiment. **c**, Estimator of $\sigma$ in the first experiment. **d**, Estimator of $\sigma$ in the second experiment.

- Nevertheless, Figure 4 makes it clear that a “characteristic” of the GMM objective function documented in all GMM Monte Carlo studies is inherited by EMM. Inference based on the overidentifying restrictions test as well as other $\chi^2$ statistics present important over rejections.

## 5 Consumption-Based Asset-Pricing Model

The last environment in which we assess the small-sample properties of EMM is the familiar consumption-based asset-pricing model (CCAPM), in which the parameters of the representative agent’s utility are estimated. This is the environment that received more attention in the GMM literature; see e.g. (Tauchen 1986; Kocherlakota 1990; Hansen, Heaton, and Yaron 1995). This is also the environment in which EMM may be more useful, because with this technique it does not impose as many restrictions as GMM in terms of the variables that the econometrician needs to observe in order to recover the structural parameters of the economy (Bansal, et al. 1995; Chumacero 1995).

### *5.1 The environment*

The CCAPM that is estimated in this section constitutes a simplified version of the estimated models studied by Hansen, Heaton, and Yaron (1995), but we use the same preference parameters of that study.

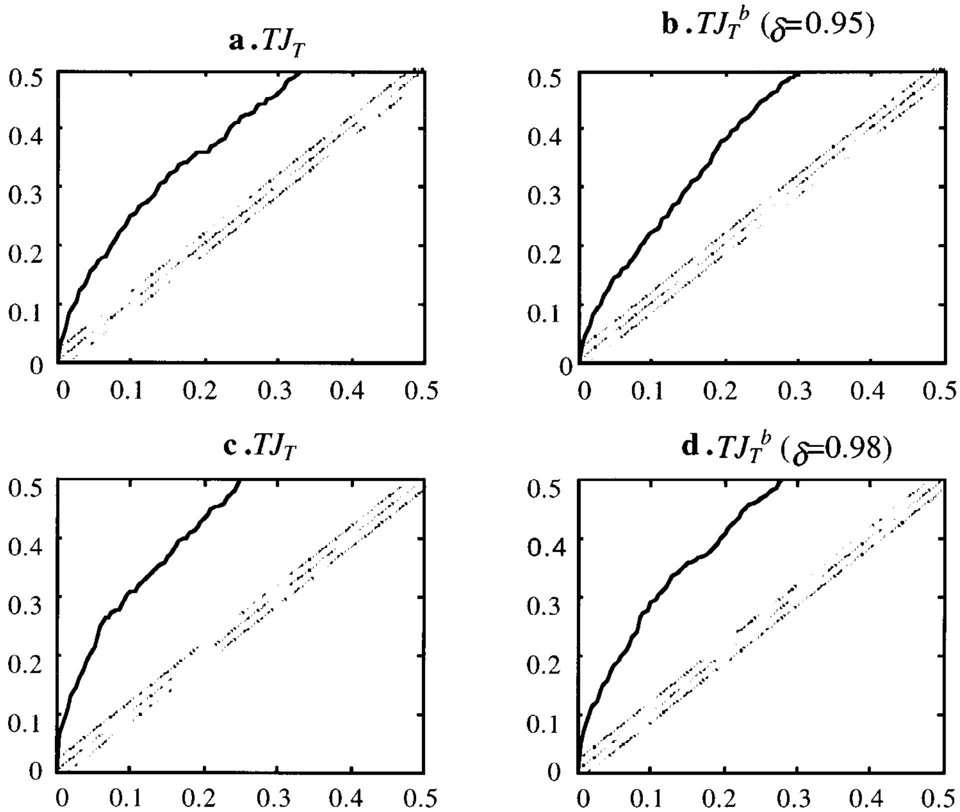


**Figure 4**
Empirical and Theoretical Distributions of Tests: Stochastic Volatility Model ($T = 500$)
The $x$-axis corresponds to the theoretical $p$-value. The $y$-axis corresponds to the empirical $p$-value. A band about the 45-degree line corresponding to the 90% confidence region based on the Kolmogorov-Smirnov test is plotted using dotted lines. **a**, $TJ_T$. Empirical distribution of the EMM estimator in the first specification with a $\chi^2(5)$ limiting distribution. **b**, $TJ_T^b$ ($\delta = 0.95$). Empirical distribution of the EMM estimator in the first specification with a $\chi^2(1)$ limiting distribution. **c**, $TJ_T$. Empirical distribution of the EMM estimator in the second specification with a $\chi^2(5)$ limiting distribution. **d**, $TJ_T^b$ ($\delta = 0.98$). Empirical distribution of the EMM estimator in the second specification with a $\chi^2(1)$ limiting distribution.

Consider a representative agent that is assumed to have preferences over consumption, given by:

$$U_0 = E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}, \qquad \gamma > 0 \tag{16}$$

where $c_t$ is consumption at date $t$. The agent can buy and sell real risk-free bonds ($b$) at price $q$ that are in zero net supply, and pay off one unit of consumption in the next period. Thus, the budget constraint that the agent faces at time $t$ is given by:

$$c_t + q_t b_{t+1} \leq y_t + b_t \qquad t = 0, 1, \ldots, \tag{17}$$

where $y_t$ is the endowment at date $t$. From this very simple setup, the familiar Euler equation that determines the price of the risk-free bond arises (after the equilibrium conditions are imposed):

$$q_t - E_t \left[ \beta \left( \frac{y_{t+1}}{y_t} \right)^{-\gamma} \right] = 0 \tag{18}$$

It is further assumed that the rate of growth of the endowment follows a first-order autoregressive process that is consistent with the first-order VAR for consumption and dividend growth described in Kocherlakota

**Table 4**
Properties of Estimators: CCAPM Model

| Estimator | $\beta = 0.97$ | | | $\gamma = 1.3$ | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Standard Deviation** | **RMSE** | **Mean** | **Standard Deviation** | **RMSE** |
| GMM | 0.964 | 0.030 | 0.031 | 0.994 | 1.030 | 1.074 |
| EMM | 0.971 | 0.003 | 0.004 | 1.284 | 0.101 | 0.102 |
| | $\beta = 1.139$ | | | $\gamma = 13.7$ | | |
| GMM | 2.322 | 0.974 | 1.532 | 8.848 | 12.054 | 12.988 |
| EMM | 1.234 | 0.236 | 0.254 | 13.347 | 1.129 | 1.182 |

Notes: The results were obtained by estimating 1,000 samples, each of size 100. RMSE = Root Mean Square Error. GMM = Results obtained with the GMM estimator. EMM = Results from EMM with an AR(1) auxiliary model and the OPG as weighting matrix.

(1990) and Hansen, Heaton, and Yaron (1995):

$$\ln\left(\frac{y_t}{y_{t-1}}\right) = 0.013 + 0.178 \ln\left(\frac{y_{t-1}}{y_{t-2}}\right) + \varepsilon_t, \qquad \varepsilon_t \approx N(0, 0.0144) \tag{19}$$

To solve Equation (18), we consider the discrete-state Markov chain model used by Tauchen (1986) and Kocherlakota (1990). The first-order Markov process that approximates Equation (19) is chosen to have 10 states, and we use Tauchen and Hussey's (1991) discretization method. As in Hansen, Heaton, and Yaron (1995), we consider two specifications for the preference parameters:

$$\begin{aligned}(\beta, \gamma) &= (0.97, 1.3)\\ (\beta, \gamma) &= (1.139, 13.7)\end{aligned} \tag{20}$$

In implementing the GMM estimators, Equation (18) is multiplied by three instrumental variables to form the moment conditions (a constant, and the first lags of the return of the risk-free bond and the growth rate of the endowment). A two-step weighting matrix was employed using the quadratic spectrum kernel and a VAR(1) prewhitening, as described in Andrews and Monahan (1992).

In the case of EMM, an AR(1) process for the return of the risk-free bond was chosen, that is:

$$\frac{1}{q_t} = \alpha_0 + \alpha_1 \frac{1}{q_{t-1}} + u_t, \qquad u_t \approx N(0, \sigma^2) \tag{21}$$

The steps followed on the EMM estimation are similar to the ones described in the previous section.

Again, to make a correspondence with the notation introduced in Section 2, the structural parameters of the model are $\rho = (\beta, \gamma)$, and the parameters of the auxiliary model are $\theta = (\alpha_0, \alpha_1, \sigma)$. The GMM and EMM estimators were computed for 1,000 replications of sample size 100.

### *5.2 The results*

The results of the Monte Carlo study for each specification presented in Equation (20) are displayed in Table 4 and in Figures 5 through 8. Again, the results of this experiment confirm that EMM outperforms GMM in every aspect. In particular:

- EMM provides unbiased estimators of the structural parameters, while GMM does not. In particular, as noted by Kocherlakota (1990), GMM seriously underestimates the relative risk-aversion coefficient, while EMM does not.
- Regardless of the specification, GMM has an RMSE associated with the parameter $\gamma$ that is at least 10 times greater than EMM. Gains in efficiency of using EMM are also apparent in the case of $\beta$.
- The unconditional distributions of the parameters estimated by EMM are well approximated by normal densities, while GMM they are not. Figures 5 and 7 also show that EMM is much more stable and concentrated in the "correct" region.

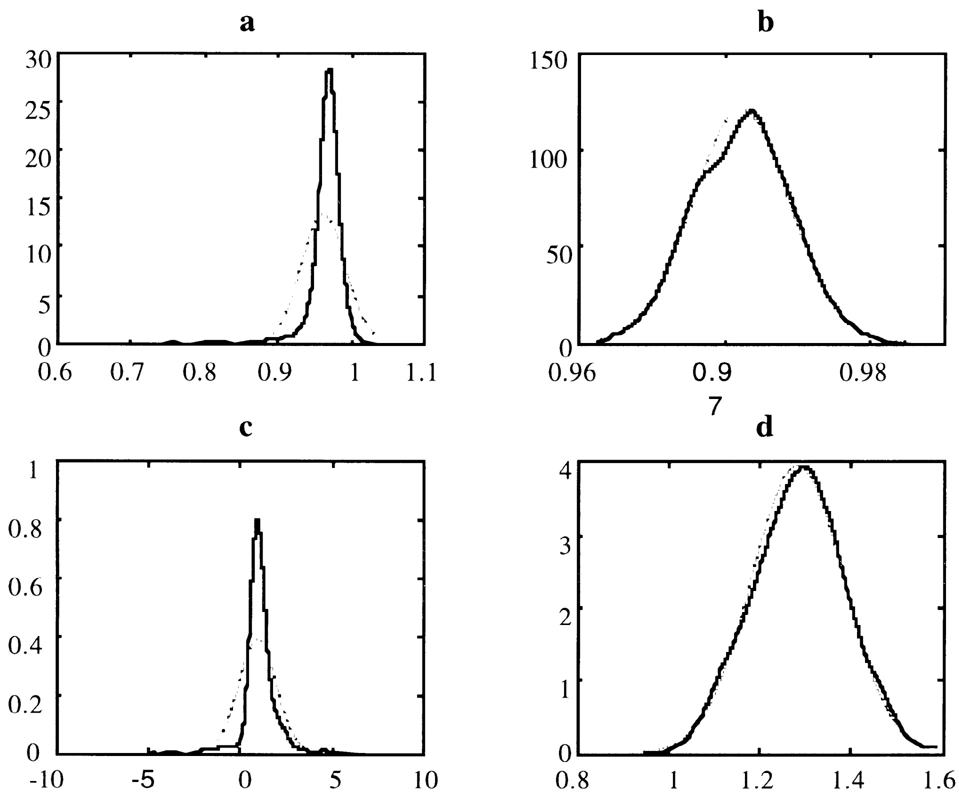


**Figure 5**
Estimated Probability Density Functions for the CCAPM Model (First Experiment)
The continuous lines correspond to Epanechnikov kernel estimators of the pdfs using the bandwidth proposed by Silverman (1986). The dotted lines correspond to the normal pdf with the same mean and variance of each of the estimators. **a**, Density of $\beta$ with the GMM estimator. **b**, Density of $\beta$ with the EMM estimator. **c**, Density of $\gamma$ with the GMM estimator. **d**, Density of $\gamma$ with the EMM estimator.

- Finally, even though inference based on Wald tests and the difference in criterion functions still present important over rejections in EMM, their magnitudes are not as important as in GMM. In particular, there is no single case in which GMM does a better job approximating the asymptotic distributions of the tests. It is also worth noticing that for this simple environment, the empirical size of the overidentifying restrictions test is very close to its theoretical counterpart for standard levels of significance.

## 6 Concluding Remarks

This paper presents three Monte Carlo experiments with varying degrees of complexity in order to assess the finite sample properties of EMM. The first one compares it with a fully efficient procedure (Maximum Likelihood) by estimating an invertible MA process. The second and third experiments compare the finite sample properties of the EMM estimators with those of GMM by using stochastic volatility models and consumption-based asset-pricing models. The general conclusions that can be drawn from these experiments are:

- EMM outperforms GMM in all of the dimensions analyzed in the Monte Carlo experiments. It provides more reliable, unbiased, efficient, and numerically stable estimates. The gains in efficiency are substantial in all of the cases, and appear to increase with the sample size.

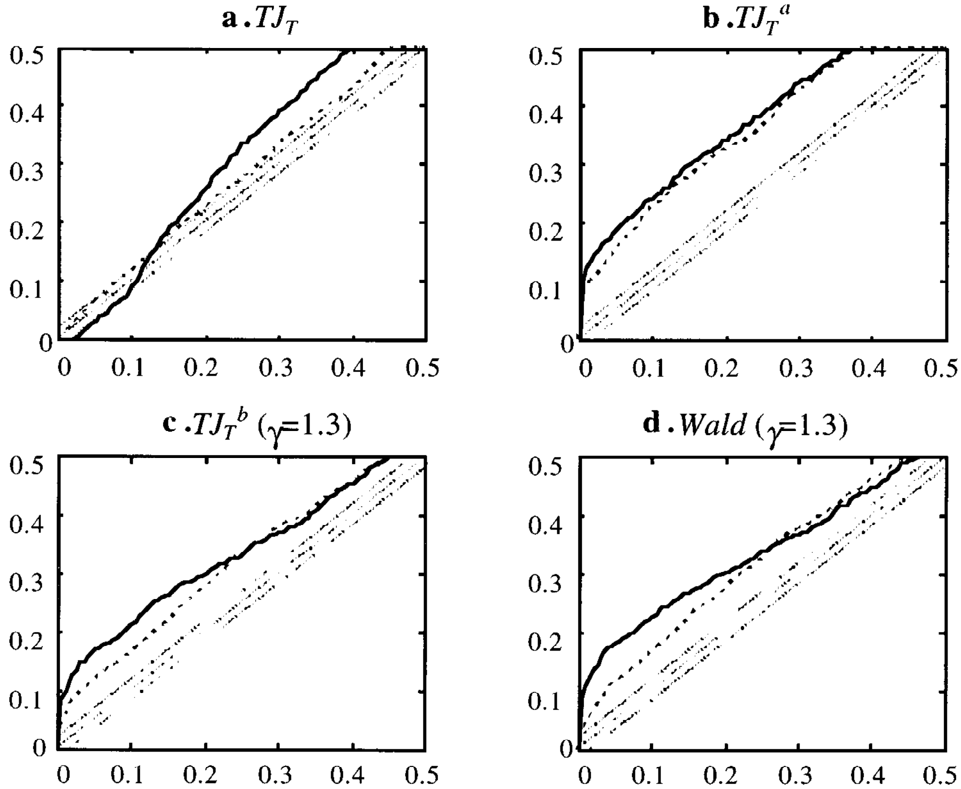


**Figure 6**
Empirical and Theoretical Distributions of Tests: CCAPM Model (First Experiment)
The $x$-axis corresponds to the theoretical $p$-value. The $y$-axis corresponds to the empirical $p$-value. A band about the 45-degree line corresponding the a 90% confidence region based on the Kolmogorov-Smirnov test is plotted using dotted lines. **a**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(1)$ limiting distribution. **b**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(2)$ limiting distribution. **c**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(1)$ limiting distribution. **d**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with an N(0, 1) limiting distribution.

- All the problems that come with the choice of weighting matrix in the case of GMM are absent in EMM when the auxiliary model is appropriately chosen.

- When the structural model to be estimated is simple, inference with EMM is as reliable as likelihood ratio tests.

- When the structural model is more complex, criterion-based inference continues to present the same problems as in GMM; however, inference based on EMM estimators is more reliable than with GMM.

In summary, EMM constitutes a very promising estimation technique for complex structural models, and "corrects" several of the problems that are present in GMM. Further research should be conducted to make criterion–function-based inference more reliable.

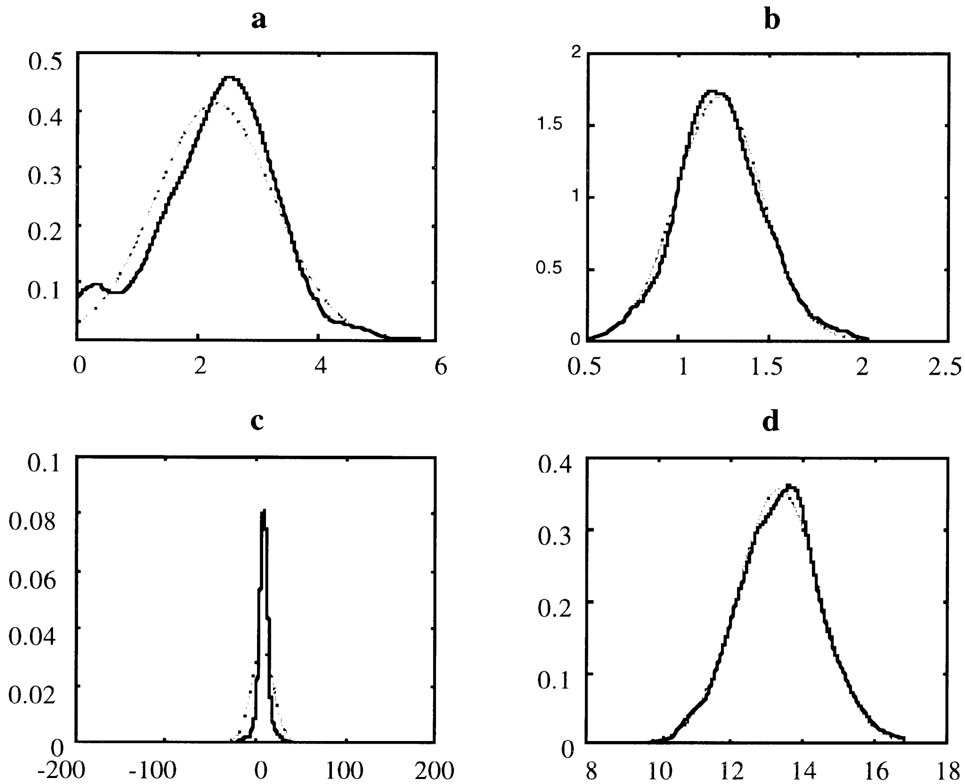


**Figure 7**
Estimated Probability Density Functions for the CCAPM Model (Second Experiment)
The continuous lines correspond to Epanechnikov kernel estimators of the pdfs using the bandwidth proposed by Silverman (1986). The dotted lines correspond to the normal pdf with the same mean and variance as each of the estimators. **a**, Density of $\beta$ with the GMM estimator. **b**, Density of $\beta$ with the EMM estimator. **c**, Density of $\gamma$ with the GMM estimator. **d**, Density of $\gamma$ with the EMM estimator.

## References


Andersen, Torben, and Bent Sorensen (1994). "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study." Working Study 175, Northwestern University. Published as: Andersen, Torben, and Bent Sorensen (1996). "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study." *Journal of Business and Economic Statistics* 14:328–352.

Andrews, Donald (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–858.

Andrews, Donald, and Christopher Monahan (1992). "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60:953–966.

Bansal, Ravi, Ronald Gallant, Robert Hussey, and George Tauchen (1994). "Computational Aspects of Nonparametric Simulation Estimation." In D. Belsley (ed.) *Computational Techniques for Econometrics and Economic Analysis.* Boston, MA: Kluwer Academic, pp. 3–22.

Bansal, Ravi, Ronald Gallant, Robert Hussey, and George Tauchen (1995). "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data." *Journal of Econometrics* 76:397–403.

Breeden, D. T. (1979). "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics* 7:262–296.

Burnside, Craig (1994). "Hansen-Jagannathan Bounds as Classical Tests of Asset-Pricing Models." *Journal of Business and Economic Statistics* 12:57–79.

Burnside, Craig, and Martin Eichenbaum (1994). "Small Sample Properties of Generalized Method of Moments-Based Wald Tests." Working Study 94-12, Federal Reserve Bank of Chicago.

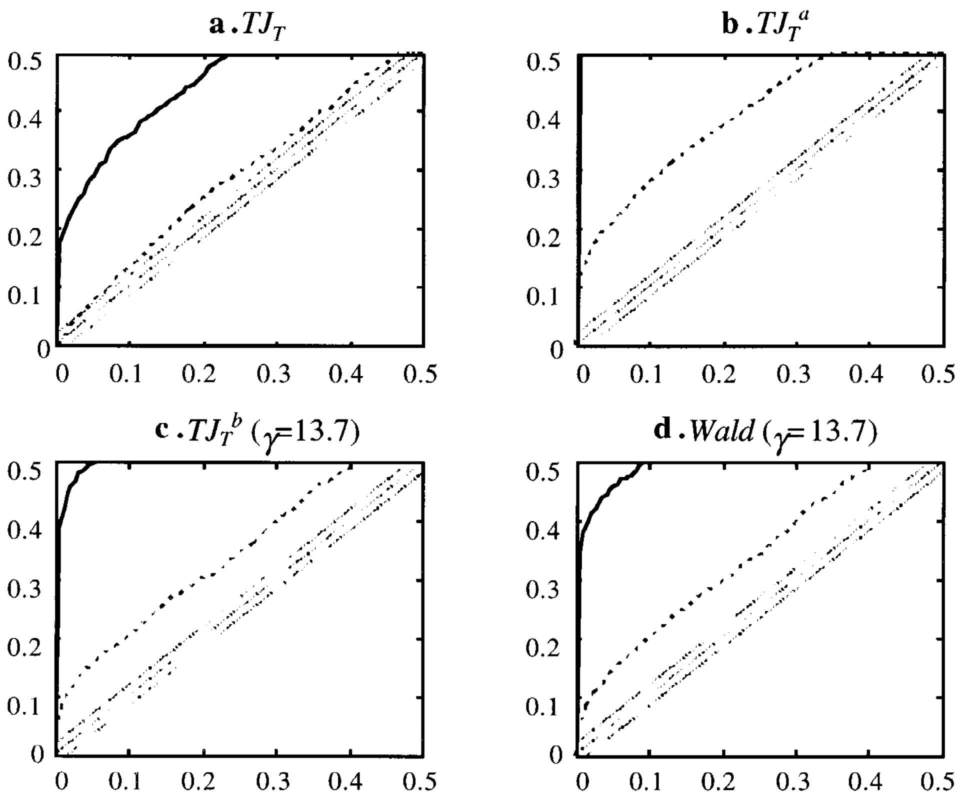


**Figure 8**
Empirical and Theoretical Distributions of Tests: CCAPM Model (Second Experiment)
The $x$-axis corresponds to the theoretical $p$-value. The $y$-axis corresponds to the empirical $p$-value. A band about the 45-degree line corresponding to the 90% confidence region based on the Kolmogorov-Smirnov test is plotted using dotted lines. **a**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(1)$ limiting distribution. **b**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(2)$ limiting distribution. **c**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with a $\chi^2(1)$ limiting distribution. **d**, The continuous line corresponds to the empirical distribution of the GMM estimator. The dashed line corresponds to the empirical distribution of the EMM estimator with an AR(1) auxiliary model and the OPG as weighting matrix, both with an N(0, 1) limiting distribution.

Christiano, Lawrence, and Wouter den Haan (1994). "Small Sample Properties of GMM for Business Cycle Analysis." Manuscript, Northwestern University.

Chumacero, Rómulo (1995). "Intertemporal Asset Pricing without Consumption Data: An Application of the EMM Estimation Technique." PhD Dissertation, Duke University.

Danielson, Jon (1994). "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood." *Journal of Econometrics* 61:375–400.

Davidson, Russell, and James MacKinnon (1994). "Graphical Methods for Investigating the Size and Power of Hypothesis Tests." Discussion Paper 903, Queens University.

Den Haan, Wouter, and Andrew Levin (1996a). "A Practitioner's Guide to Robust Covariance Matrix Estimation." Manuscript, University of California at San Diego.

Den Haan, Wouter, and Andrew Levin (1996b). "Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures." Manuscript, University of California at San Diego.

Gallant, A. Ronald (1987). *Nonlinear Statistical Models*. New York: Wiley.

Gallant, A. Ronald, David Hsieh, and George Tauchen (forthcoming). "Estimation of Stochastic Volatility Models with Suggestive Diagnostics." *Journal of Econometrics*.

Gallant, A. Ronald, and George Tauchen (1996). "Which Moments to Match?" *Econometric Theory* 12:657–681.

Gourieroux, Christian, and Alain Monfort (1993). "Pseudo-Likelihood Methods." In G. Maddala, C. Rao, and H. Vinod (eds.) *Handbook of Statistics*, 11. North-Holland, pp. 335–362.

Gourieroux, Christian, Alain Monfort, and E. Renault. (1993). "Indirect Inference." *Journal of Applied Econometrics* 8:S85–S118.

Hamilton, James (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, Lars P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029–1053.

Hansen, Lars P., John Heaton, and Amir Yaron (1995). "Finite Sample Properties of Some Alternative GMM Estimators." Manuscript, University of Chicago.

Hansen, Lars P., and Kenneth Singleton (1982). "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50:1269–1286.

Jacquier, Eric, Nicholas Polson, and Peter Rossi (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics* 12:371–417.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1996). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." Manuscript, Nuffield College.

Kocherlakota, Narayana (1990). "On Tests of Representative Consumer Asset Pricing Models." *Journal of Monetary Economics* 26:285–304.

Lee, Bong-Soo, and Beth F. Ingram (1991). "Simulation Estimation of Time-Series Models." *Journal of Econometrics* 47:197–205.

Melino, A., and S. Turnbull (1990). "Pricing Foreign Currency Options with Stochastic Volatility." *Journal of Econometrics* 45:239–265.

Rao, Radhakrishna (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.

Tauchen, George (1986). "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data." *Journal of Business and Economic Statistics* 4:397–416.

Tauchen, George (1996). "New Minimum Chi-Square Methods in Empirical Finance." In D. Kreps and K. Wallis (eds.) *Advances in Economic Theory and Econometrics: Seventh World Congress*. Cambridge University Press.

Tauchen, George, and Robert Hussey (1991). "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models." *Econometrica* 59:371–396.

Thisted, Ronald (1988). *Elements of Statistical Computing*. New York: Chapman & Hall.

**Editorial Policy**

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826