Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
July 2001, Volume 5, Number 2
The MIT Press

Studies in Nonlinear Dynamics and Econometrics (E-ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142-1407. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142-1407; (617)253-2889; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $135.00. Canadians add additional 7% GST. Submit all claims to: journals-claims@mit.edu. Prices subject to change without notice.

Estimating ARMA Models Efficiently

Rómulo A. Chumacero
Department of Economics
University of Chile
and
Research Department
Central Bank of Chile
rchumace@econ.uchile.cl

Abstract. *This paper presents the asymptotic and finite sample properties of the efficient method of moments and indirect inference, when applied to estimating stationary ARMA models. Issues such as identification, model selection, and testing are also discussed. The properties of these estimators are compared to those of maximum likelihood using Monte Carlo experiments for both invertible and noninvertible ARMA models.*

Keywords. Monte Carlo, efficient method of moments, indirect inference, ARMA, identification, model selection

Acknowledgments. This research received support from the Research Fund of the Department of Economics of the University of Chile. I would like to thank Miguel Basch, Bruce Mizrach, George Tauchen, the participants in the Economics Workshop at the University of Chile, the Fifth International Conference of the Society for Computational Economics at Boston College, the 17th Meeting of the Latin American Econometric Society, and an anonymous reviewer for helpful comments. The usual disclaimer applies.

1 Introduction

There is a long-standing tradition of estimating stationary ARMA models using likelihood-based methods, the estimation of a sequence of long autoregressions, or nonlinear least squares.[1] Gallant and Tauchen (1996) developed a kind of minimum chi-square estimator (called the efficient method of moments, or EMM) that is also suitable for estimating stationary ARMA models. In fact, Gouriéroux, Monfort, and Renault (1993) estimate an invertible MA(1) model using a method that is similar in spirit to EMM (indirect inference, or II).

Nevertheless, simulation-based methods (such as EMM and II) are not routinely applied to estimate ARMA models because of evident shortcomings: first, the maximum-likelihood estimator (MLE) performs efficiently (in the root-mean-square error [RMSE] sense), so it is not clear why less-efficient alternative methods should be considered. Second, simulation-based methods are costly to use, because they require more computer time than simpler alternatives. Finally, tests based on simulation-based estimators may be cumbersome and, as reported elsewhere (Chumacero 1997), may present important size problems.

The conventional view, therefore, is that although EMM and II may be useful in different setups (particularly when the alternative is the conventional method of moments or when MLE is unfeasible), they do not appear to offer any practical advantage when estimating ARMA models.

[1] See Ghysels, Khalaf, and Vodounou 1994 or Galbraith and Zinde-Walsh 1994 or 1997 for references.

This article challenges that view by showing that if the asymptotic properties of both the EMM and II moment conditions are exploited, computationally efficient algorithms can be developed to estimate ARMA models. In this sense, efficiency could therefore be defined more broadly. The article also presents other contributions that may be useful to EMM and II users in setups other than those considered here. In particular, issues such as identification, testing, and model selection are explicitly discussed.

The article is organized as follows. Section 2 presents a brief description of EMM and II. Section 3 describes the class of estimators considered, discusses the issue of identification, and derives the asymptotic properties of the estimators. Section 4 presents the results of Monte Carlo experiments to assess the finite-sample properties of the estimators described in the previous section. Finally, Section 5 summarizes the main findings.

2 The Estimation Methods

Consider a stationary stochastic process $p(y_t|x_t, \rho)$, describing y_t in terms of exogenous variables (x_t) and structural parameters (ρ), which the econometrician is interested in estimating. Consider also an auxiliary model $f(y_t|x_t, \theta)$ that can be expressed analytically, whereas $p(y_t|x_t, \rho)$ cannot. Gallant and Tauchen (1996) proposed using the scores of the auxiliary model:

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

to generate the generalized method of moments (GMM) moment conditions

$$m_T(\rho, \hat{\theta}_T) = \int \int (\partial/\partial\theta) \ln f(y_t|x_t, \hat{\theta}_T)\, p(y|x, \rho)\, dy\, p(x|\rho)\, dx \tag{1}$$

where $\hat{\theta}_T$ is defined as the MLE of $f(\cdot)$ for a sample of size T; that is:

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \ln f(y_t|x_t, \hat{\theta}_T) \tag{2}$$

When analytical expressions for (1) are not available, simulations may be required to compute them, in which case we approximate the moments by

$$m_T(\rho, \hat{\theta}_T) \cong m_N(\rho, \hat{\theta}_T) = \frac{1}{N} \sum_{n=1}^{N} (\partial/\partial\theta) \ln f(\tilde{y}_n(\rho)|\tilde{x}_n(\rho), \hat{\theta}_T) \tag{3}$$

where N is the sample size of the Monte Carlo integral approximation drawn from a sample of y and x for a given value of ρ. When (3) is used to approximate the moments, the GMM estimator of ρ, with an efficient weighting matrix, is given by

$$\hat{\rho} = \arg\min_{\rho \in R} m'_N(\rho, \hat{\theta}_T)(\hat{I}_T)^{-1} m_N(\rho, \hat{\theta}_T) \tag{4}$$

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

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^{T} [(\partial/\partial\theta \ln f(y_t|x_t, \hat{\theta}_T))][(\partial/\partial\theta \ln f(y_t|x_t, \hat{\theta}_T))]' \tag{5}$$

Gallant and Tauchen (1996) demonstrated the strong convergence and asymptotic normality of the estimator presented in (4):[2]

$$\sqrt{T}(\hat{\rho} - \rho_0) \xrightarrow{D} N[0, (d'_\rho I^{-1} d_\rho)^{-1}]$$

where $d_\rho = \partial m(\rho_0, \theta_0)/\partial\rho'$, $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$, and $\hat{I}_T \xrightarrow{a.s.} I$.

[2]See Tauchen 1996 for a step-by-step derivation of these results.

By standard arguments, the asymptotic distribution of the objective function that $\hat{\rho}$ minimizes is given by

$$TJ_T = Tm'_N(\hat{\rho}, \hat{\theta}_T)(\hat{I}_T)^{-1} m_N(\hat{\rho}, \hat{\theta}_T) \xrightarrow{D} \chi^2_{h-r} \tag{6}$$

with r and h denoting the dimensions of ρ and θ, respectively.

Equation (6) corresponds to the familiar overidentifying restrictions test described by Hansen (1982). As in GMM, the order condition for identification requires that $h \geq r$. The rank condition is more involved, given that in this setup we require the existence of a unique function linking ρ and θ in a sense that will be defined more precisely below.

Given the results described, and provided identification conditions are met, statistical inference may be carried out the same way as in GMM. Depending on the complexity of the auxiliary model, however, it may be difficult to construct Wald-type tests based on the variance-covariance matrix obtained by differentiating the moments (Chumacero 1997).

If (1) can be obtained analytically, all the expressions using $m_N(\cdot)$ should be replaced by $m_T(\cdot)$. As the next section indicates, a simple analytical expression for (1) is available when estimating Gaussian ARMA models, thus making simulation-based methods for computing (3) both unnecessarily costly and inefficient.

The II estimator is similar to EMM, the main difference being the choice of moment conditions. II mimics the optimization underlying (2), instead of the first-order conditions, in which case we define

$$\tilde{\theta}_N(\rho) = \arg\max_{\theta \in \Theta} \sum_{n=1}^{N} \ln f(\tilde{y}_n(\rho) | \tilde{x}_n(\rho), \theta) \tag{7}$$

That is, we find the MLE of the auxiliary model for a given value of ρ and an artificial realization of size N of y and x. The moment conditions that II uses are given by (8) instead of (3):

$$m_N(\rho, \hat{\theta}_T) = [\tilde{\theta}_N(\rho) - \hat{\theta}_T] \tag{8}$$

There is a major difference between EMM and II. Whereas the former carries out only one optimization of the auxiliary model and uses the parameters estimated there while evaluating the scores with simulated data, II requires optimization of the auxiliary model for each value of the structural model under consideration, making it computationally more demanding.

In the next section, we show that in the case of Gaussian ARMA models, simulations are not required, because there is an analytical expression for the relationship between the parameters of the auxiliary and structural models.

3 Estimating ARMA Models with EMM and II

3.1 The general case

Consider a stationary Gaussian ARMA(p, q) model with $q \neq 0$ that we are interested in estimating, and denote by $\gamma_i (i = 0, 1, \ldots)$ its autocovariances:

$$y_t = \sum_{i=1}^{p} \delta_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

The auxiliary model used for estimating this process is given by

$$y_t = \sum_{i=1}^{j} \beta_i y_{t-i} + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

So that we may remain consistent with the notation system used in the previous section, the structural parameters of this model are $\rho = (\delta_1, \ldots, \delta_p, \alpha_1, \ldots, \alpha_q, \sigma^2)$, and the auxiliary model's parameters are $\theta = (\beta_1, \beta_2, \ldots, \beta_j, \sigma_v^2)$.

The order condition for identification requires $j \geq p + q$. The rank condition can be studied by evaluating the asymptotic properties of the estimators of the AR(j) auxiliary model.

In this case, it is a simple matter to verify that

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_j \end{bmatrix} \overset{a.s.}{\rightarrow} \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{j-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{j-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{j-1} & \gamma_{j-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_j \end{bmatrix} = \begin{bmatrix} \beta_{0,1} \\ \beta_{0,2} \\ \vdots \\ \beta_{0,j} \end{bmatrix} = \beta_0$$

$$\hat{\sigma}_v^2 = \frac{\sum_{t=j+1}^{T} \left(y_t - \sum_{i=1}^{j} \hat{\beta}_i y_{t-i} \right)^2}{T - j} \overset{a.s.}{\rightarrow} \gamma_0 \left(1 + \sum_{i=1}^{j} \beta_{0,i}^2 \right) + 2 \sum_{i=1}^{j} \gamma_i \left(\sum_{l=1}^{j-i} \beta_{0,l} \beta_{0,i+l} - \beta_{0,i} \right) \tag{9}$$

Thus, all the auxiliary model's parameters are (asymptotically speaking) a nonlinear function of the first j autocovariances of y. As (9) makes clear, approximating moment conditions for II is unnecessary, given that this equation can be used directly instead of (7).

In the case of EMM, identification requires a unique set of parameters in the structural model that, accompanied with the consistent auxiliary model estimators, make (1) equal to 0. As will be shown shortly, this condition is equivalent to the identifiability condition that must also be imposed for estimating ARMA models with maximum likelihood (ML) (see Hamilton 1994 or Tanaka 1996). Given that for every invertible stationary ARMA model there is an observationally equivalent noninvertible ARMA model, the same requirements imposed on any stationary ARMA model estimated using maximum likelihood should apply when estimating with EMM or II.

As the structural model is assumed to be Gaussian, the scores of the auxiliary model can be expressed as

$$m_T(\rho, \hat{\theta}_T) = \begin{bmatrix} \dfrac{\left(\gamma_i + \sum_{l=1}^{i} \hat{\beta}_l \gamma_{i-l} + \sum_{l=i+1}^{j} \hat{\beta}_l \gamma_{l-i} \right)}{\hat{\sigma}_v^2}, \quad i = 1, \ldots, j, \\ -\dfrac{1}{2\hat{\sigma}_v^2} + \dfrac{\gamma_0 \left(1 + \sum_{i=1}^{j} \hat{\beta}_i^2 \right) + 2 \sum_{i=1}^{j} \gamma_i \left(\sum_{l=1}^{j-i} \hat{\beta}_l \hat{\beta}_{i+l} - \hat{\beta}_i \right)}{2\hat{\sigma}_v^4} \end{bmatrix} \tag{10}$$

where the dimension of (10) is $(j + 1) \times 1$ and corresponds to the unconditional expectation of the scores of the auxiliary model. The last moment condition is equal to zero when $\hat{\beta}$ is replaced by β_0. This is also the case for the first j moments (up to a reparameterization discussed below). As with II, approximating the moment conditions using Monte Carlo simulations (as in Equation 3) is unnecessary, because these can be derived analytically.

Gouriéroux, Monfort, and Renault (1993), Ghysels, Khalaf, and Vodounou (1994), Chumacero (1997), and Michaelides and Ng (1997) used Monte Carlo methods to approximate either (9) or (10) for the case of Gaussian MA(1) models. Nevertheless, as these equations show, the moments can be obtained directly. To estimate the parameters of the ARMA process one can simply replace the autocovariances with functions of the structural model's parameters.

Next, we apply these results to two particular processes, an MA(1) and an ARMA(1,1) and then go on to conduct Monte Carlo experiments to assess the finite sample properties of different EMM and II estimators, comparing them with those of ML.

3.2 The MA(1) model

The model that we are interested in estimating is

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

and the auxiliary model is again an AR(j) process. In this case the autocovariances are given by

$$\gamma_0 = \sigma^2(1+\alpha^2), \quad \gamma_1 = \sigma^2\alpha, \quad \gamma_i = 0, \text{ for } i > 1$$

It is trivial to verify that for any AR(j) auxiliary model, estimates converge to

$$\begin{bmatrix} \hat{\beta}_i \\ \hat{\sigma}_v^2 \end{bmatrix} \overset{a.s.}{\rightarrow} \frac{1}{1-\alpha^{2(j+1)}} \begin{bmatrix} (-1)^{i-1}\alpha^i(1-\alpha^{2(j+1-i)}) \\ \sigma^2(1-\alpha^{2(j+2)}) \end{bmatrix} \quad \text{for } i = 1, \ldots, j$$

when $|\alpha| \neq 1$, and to

$$\begin{bmatrix} \hat{\beta}_i \\ \hat{\sigma}_v^2 \end{bmatrix} \overset{a.s.}{\rightarrow} \frac{1}{j+1} \begin{bmatrix} (-1)^{i-1}\alpha^i(j+1-i) \\ \sigma^2(j+2) \end{bmatrix} \quad \text{for } i = 1, \ldots, j$$

in the unit root case. Once analytical expressions for the asymptotic values of auxiliary model's parameters are found, it is easy to verify (by direct substitution) that (9) is equal to zero.

Note, however, that given that the auxiliary model's parameters are functions of the autocovariances, there is always an invertible MA model that is observationally equivalent to a noninvertible MA model. In this case, the following MA(1) model reproduces the same autocovariances and the same estimates for the auxiliary model, thus also satisfying (10):

$$y_t = \varepsilon_t + \alpha^*\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma_*^2)$$

with $\alpha^* = 1/\alpha$ and $\sigma_*^2 = \sigma^2\alpha^2$. Thus, the only case in which EMM or II can satisfy the rank condition (exactly) is when $\alpha = \pm 1$.

3.3 The ARMA(1,1) model

The other model that we are interested in estimating is

$$y_t = \delta y_{t-1} + \varepsilon_t + \alpha\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

while the auxiliary model is once again an AR(j) process. In this case the autocovariances are given by

$$\gamma_0 = \frac{\sigma^2(1+\alpha^2+2\alpha\delta)}{1-\delta^2}, \quad \gamma_1 = \gamma_0\delta + \sigma^2\alpha, \quad \gamma_i = \delta\gamma_{i-1} \text{ for } i > 1$$

This process yields auxiliary model and moment conditions estimates expressed by (9) and (10). As with the MA(1) process discussed above, when $\alpha^* = 1/\alpha$ and $\sigma_*^2 = \sigma^2\alpha^2$ are replaced by α and σ^2, both processes are observationally equivalent. Again, identification requires a stance with respect to the invertibility of the process.

The next section develops several Monte Carlo experiments to assess the finite-sample properties of several EMM and II estimators for these two cases, discusses choosing the auxiliary model and testing, and compares results with those obtained using ML.

4 The Monte Carlo Experiments

There have been at least four Monte Carlo experiments to assess the finite-sample properties of simulation-based methods for invertible MA(1) processes. Gouriéroux, Monfort, and Renault 1993 used II with

AR(1), AR(2), and AR(3) auxiliary models, setting α equal to -0.5. Ghysels, Khalaf, and Vodounou (1994) compared several simulation-based estimators for different specifications of the invertible MA(1) but used simulations to approximate the moment conditions. Michaelides and Ng (1997) also performed Monte Carlo experiments for different sample sizes and different values of N to approximate the moment conditions in (3). Nevertheless, as their histograms show, even though they are estimating an invertible MA(1) model, they allow for estimates of α greater than 1 (in absolute value) and so do not impose the necessary identification conditions. In their study, they set the auxiliary model at AR(3). Chumacero (1997) also studied the case of the finite-sample properties of the invertible MA(1) model with fixed AR auxiliary models ((2) and (3)), along with two other Monte Carlo experiments for more complex setups, which show the superior performance of EMM over that of GMM in several counts. In a different setup, Gallant and Tauchen (1999) also presented Monte Carlo evidence showing the gains in efficiency of using EMM over the conventional method of moments estimators. These articles, however, leave a number of questions unanswered that are addressed here:

- From a practical standpoint, how should one choose the auxiliary model? Do any model selection criteria offer better results in terms of efficiency?
- Chumacero (1997) and Michaelides and Ng (1997) presented evidence that the finite-sample properties of the overidentifying-restrictions test described in (6) present problems of size. In particular, there is strong evidence of over-rejections. Is there a simple way to correct this problem and provide a better approximation of the asymptotic distribution of this test?
- How well do Wald-type tests perform under different specifications for ARMA models?
- How well do EMM and II perform when estimating noninvertible ARMA models?
- What are the gains from using the exact moment conditions instead of the simulation-based approximations?
- Is there any gain from estimating stationary ARMA models with EMM or II instead of ML?

4.1 Design of the experiments

To answer the questions posed above, two types of experiments were implemented: one for different specifications of an MA(1) model (both invertible and noninvertible), and the other for different specifications of an ARMA(1,1) model. In each case, results were compared with the properties of MLEs that were estimated using the conditional likelihood, when the true process was invertible, and the unconditional (exact) likelihood, when the process was not.

Each setup used 1,000 samples, each of size $T = 100$ and $T = 200$, values chosen to allow comparison with previous studies (particularly Ghysels, Khalaf, and Vodounou 1994 and Michaelides and Ng 1997). In the case of EMM and II, two different types of estimators were chosen, one using the exact moment conditions and the other approximating them, using $N = 2{,}500$. Finally, three selection criteria for the choice of the lag length of the auxiliary model were used. These are the Akaike information criterion (AIC), the Schwarz criterion (BIC), and the Hannan & Quinn criterion (HQ). Given that the auxiliary model assumes normality, they are defined as follows:

$$\begin{aligned}
\mathrm{AIC}(j) &= \ln \hat{\sigma}_v^2 + (2j/T) \\
\mathrm{BIC}(j) &= \ln \hat{\sigma}_v^2 + (j \ln T/T) \\
\mathrm{HQ}(j) &= \ln \hat{\sigma}_v^2 + (2j \ln(\ln T)/T) \qquad (11)
\end{aligned}$$

where j is chosen to minimize (11) in all cases.

Table 1
Properties of the estimators: MA(1) model ($T = 100$)

	EMM			II			
$\alpha = -0.5$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	***j***
ML	−0.508	0.098	1.038	−0.508	0.098	1.038	
AIC	−0.506	0.136	1.000	−0.540	0.144	1.113	2.927
AIC-N	−0.506	0.137	6.519	−0.540	0.146	2.982	2.927
$\alpha = -0.95$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	***j***
ML	−0.910	0.072	1.536	−0.910	0.072	1.536	
AIC	−0.908	0.113	1.000	−0.965	0.068	5.687	7.598
AIC-N	−0.905	0.115	7.569	−0.964	0.068	17.777	7.598
$\alpha = -1.05$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	***j***
MLX	−1.041	0.045	2.231	−1.041	0.045	2.231	
AIC	−1.116	0.161	1.000	−1.042	0.086	5.188	7.660
AIC-N	−1.119	0.165	7.775	−1.043	0.086	16.242	7.660
$\alpha = -1.5$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	***j***
MLX	−1.508	0.210	2.808	−1.508	0.210	2.808	
AIC	−1.509	0.346	1.000	−1.394	0.300	1.986	4.110
AIC-N	−1.511	0.346	6.781	−1.381	0.321	2.155	4.110

Note: The results were obtained by estimating 1,000 samples. RMSE = root-mean-square error; Time = mean of the ratio between time to convergence of a method and time to convergence of EMM using analytical moments and the AIC information criterion; j = average lag length of the auxiliary model; ML = results obtained with the conditional MLE; MLX = results obtained with the unconditional (exact) MLE; AIC = results using AIC as information criterion and the analytical moment conditions; AIC-N = results using AIC as information criterion and a numerical approximation for the moments.

Chumacero (1997) showed that the choice of weighting matrix is not as crucial in EMM as in GMM and that (5) provides results basically identical to other procedures requiring the computing of HAC matrices, when the auxiliary model is chosen correctly, in the sense of providing a good statistical description of all data features. Here, therefore, we concentrate on comparing model selection criteria described in (11) and their implications for the properties of the coefficients and the statistics commonly used for inference. In all the samples the maximum and minimum lags for the auxiliary model were set to 20 and 2 respectively. The minimum lag was set in order to force the model to be overidentified. For each sample, each expression in (11) was minimized, and j was chosen accordingly.

4.2 Results for the MA(1) model

Tables 1 to 4 report the results for different specifications of the MA(1) model,[3] comparing results obtained using EMM and II to those obtained by ML. The main findings are as follows:

- As known, BIC tends to select more parsimonious models, followed by HQ, and AIC always chooses larger auxiliary models.
- The three selection criteria provide no discernable differences in terms of bias. Nevertheless, AIC tends to provide more efficient estimates, followed by HQ, and finally BIC. This is because BIC is too conservative in the choice of lag length, particularly in cases near a unit root.
- In this simple example, a choice of $N = 2{,}500$ approximates the exact moment conditions well. Because moments can be computed exactly for any stationary ARMA model, however, these approximations are

[3]Results for alternative data-generating processes are available upon request.

Table 2
Properties of the estimators: MA(1) model ($T = 200$)

	EMM			II			
$\alpha = -0.5$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	j
ML	−0.505	0.065	2.717	−0.505	0.065	2.717	
AIC	−0.505	0.081	1.000	−0.526	0.085	1.006	3.455
BIC	−0.505	0.085	0.985	−0.518	0.089	0.982	2.159
HQ	−0.509	0.084	0.985	−0.525	0.089	0.983	2.456
$\alpha = -0.95$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	j
ML	−0.926	0.046	2.741	−0.926	0.046	2.741	
AIC	−0.936	0.069	1.000	−0.972	0.052	19.818	10.313
BIC	−0.930	0.092	0.991	−0.964	0.068	17.777	5.114
HQ	−0.938	0.076	0.976	−0.973	0.056	19.381	7.398
$\alpha = -1.05$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	j
MLX	−1.045	0.031	5.106	−1.045	0.031	5.106	
AIC	−1.072	0.084	1.000	−1.031	0.057	17.955	10.338
BIC	−1.085	0.122	1.005	−1.042	0.084	16.242	5.137
HQ	−1.073	0.096	0.993	−1.031	0.063	17.930	7.432
$\alpha = -1.5$	**Mean**	**RMSE**	**Time**	**Mean**	**RMSE**	**Time**	j
MLX	−1.501	0.133	7.993	−1.501	0.133	7.993	
AIC	−1.498	0.190	1.000	−1.432	0.184	1.407	4.873
BIC	−1.470	0.241	1.104	−1.414	0.238	2.155	2.780
HQ	−1.477	0.208	1.034	−1.417	0.202	1.666	3.516

Note: MLX = results obtained with the unconditional (exact) MLE; AIC, BIC, HQ = results using AIC, BIC, or HQ, respectively, as information criterion and the analytical moment conditions.

Table 3
Properties of the overidentifying restrictions test: MA(1) model ($T = 100$)

	EMM						II					
	TJ_T			TJ_{T-q}			TJ_T			TJ_{T-q}		
$\alpha = -0.5$	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
AIC	4.2	6.0	12.7	12.9	4.2	10.2	0.1	33.5	35.5	1.4	31.2	34.5
AIC-N	3.8	6.8	13.1	11.6	4.4	10.5	0.9	33.0	35.8	1.7	31.1	34.4
$\alpha = -0.95$	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
AIC	0.1	9.8	18.9	13.6	4.5	9.4	0.1	70.1	73.3	0.1	67.9	71.3
AIC-N	0.1	10.6	19.0	8.7	4.7	10.0	0.1	69.8	72.4	0.1	67.4	70.7
$\alpha = -1.05$	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
AIC	0.1	10.3	18.7	13.3	4.3	8.4	0.1	70.7	73.9	0.1	68.9	71.6
AIC-N	0.1	10.2	18.6	8.8	4.5	9.8	0.1	70.2	72.7	0.1	67.8	71.1
$\alpha = -1.5$	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
AIC	3.3	7.6	13.4	15.1	5.4	9.4	0.3	18.1	20.9	0.2	16.0	19.2
AIC-N	2.3	8.0	14.2	11.9	5.6	10.1	0.3	17.7	20.9	0.2	16.2	19.0

Note: TJ_T = overidentifying restrictions test; $(T-q)J_{T-q}$ = overidentifying restrictions test adjusted by degrees of freedom; $M = p$-value of the Mann-Whitney-Wilcoxon test; 5%, 10% = size of the test for the nominal counterpart; AIC = results using AIC as information criterion and the analytical moment conditions; AIC-N = results from EMM using AIC as information criterion and a numerical approximation for the moments.

Table 4
Properties of the Wald test: MA(1) model (EMM)

	$T = 100$						$T = 200$					
$\alpha = -0.5$	W_T			W_{T-q}			W_T			W_{T-q}		
Estimator	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
ML	20.0	7.2	12.6	20.0	7.2	12.6	36.6	5.2	12.0	36.6	5.2	12.0
AIC-N	15.2	8.4	13.3	19.4	8.1	13.0	13.0	7.9	13.3	15.1	7.6	12.5
AIC	17.7	9.1	13.0	22.4	8.3	12.7	19.5	6.7	11.5	22.3	6.5	11.2
BIC-N	47.4	6.3	11.0	48.1	6.0	10.9	25.5	6.5	10.8	27.5	6.1	10.7
BIC	49.1	6.8	10.6	44.5	6.6	10.6	39.2	5.0	8.8	41.7	4.8	8.8
HQ-N	33.4	6.4	11.5	38.1	6.2	11.5	17.9	6.7	11.4	19.7	6.5	11.3
HQ	36.0	7.3	11.3	40.9	7.0	11.3	26.9	5.3	9.6	29.4	5.2	9.6
$\alpha = -1.5$	W_T			W_{T-q}			W_T			W_{T-q}		
Estimator	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**	**M**	**5%**	**10%**
MLX	38.7	8.6	11.1	38.7	8.6	11.1	47.2	7.5	11.3	47.2	7.5	11.3
AIC-N	8.0	1.6	4.8	5.9	1.2	3.5	48.3	4.9	10.7	48.1	4.8	9.4
AIC	7.5	1.8	5.0	5.6	1.2	4.4	15.4	4.2	9.0	41.8	3.9	8.8
BIC-N	0.2	0.3	1.9	0.2	0.3	1.7	5.8	1.5	4.5	5.2	1.2	4.0
BIC	0.2	0.3	2.0	0.2	0.3	1.8	4.6	0.9	3.9	4.0	0.9	3.8
HQ-N	0.6	0.5	2.4	0.4	0.4	1.9	23.1	2.8	6.3	21.0	2.4	5.5
HQ	0.6	0.5	2.3	0.5	0.5	2.0	15.9	2.3	6.1	14.3	2.1	5.9

Note: W_T = Wald test for the null; W_{T-q} = Wald test for the null adjusted by degrees of freedom; M = p-value of the Mann-Whitney-Wilcoxon Test; 5%, 10% = size of the test for the nominal counterpart; ML = results obtained using the conditional MLE; MLX = results obtained using the unconditional (exact) MLE; AIC-N, BIC-N, HQ-N = results from EMM using AIC, BIC, or HQ, respectively, as information criterion and a numerical approximation for the moments ($N = 2{,}500$); AIC, BIC, HQ = results from EMM using AIC, BIC, or HQ, respectively, as information criterion and the exact moment conditions.

unnecessary.[4] Aside from the fact that using the exact moment conditions yields better finite sample results, there is also a practical reason to do so: the gains in computing time are substantial. In fact, even in this simple setup, and even for moderate sample sizes, obtaining EMM estimates using the exact moment conditions can be up to eight times faster than using ML. Although inefficient (in the RMSE sense), numerical approximation is also very costly in terms of computing time, particularly in the case of II, which can take 22 times longer to compute than when analytical moments are used.[5]

- When comparing EMM and II, the latter tends to dominate EMM in the RMSE for the case of MA(1) models. There is, however, an important caveat. As Tauchen (1996) noted, EMM provides a numerically stable environment for optimizing the GMM objective function. This is not the case for II, particularly when we are dealing with MA processes close to the unit circle. In fact, for these specifications, convergence had to be "forced" in about 52% of the cases, where the objective function turned out to be unstable even when analytical moments were supplied. This is not the case for EMM, where convergence was never "forced."

- As documented in Chumacero 1997 and Michaelides and Ng 1997, in this setup, EMM tends to over-reject the null when using the standard over-identifying restrictions test. Furthermore, the less parsimonious the auxiliary model, the greater the size problem. Thus, AIC tends to over-reject more than HQ, which does the same compared to BIC. The magnitude of over-rejections is rather important; thus a simple correction is suggested. Instead of using the actual sample size, $T - q$ should be used for scaling (6). As Table 3

[4]All previous Monte Carlo experiments conducted in this setup use the approximation to the moment conditions instead of the exact moments.
[5]These gains are even more impressive when we are dealing with more complex ARMA models.

shows, this simple modification allows for an almost complete correction of the size distortion around standard levels of significance. At any rate, size distortions are still important for levels of more than 15%. Due in part to the unstable nature of the II objective function, its size distortions are more significant and cannot be corrected with the simple procedure described above.

- With respect to Wald-type tests, when we are using EMM and when the parameter is relatively distant from the unit circle, they behave well (Table 4). Nonetheless, they tend to become unstable near the unit root, thus making the inversion of chi-square tests preferable. Note, however, that in all these cases, a correction such as the one described above is also needed. For II, results were not even reported, because even when it was distant from the unit circle, the objective function was so unstable that singularities were common.

To conclude, EMM tends to perform well when estimating MA(1) models, particularly for cases approaching a unit root where the bias usually associated with conditional ML is less. At any rate, ML is still more efficient in terms of RMSE. There is, however, one important advantage to using EMM, and that involves the time needed to estimate these models. In fact, EMM is at least twice as fast as ML. When the overidentifying-restrictions test is suitably transformed, AIC should become the selection criterion of choice, because it usually provides more efficient estimates. AIC should not be chosen for inference, however, if the objective function is not transformed. In the case of II, even when the reported RMSEs are generally smaller than with EMM, the objective function is particularly unstable, and in most cases convergence had to be forced. Given the unstable nature of this objective function, overidentifying-restrictions tests and Wald tests present significant size distortions that cannot be corrected.

4.3 Results for the ARMA(1,1) model

As in the previous exercise, several specifications for the ARMA(1,1) process were estimated. In all cases, the parameter associated with the autoregressive coefficient was set equal to −0.8 and the coefficient associated with the MA component was allowed to vary. Once again 1,000 samples each of sizes 100 and 200 were artificially generated and estimated by ML and three EMM and II estimators. In the latter cases, we used the exact moment conditions, choosing the auxiliary models using the three selection criteria discussed above.

The results of these experiments are reported in Tables 5 and 6 and can be summarized as follows:

- The selection criterion that renders the smallest RMSE for the autoregressive coefficient is BIC, followed by HQ, and finally AIC. This order is reversed when we compare the RMSE for the MA coefficient. The inference is clear: particularly where the process is invertible, the auxiliary model that best captures the dynamics of the MA coefficient requires several lags. To better capture the characteristics of the AR component, parsimony is preferred (BIC).

- All models do equally well in terms of bias. As discussed above, particularly where the MA coefficient is close to the unit circle, EMM reduces ML-associated bias.

- The gains in computing time rise substantially when EMM is used instead of ML. In particular, the average estimation for a sample size of 200 is 20 times faster with EMM than with ML. These gains are increased to a factor of 35 if the exact likelihood is used when estimating using ML.

- There is another compelling reason, besides computational efficiency, to use EMM. The EMM objective function is numerically stable, thus one could use EMM (with analytical moments) to compute starting values and then proceed with ML.

- Compared to EMM, II still shows significant numerical stability problems. This time, however, EMM tended to outperform II in the RMSE for several specifications. Again, II required “forced” convergence in more than 60% of the cases.

Table 5
Properties of the estimators: ARMA(1,1) model ($\delta = -0.8$, $T = 100$)

	EMM				II				
Estimator	δ		α		δ		α		
$\alpha = -0.7$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.793	0.064	−0.704	0.085	−0.793	0.064	−0.704	0.085	
AIC	−0.771	0.125	−0.742	0.170	−0.743	0.336	−0.832	0.216	5.40
BIC	−0.774	0.096	−0.755	0.185	−0.813	0.169	−0.796	0.225	3.48
HQ	−0.775	0.102	−0.754	0.175	−0.774	0.228	−0.828	0.225	4.06
$\alpha = -0.8$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.795	0.063	−0.797	0.075	−0.795	0.063	−0.797	0.075	
AIC	−0.775	0.127	−0.829	0.155	−0.812	0.323	−0.922	0.173	6.42
BIC	−0.779	0.097	−0.843	0.167	−0.863	0.198	−0.891	0.189	3.88
HQ	−0.777	0.105	−0.842	0.155	−0.838	0.248	−0.919	0.180	4.83
$\alpha = -0.9$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.797	0.061	−0.880	0.067	−0.797	0.061	−0.880	0.067	
AIC	−0.777	0.135	−0.893	0.148	−0.961	0.390	−0.967	0.106	7.89
BIC	−0.783	0.097	−0.897	0.139	−0.941	0.267	−0.939	0.136	4.28
HQ	−0.782	0.110	−0.901	0.127	−0.960	0.332	−0.962	0.111	5.67
$\alpha = -0.95$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.801	0.061	−0.909	0.072	−0.801	0.061	−0.909	0.072	
AIC	−0.779	0.139	−0.912	0.133	−0.997	0.479	−0.976	0.091	8.72
BIC	−0.785	0.100	−0.909	0.140	−0.971	0.301	−0.950	0.117	4.46
HQ	−0.783	0.114	−0.914	0.165	−0.997	0.388	−0.973	0.131	5.99

Note: See Table 2 for definitions.

Table 6
Properties of the estimators: ARMA(1,1) model ($\delta = -0.8$, $T = 200$)

	EMM				II				
Estimator	δ		α		δ		α		
$\alpha = -0.7$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.797	0.044	−0.702	0.057	−0.797	0.044	−0.702	0.057	
AIC	−0.784	0.064	−0.724	0.097	−0.689	0.273	−0.797	0.156	6.13
BIC	−0.786	0.058	−0.746	0.132	−0.745	0.145	−0.805	0.178	3.94
HQ	−0.785	0.060	−0.737	0.113	−0.714	0.189	−0.809	0.167	4.72
$\alpha = -0.8$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.798	0.043	−0.799	0.049	−0.798	0.043	−0.799	0.049	
AIC	−0.785	0.068	−0.823	0.093	−0.711	0.254	−0.899	0.139	7.93
BIC	−0.788	0.057	−0.853	0.129	−0.785	0.144	−0.918	0.161	4.69
HQ	−0.787	0.060	−0.840	0.113	−0.752	0.184	−0.912	0.151	5.84
$\alpha = -0.9$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.799	0.043	−0.889	0.042	−0.799	0.043	−0.889	0.042	
AIC	−0.788	0.070	−0.911	0.076	−0.844	0.255	−0.970	0.090	10.42
BIC	−0.791	0.064	−0.924	0.100	−0.884	0.189	−0.969	0.099	5.59
HQ	−0.790	0.062	−0.921	0.086	−0.877	0.217	−0.974	0.094	7.45
$\alpha = -0.95$	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	j
ML	−0.801	0.042	−0.926	0.046	−0.801	0.042	−0.926	0.046	
AIC	−0.790	0.073	−0.942	0.065	−0.967	0.337	−0.982	0.077	11.94
BIC	−0.793	0.063	−0.941	0.089	−0.936	0.235	−0.976	0.067	5.97
HQ	−0.793	0.062	−0.944	0.073	−0.960	0.283	−0.985	0.090	8.19

Note: See Table 2 for definitions.

5 Concluding Remarks

This article develops a methodology for estimating stationary Gaussian ARMA models (both invertible and noninvertible) using EMM and II. In contrast to the prevailing practice, simulation is not required to compute the moment conditions used by these methods. The gains in terms of efficiency and computing time are substantial.

Where the ARMA process is close to the unit circle, EMM may be preferred because it reduces bias. In any case, ML still yields lower RMSE than EMM and II.

This article also addresses how to choose the auxiliary model, examining three automatic selection criteria. This examination revealed that AIC tends to perform better than BIC and HQ (in terms of RMSE) when one is estimating pure MA models.

Experiments performed also indicated a simple way of correcting the over-rejection problem that typically occurs when using Hansen's (1982) test. If this correction is not performed, large-scale auxiliary models will present significant size distortions. This correction does not apply when using II, because its objective function is numerically unstable near the unit root.

When dealing with ARMA models that are relatively close to the unit circle, Wald-type tests do not perform adequately when compared to their asymptotic properties. Thus, inverting chi-square tests are preferable once a $(T - q)$ correction is performed.

References

Chumacero, R. (1997). "Finite sample properties of the efficient method of moments." *Studies in Nonlinear Dynamics and Econometrics*, 2: 35–51.

Galbraith, J. W., and V. Zinde-Walsh (1994). "A simple non-iterative estimator for moving-average models." *Biometrika*, 81: 143–155.

Galbraith, J., and V. Zinde-Walsh (1997). "On simple auto-regressive-based estimation and identification techniques for ARMA models." *Biometrika*, 84: 685–696.

Gallant, R., and G. Tauchen (1996). "Which moments to match?" *Econometric Theory*, 12: 657–681.

Gallant, R., and G. Tauchen (1999). "The relative efficiency of method of moments estimators." *Journal of Econometrics*, 92: 149–172.

Ghysels, E., L. Khalaf, and C. Vodounou (1994). "Simulation based inference in moving average models." Unpublished manuscript. Montreal, Canada: Centre Interuniversitaire de Recherche en Analyse des Organisations.

Gouriéroux, C., A. Monfort, and E. Renault (1993). "Indirect inference." *Journal of Applied Econometrics*, 8: S85–S118.

Hamilton, J. (1994). *Time Series Analysis*. Princeton, N.J.: Princeton University Press.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators." *Econometrica*, 50: 1029–1053.

Michaelides, A., and S. Ng (1997). "Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators." Unpublished manuscript. Boston: Boston College.

Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: John Wiley & Sons.

Tauchen, G. (1996). "New minimum chi-square methods in empirical finance." Unpublished manuscript. Raleigh, N.C.: Duke University.

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