ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 69 – 2003

Simulation Based Inference in Moving Average Models

Éric GHYSELS, Lynda KHALAF, Cosmé VODOUNOU *

ABSTRACT. – We examine several autoregressive-based estimators for the parameters of a moving average process, including the estimators initially proposed by GALBRAITH and ZINDE-WALSH [1994] and GOURIÉROUX, MONFORT and RENAULT [1993]. We also propose over-identified asymptotic-least-squares based variants of the former, and extensions of the latter based on GALLANT and TAUCHEN'S [1996] simulated method of moments. The relative performance of these estimators is assessed, with emphasis on the near-uninvertibility region. We find that, although no formal local-to-one arguments are taken into consideration, the *Wald*-type indirect inference method performs best at the boundary, with practically just one calibration.

Méthodes d'inférence basées sur des simulations dans les modèles de moyenne mobile

RÉSUMÉ. – Nous examinons plusieurs estimateurs basés sur des autorégressions pour des modèles de moyenne mobile et principalement ceux qui généralisent les procédures récentes avancées par GALBRAITH et ZINDE-WALSH [1994] et GOURIÉROUX, MONFORT et RENAULT [1993]. Nous proposons des variantes sur-identifiées de l'estimateur de *Galbraith* et *Zinde-Walsh*, obtenues par la méthode des Moindres Carrés Asymptotiques, et des extensions de l'estimateur de *Gouriéroux*, *Monfort* et *Renault*, fondées sur la méthode des moments simulés (GALLANT et TAUCHEN [1996]). Nous évaluons respectivement leur performance en étudiant spécifiquement leur comportement dans la région des valeurs proches de la non-inversibilité. Bien que nous ne considérions pas formellement d'arguments asymptotiques locaux au cercle unitaire, nos résultats suggèrent que la méthode d'inférence indirecte du type *Wald* a la meilleure performance à la limite de l'espace des paramètres, même avec une seule calibration.

* É. GHYSELS: Department of Economics, University of North Carolina; L. KHALAF: Département d'économique Université Laval, Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN), and Centre interuniversitaire de recherche en économie quantitative (CIREQ); C. VODOUNOU: Institut National de la Statistique et de l'Analyse Économique, Bénin.

1 Introduction

Simulation-based estimation procedures have become popular in recent years. For instance, in the context of dynamic models, LAROQUE and SALANIÉ [1989], SMITH [1993], INGRAM and LEE [1991], DUFFIE and SINGLETON [1993], GOURIÉROUX, MONFORT and RENAULT [1993], GALLANT and TAUCHEN [1996], GOURIÉROUX and MONFORT [1996], HARRIS [1999], DRIDI [2000] and DRIDI and RENAULT [2000], GALBRAITH and ZINDE-WALSH [2001] provide extensive coverage of simulation-based methods. As the computational burden is becoming less of an issue one observes a change in the practice of econometrics. There are several reasons for the increased popularity of simulation-based methods in applied work. For instance, simulations are often the key to feasible estimation in various non-linear contexts. Moreover, these procedures are shown to circumvent finite sample problems such as estimation bias and tests size control.

The purpose of this paper is to assess the usefulness of simulation and instrumental based inference methods, such as those proposed by GOURIÉROUX, MONFORT and RENAULT [1993] and GALLANT and TAUCHEN [1996], in boundary conditions. TAUCHEN [1998] studies the role of stability constraints in estimation of dynamic models and analyses the behavior of the objective function on either side of the boundary of the stability region of the parameter space. His paper is an excellent example of the remarkable performance of simulation-based methods when parameters of interest are near boundaries. We examine similar issues though focus more specifically on the estimation of MA(1) models for which we study new and existing autoregressions (AR) based instrumental estimators. There are several reasons why we focus on the MA(1) model. First, the difficulties associated with maximum-likelihood (ML) estimation – despite the model's very simple structure – are well documented.[1] Second, it is well known that the performance of standard ML and NLS estimators and test statistics may be very poor in near-boundary conditions, which is unfortunate as close-to-boundary MA(1) models are of considerable interest in econometric practice; DAVIS and DUNSMUIR [1996] discuss several examples. Local-to-boundary modeling has recently been proposed to deal with such problems.[2] However, the complexities of the associated asymptotic distributions are a barrier to practical applications.

There are several existing instrumental AR-based procedures for the estimation of MA models, and here we provide several extensions which are built on these estimators. DURBIN [1959] provided a first example of such a procedure. This relatively straightforward approach has attracted considerable attention in the literature; see, for example, MCCLAVE [1973] or MENTZ [1977]. HANNAN and RISSANEN [1982] and KOREISHA and PUKKILA [1990] suggested an alterna-

1. For likelihood-based or nonlinear least squares (NLS) estimators, see, for instance, BOX and JENKINS [1976], FULLER [1976], GODOLPHIN [1977], OSBORN [1977], ANSLEY [1979] and ANSLEY and NEWBOLD [1980].

2. References relevant to the model considered in this paper may be found in HARRIS [1999], SHEPPARD [1993], DAVIS, CHEN and DUNSMUIR [1994], DAVIS and DUNSMUIR [1996] and MCCABE and LEYBOURNE [1998].

tive method which relies on a sequence of long autoregressions. GALBRAITH and ZINDE-WALSH [1994] (G-ZW) proposed a related estimator which involves minimum *Hilbert Norm* distances.[3] GOURIÉROUX, MONFORT and RENAULT [1993] (GMR) considered a simulation-AR based indirect estimator. Although existing results on all these estimators are promising in MA(1) contexts, little is known about their relative merits, particularly for the parameter close to the boundary. It also appears that a method-of-moments (MM) technique which exploits the overidentification implicit in the well-known *Yule-Walker* equations has not been formally studied in this simple model. Here, we propose several instrumental AR-based estimators applying: (i) the asymptotic least squares (ALS) method (GOURIÉROUX, MONFORT and TROGNON [1985]) and, (ii) the simulated methods-of-moments (EMM) method (GALLANT and TAUCHEN [1996]). We study these estimators along with the various other instrumental AR-based estimators introduced in the literature.

Our Monte Carlo results show the following: (i) the simple G-ZW estimator seems preferable – in general – to its over-identified ALS counterpart, (ii) the EMM estimator is dominated by the GMR estimator throughout the parameter space, (iii) the GMR estimator outperforms all estimators studied near unity, (iv) whereas most estimators are more erratic at the boundary, the GMR estimator seems not to suffer from this problem. The GMR emerges as the best choice within the class of estimators studied. Note that we do not assess the impact of estimating the variance of the model's innovations.

The paper is organized as follows. We set notation in section 2. In section 3, we introduce the different estimators. Section 4 reviews briefly the estimators' asymptotic properties. In section 5, we report simulation results comparing the finite sample performance of the different estimators. Section 6 concludes.

2 Framework

We consider the process:

$$X_t = u_t + \theta\, u_{t-1},\ t = 1,\ldots,T, \tag{1}$$

where $|\theta| < 1$ and u_t, $t = 1,\ldots,T$, are independent, identically distributed, normal random variables with $Eu_t = 0$, $Eu_t^2 = 1$. In this framework, we are naturally led to consider the approximating AR(p) model:

$$X_t = \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \varepsilon_t,\ t = 1,\ldots,T, \tag{2}$$

where ε_t is a white noise. As is well-known, if $p \to \infty$, then the AR coefficients are linked to θ through the following recursion:

$$\beta_1 = \theta, \tag{3}$$

$$\beta_j = -\theta\beta_{j-1},\ \ j > 1. \tag{4}$$

3. See also GALBRAITH and ZINDE-WALSH [1997].

Let,

$$(5) \qquad \hat{\beta}^{p}_{(T)} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)' = \underset{\beta}{\text{Argmin}}\, Q^{p}_{(T)}(\beta),$$

$$(6) \qquad Q^{p}_{(T)}(\beta) = \Sigma^{T}_{t=p+1}\left[X_t - \beta_1 X_{t-1} - \ldots - \beta_p X_{t-p}\right]^2,$$

refer to the OLS estimates of the AR parameter $\beta = \left(\beta_1, \ldots, \beta_p\right)'$, in the regression (2). The usual central limit theorem arguments yield that:

$$(7) \quad \text{plim}\, \hat{\beta}_i = (-1)^{i-1}\,\theta^i \left[\frac{1-\theta^{2(p-i+1)}}{1-\theta^{2(p+1)}}\right] \equiv \beta^{p}_{i}(\theta),\ i = 1,\ldots,p.$$

For further reference, let:

$$(8) \qquad \beta^{p}(\theta) = \left(\beta^{p}_{1}(\theta), \ldots, \beta^{p}_{p}(\theta)\right)'.$$

Henceforth, we will also use the weighting outer-product matrix:

$$(9) \qquad \hat{\Omega}_{(T)} = \left[\left(\frac{\partial Q^{p}_{(T)}}{\partial \beta}(\beta)\right)\left(\frac{\partial Q^{p}_{(T)}}{\partial \beta}(\beta)\right)'\right]^{-1}$$

evaluated at $\hat{\beta}^{p}_{(T)}$.

The inference procedures we study here may be classified into two categories, based on whether (3)-(4) and (7) are implicitly or explicitly used to derive the estimators. For instance, G-ZW obtain their approximate estimator applying (3)-(4) and compute its asymptotic bias using (7). In contrast, the GMR estimator does not require such explicit relations between the MA and AR parameters.[4] The lag truncation bias is rather handled through simulations, as will become clear from our presentation below.

Our framework sets the variance parameter to one, and the underlying distribution to the Gaussian case. It is easy to see that if we assume that the process u_t is $i.i.d.\ N(0,\sigma^2)$, then the OLS AR-based variance estimator $\widehat{\sigma}^2_{p(T)}$ satisfies,

$$(10) \qquad \text{plim}\, \widehat{\sigma}^2_{p(T)} = \frac{\sigma^2\left(1-\theta^{2(p+2)}\right)}{1-\theta^{2(p+1)}} \equiv \sigma^2_p\left(\theta,\sigma^2\right),$$

given the same regularity conditions underlying (7). This caracterizes a *binding function* which may serve to identify σ^2. Note, however, that whether $\sigma^2 \neq 1$ or $\sigma^2 = 1$, $\text{plim}\, \hat{\beta}_i = \beta^{p}_{i}(\theta)$, $i = 1,\ldots,p$, where $\beta^{p}_{i}(\theta)$ is as defined above, *ie,* is free of σ. In other words, the effective *binding function* is

4. Of course, (7) provides an implicit *binding function* which justifies the use of the AR process as an instrumental model. The concept of a *binding function* is formerly defined in Gouriéroux, Monfort and Renault [1993].

$\beta^p(\theta)$, which justifies the simplifying assumption $\sigma^2 = 1$. Our point here is that indirect inference theory, as originally proposed (GMR, GALLANT and TAUCHEN [1996]), does not take nuisance parameters formally in consideration. The same holds true for the G-ZW approach. In view of this, if θ is viewed as the parameter of interest, then our framework is not unduly restrictive. Of course, (10) implies that the auxiliary function σ_p^2 may provide some information about θ; our Monte Carlo design does not allow to assess this effect. It is important to note, however, that the indirect inference theory has evolved towards a semi-parametric setting (see, for example, DRIDI and RENAULT [2000]) which allows to treat the scale factor and possibly the full error distribution as a nuisance parameter. Semi-parametric indirect inference (denoted $SPII$) is beyond the scope of the present paper, yet we should note that if nuisance parameters (*eg,* moments of the error distribution, or the error distribution *per se*) are of concern, $SPII$ seems a more appropriate estimation method.

In the following, we shall first consider the indirect inference based estimators. To introduce the GMR procedure, we shall define simulated paths and calibrated simulated criteria, which will also serve to motivate the EMM estimators we propose. We shall next re-examine the G-ZW estimator and provide an extension which exploits the over-identified relations (7).

3 Autoregressive Based Estimators

3.1 Wald-type Indirect Inference Estimators

For a given value of the MA parameter and using the structural MA(1) model (equation 1), we can obtain H simulated paths of size T

(11) $$\left[\tilde{X}_1^h(\theta), \ldots, \tilde{X}_T^h(\theta)\right], h = 1, \ldots, H,$$

based on independent drawings from *iid* standard normal variates $\left[\tilde{u}_1^h, \ldots, \tilde{u}_T^h\right]$, $h = 1, \ldots, H$. For each of these paths, and conformably with (5)-(6), define:

(12) $$\tilde{\beta}_{(T)}^{hp}(\theta) = \left(\tilde{\beta}_1^h(\theta), \ldots, \tilde{\beta}_p^h(\theta)\right)' = \underset{\beta}{\text{Argmin}}\, \tilde{Q}_{(T)}^{hp}(\beta, \theta),$$

(13) $$\tilde{Q}_{(T)}^{hp}(\beta, \theta) = \Sigma_{t=p+1}^T \left[\tilde{X}_t^h(\theta) - \beta_1 \tilde{X}_{t-1}^h(\theta) - \ldots - \beta_p \tilde{X}_{t-p}^h(\theta)\right]^2$$

$$h = 1, \ldots, H.$$

In other words, $\tilde{\beta}^{hp}_{(T)}(\theta)$ minimizes the least squares criterion where the observed sample is replaced by a simulated sample conditional upon θ. To "calibrate" the procedure overall simulated path, let

$$(14) \qquad \tilde{\beta}^{p}_{(T,H)}(\theta) = \frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}^{hp}_{(T)}(\theta).$$

The GMR estimator solves for the value of θ which minimizes a quadratic form based on $\left(\hat{\beta}^{p}_{(T)} - \tilde{\beta}^{p}_{(T,H)}(\theta)\right)$. Formally, the estimator is:

$$(15) \quad \hat{\theta}^{SW}_{(T,H,p)} = \underset{\theta}{\text{Argmin}}\left(\hat{\beta}^{p}_{(T)} - \widetilde{\beta}^{p}_{(T,H)}(\theta)\right)'\hat{\Omega}_{(T)}\left(\hat{\beta}^{p}_{(T)} - \widetilde{\beta}^{p}_{(T,H)}(\theta)\right)$$

where the weighting matrix $\hat{\Omega}_{(T)}$ is obtained using the observed sample as in (9).

Alternatively, we can obtain a sequence $\tilde{X}_1(\theta),\ldots,\tilde{X}_{TH}(\theta)$, of TH simulated values based on TH drawings from the white noise process u_t, to which one would apply the following estimation criterion:

$$(16) \qquad \tilde{\beta}^{p}_{(TH)}(\theta) = \left(\tilde{\beta}^{(TH)}_{1}(\theta),\ldots,\tilde{\beta}^{(TH)}_{p}(\theta)\right)'$$

$$= \text{Argmin}\, \tilde{Q}^{p}_{(TH)}(\beta,\theta),$$

$$(17) \quad \tilde{Q}^{p}_{(TH)}(\beta,\theta) = \Sigma^{TH}_{t=p+1}\left[\tilde{X}_t(\theta) - \beta_1\tilde{X}_{t-1}(\theta),\ldots,\beta_p\tilde{X}_{t-p}(\theta)\right]^2.$$

This leads to another quadratic form based on the difference $\left(\hat{\beta}^{p}_{(T)} - \tilde{\beta}^{p}_{(TH)}(\theta)\right)$ and an alternative estimator of the GMR type as follows

$$(18) \quad \hat{\theta}^{SW}_{(TH,p)} = \underset{\theta}{\text{Argmin}}\left(\hat{\beta}^{p}_{(T)} - \widetilde{\beta}^{p}_{(TH)}(\theta)\right)'\hat{\Omega}_{(T)}\left(\hat{\beta}^{p}_{(T)} - \widetilde{\beta}^{p}_{(TH)}(\theta)\right).$$

A notational point: the estimators in (15)-(18) bear the superscript SW indicating that they are simulation-based exploiting a *Wald* principle.

GOURIÉROUX *et al.* [1993] show that both estimators have the same asymptotic distributions given appropriate regularity conditions. They also argue that, in this particular application, the efficiency loss resulting from replacing $\hat{\Omega}_T$ by an identity matrix is marginal and, therefore, suggest to use the latter in practical applications. We proceed next to extend the indirect inference methodology in the EMM direction.

3.2 Efficient Method of Moments Estimators

Following the procedure set forth in GALLANT and TAUCHEN [1996], we introduce a third version of the indirect estimator, as follows. Consider the score vector $\partial\tilde{Q}^{p}_{(TH)}/\partial\beta(\beta,\theta)$ where $\tilde{Q}^{p}_{(TH)}(\beta,\theta)$ is defined in (17). To implement the GALLANT and TAUCHEN [1996] procedure, we propose to construct a quadratic form based on the difference between the latter simulated score evaluated at the observed $\hat{\beta}^p_{(T)}$, which we denote $\partial \tilde{Q}^p_{(TH)}/\partial\beta\left(\hat{\beta}^p_{(T)},\theta\right)$, and its data-based counterpart which is zero by construction.

$$(19)\quad \hat{\theta}^{SS}_{(TH,p)} = \underset{\theta}{\text{Argmin}}\left[\frac{\partial \tilde{Q}^{(p)}_{TH}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right)\right]' \hat{\Omega}_{(T)}\left[\frac{\partial \tilde{Q}^{(p)}_{TH}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right)\right].$$

The advantage of this estimator is that it does not involve estimating β from the simulated data as it relies on the score function, hence the superscript SS. We also study a calibrated score version of the estimator as in (15). Let:

$$(20)\quad \frac{\partial \tilde{Q}^{(p)}_{T,H}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right) = \frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{Q}^{hp}_{(T)}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right),$$

where $\partial \tilde{Q}^{hp}_{(T)}/\partial\beta\left(\hat{\beta}^p_{(T)},\theta\right)$ corresponds to the score associated with a single simulated path, evaluated at $\hat{\beta}^p_{(T)}$. Constructing the corresponding EMM quadratic form leads to the estimator:

$$(21)\quad \hat{\theta}^{SS}_{(T,H,p)} = \underset{\theta}{\text{Argmin}}\left[\frac{\partial \tilde{Q}^{(p)}_{T,H}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right)\right]' \hat{\Omega}_{(T)}\left[\frac{\partial \tilde{Q}^{(p)}_{T,H}}{\partial\beta}\left(\hat{\beta}^p_{(T)},\theta\right)\right].$$

From the results in GMR and GALLANT and TAUCHEN [1996], these estimators are asymptotically equivalent to the GMR estimators (see section 4 for a brief discussion of the asymptotic distributions involved). The EMM estimators (19)-(21) also leads to a class of estimators not involving simulation yet are based on methods of moments principles. We consider these methods in the next section.

3.3 Asymptotic Least Squares Estimators

The estimators considered in this class use (3)-(4) and/or (7) explicitly. The G-ZW estimators defined as:

$$(22)\quad \hat{\theta}^{G-ZW}_{(T,p)} = \hat{\beta}_1$$

is based on (3). The latter is clearly straightforward to apply and does not require the assumption of Gaussian errors. Here, we consider an over-identified version, of this estimator which follows the ALS principle (GOURIÉROUX, MONFORT and TROGNON [1985]):

$$(23)\quad \hat{\theta}^{ALS}_{(T,p)} = \underset{\theta}{\text{Argmin}}\left(\hat{\beta}^p_{(T)} - \beta^p(\theta)\right)' \hat{\Omega}_{(T)}\left(\hat{\beta}^p_{(T)} - \beta^p(\theta)\right)$$

where $\beta^p(\theta)$ is defined in (7)-(8). Again, we also consider replacing the (asymptotically) optimal weighting matrix $\hat{\Omega}_T$ by an identity matrix as suggested by GMR.

Since the *binding function* (7) is fully exploited in the formulae for $\hat{\theta}^{ALS}_{(T,p)}$, efficiency gains with respect to the G-ZW estimator are possible, at least from an asymptotic perspective. Whether this proves true in finite samples is an open question. It is also interesting to compare the finite sample performance of this estimator to the simulation-based ones. The principle difference between $\hat{\theta}^{ALS}_{(T,p)}$ and the simulation-based estimators, say $\hat{\theta}^{SW}_{(T,H,p)}$ for instance, is readily seen on comparing (23) and (15). In a way, both estimators rely on the distance between the data based AR estimates and its hypothesized value. The latter is obtained using standard asymptotics in the case of $\hat{\theta}^{ALS}_{(T,p)}$ and through averages of simulated estimators in the case of $\hat{\theta}^{SW}_{(T,H,p)}$. The simulation experiment we conduct in the sequel is designed to assess these questions.

4 Asymptotic Distributions of Suggested Estimators: a Unified Framework

In this section, we briefly summarize the asymptotic distributions of the estimators we consider.[5] We will present results relating to any weighting matrix S_T satisfying

$$S_0 = \lim_{T \to \infty} S_T \tag{24}$$

so that the optimal weight will be clearly suggested by the formulae for the asymptotic variances provided. For clarity, we append the symbol (S_T) to each estimator's label when required. The properties of the proposed estimators are of course motivated by the following result dealing with the data-based OLS AR estimator $\hat{\beta}^p_{(T)}$.

PROPOSITION 1. *$T^{\frac{1}{2}}\left(\hat{\beta}^p_{(T)} - \beta^p(\theta_0)\right)$ has Ω_0 as an asymptotic covariance matrix, where $\beta^p(\theta)$ is given by (7)-(8), $\hat{\beta}^p_{(T)}$ is the data-based AR OLS estimator (5), θ_0 is the true value of the MA parameter, and Ω_0 is the information matrix based on the score function which underlies $\hat{\beta}^p_{(T)}$.*

5. For formal proofs, relevant regularity conditions and further discussion, the reader may refer to GOURIÉROUX, MONFORT and RENAULT [1993], GOURIÉROUX, MONFORT and TROGNON [1985], GALLANT and TAUCHEN [1996] and DRIDI [2000].

Furthermore, the simulation-based estimators rely on a limiting *binding function*, which following GMR, can in our case be defined as follows:

$$b^p(\theta) = \lim_{T\to\infty} \tilde{\beta}^{hp}_{(T)}(\theta), \tag{25}$$

where $\tilde{\beta}^{hp}_{(T)}(\theta)$ is the OLS AR based estimator associated with a simulated path, see (12). For ease of presentation, we have dropped the h index from the LHS of the latter equation, since the limit in question is for any arbitrary simulated path. From (7)-(8), we have $b^p(\theta) = \beta^p(\theta)$.[6] Conformably, define:

$$q^p(\beta,\theta) = \lim_{T\to\infty} \widetilde{Q}^{hp}_{(T)}(\beta,\theta).$$

The asymptotic properties of the above estimators are best cast in terms of the unified SALS framework set forth in DRIDI [2000]. For the problem at hand, a SALS estimator – with weighing matrix $\hat{\Omega}_{(T)}$ – may be obtained as:

$$\hat{\theta}^{SALS}_{(T,H,p)} = \underset{\theta}{\text{Argmin}}\ \widetilde{\mathcal{H}}_{TH}\left(\hat{\beta}^p_{(T)},\theta\right)' \hat{\Omega}_{(T)} \widetilde{\mathcal{H}}_{TH}\left(\hat{\beta}^p_{(T)},\theta\right)$$

where $\widetilde{\mathcal{H}}_{TH}(\beta,\theta)$ is a simulation-based estimate (with H replications) of the *estimating equations* which link the auxiliary parameters to those of interest. In other words, $\widetilde{\mathcal{H}}_{TH}(\beta,\theta)$ approximates a function $\mathcal{H}(\beta,\theta)$ defined by $\mathcal{H}(\beta_0,\theta_0) = 0$ where β_0 and θ_0 refer to the true values of β and θ. For the problem at hand, two estimating equations are relevant:

$$\begin{aligned} \mathcal{H}_{TH}(\beta,\theta) &= \beta - \beta^p(\theta), && \text{in the case of } \textit{Wald}\text{-type criteria,} \\ &= E\left(\frac{\partial q^p}{\partial \beta}(\beta,\theta)\right), && \text{in the case of score-type criteria.} \end{aligned}$$

The ALS estimator makes use of a close-form expression for $\mathcal{H}(\beta,\theta)$:

$$\widetilde{\mathcal{H}}_{TH}(\beta,\theta) = \mathcal{H}(\beta,\theta)$$

whereas the simulation based estimators use:

$$\begin{aligned} \widetilde{\mathcal{H}}_{TH}(\beta,\theta) &= \beta - \widetilde{\beta}^p_{(T,H)}(\theta), && \text{in the case of } \hat{\theta}^{SW}_{(T,H,p)}, \\ &= \beta - \widetilde{\beta}^p_{(TH)}(\theta), && \text{in the case of } \hat{\theta}^{SW}_{(TH,p)}, \\ &= \frac{\partial \tilde{Q}^{(p)}_{T,H}}{\partial \beta}(\beta,\theta), && \text{in the case of } \hat{\theta}^{SS}_{(T,H,p)}, \\ &= \frac{\partial \tilde{Q}^{(p)}_{TH}}{\partial \beta}(\beta,\theta), && \text{in the case of } \hat{\theta}^{SS}_{(TH,p)}, \end{aligned}$$

6. One may also argue that (3)-(4) – *ie*, an AR(∞) – provides an implicit *binding function*. Indeed, indirect inference methods typically rely on a numerical mapping between the β vector and θ. In this case, $b^p(\theta)$ may not *explicitly* coincide with $\beta^p(\theta)$.

where, $\widetilde{\beta}^{p}_{(T,H)}(\theta)$, $\widetilde{\beta}^{p}_{(TH)}(\theta)$, $\tilde{Q}^{(p)}_{T,H}(\beta,\theta)$, and, $\dfrac{\partial\tilde{Q}^{(p)}_{TH}}{\partial\beta}(\beta,\theta)$ are defined in (14), (16), (17) and (20).

DRIDI [2000] derives the limiting distributions of the SALS under fairly general regularity conditions which contain the framework of GOURIÉROUX, MONFORT and TROGNON [1985] and GMR. In terms of the ALS estimator, these results imply the following.

PROPOSITION 2. *The ALS estimator $\hat{\theta}^{ALS}_{(T,p)}$ defined in (23) has, under suitable regularity conditions, the following asymptotic distribution:*

$$T^{\frac{1}{2}}\left(\hat{\theta}^{ALS}_{(T,p)}(S_T)-\theta_0\right)\underset{T\to\infty}{\overset{d}{\to}}N\left(0,\Sigma^{ALS}_{p}(S_0)\right),$$

where,

$$\Sigma^{ALS}_{p}(S_0)=W_\beta(S_0)\frac{\partial\beta^p(\theta)'}{\partial\theta}S_0\Omega_0S_0\frac{\partial\beta^p(\theta)}{\partial\theta}W_\beta(S_0),\tag{26}$$

$$W_\beta(S_0)=\left[\frac{\partial\beta^p(\theta)'}{\partial\theta}S_0^{-1}\frac{\partial\beta^p(\theta)}{\partial\theta}\right]^{-1}.$$

It is evident from the latter proposition that the optimal choice for S_0 is Ω_0^{-1}. With respect to the simulation-based estimators, the following limiting distributions hold.

PROPOSITION 3. *Under suitable regularity conditions, the Wald-type simulation based estimators $\hat{\theta}^{SW}_{(T,H,P)}$ and $\hat{\theta}^{SW}_{(TH,p)}$ defined by (15) and (18) have the following asymptotic distribution:*

$$T^{\frac{1}{2}}\left(\hat{\theta}^{SW}_{i}(S_T)-\theta_0\right)\underset{T\to\infty}{\overset{d}{\to}}N\left(0,\Sigma^{SW}_{p}(S_0)\right)$$

where $i=(TH,_P),(T,H_{,}P)$ and,

$$\Sigma^{SW}_{p}(S_0)=\left(1+\frac{1}{H}\right)W_b(S_0)\frac{\partial b^p(\theta)'}{\partial\theta}S_0\Omega_0S_0\frac{\partial b^p(\theta)}{\partial\theta}W_b(S_0),\tag{27}$$

$$W_b(S_0)=\left[\frac{\partial b^p(\theta)'}{\partial\theta}S_0^{-1}\frac{\partial b^p(\theta)}{\partial\theta}\right]^{-1}.$$

Again, we see that the optimal choice for S_0 is Ω_0^{-1}. Finally, from the results in GMR, we can establish the asymptotic equivalence of the EMM estimators we propose and $\hat{\theta}^{SW}_{(T,H,P)}$ and $\hat{\theta}^{SW}_{(TH,p)}$. For completion, this is formally stated in the following proposition.

PROPOSITION 4. *Under suitable regularity conditions and using usual central limit theorem arguments, $\hat{\theta}^{SS}_{(TH,p)}$ – defined by (19) – is asymptotically equivalent to $\hat{\theta}^{SW}_{(TH,p)}$ (18). Further, if H is large, the estimator $\hat{\theta}^{SS}_{(T,H,p)}(S_T)$ – defined by (21) – has the same limiting distribution as $\hat{\theta}^{SS}_{(TH,p)}(S_T)$.*

If we compare the asymptotic variances which appear in (26) and (27), two points are worth noting. First, consider the scaling factor $\left(1+H^{-1}\right)$ which intervenes in the case of the simulation-based estimators. Such a factor is expected, as it accounts for the additional noise or randomness arising from simulation. However, as argued by GMR, this scaling effect is relatively minor for values of H say greater than 3 and below 10. Secondly, notice that the major difference between the asymptotic variances at hand is that $b^p(\theta)$ – defined by (25) – substitutes for the *explicit* $\beta^p(\theta)$ function – defined in (7)-(8) –, in the case of the simulation-based estimators. As emphasized above, indirect inference methods typically rely on a numerical mapping between the β vector and θ. At any rate, it is important to note that under the above SALS formulation of the indirect inference estimators, where $b^p(\theta) = \beta^p(\theta)$,

$$\Sigma^{SW}_p(S_0) = \left(1+\frac{1}{H}\right)\Sigma^{ALS}_p(S_0). \tag{28}$$

5 Monte Carlo study

From the asymptotic results – in particular given (28) –, one may think that the ALS estimators are more efficient than those involving simulations. Yet the finite sample evidence in GOURIÉROUX, MONFORT and RENAULT [1993] suggests that the GMR estimator may outperform the asymptotically most efficient ML estimator (though the experiment reported by GMR is restricted to $\theta = .5$). It is not known, however, whether such a superior performance would hold throughout the parameter space. We use a Monte Carlo setup to further explore the estimators' finite sample properties. All simulations were performed with Gauss Version 3.0 using 1,000 replications. We report bias and RMSE in Tables 1 and 2.

We considered $p = 8$ AR lags, with sample sizes $T = 50$ and 200.[7] These settings correspond to those examined by G-ZW and were used for comparison purposes.[8] Besides conformity with existing Monte Carlo studies, empirical relevance justifies our sample size choices. Although, one may argue

7. We have also considered 12-lags. For the sake of space, and since the 12-lags results are no more informative than the case of 8-lags, we only report results with $p = 8$.

8. GMR consider $T = 250$, $H = 1$, $p = 3$ and $\theta = .5$.

that indirect inference methods are basically intended for *large* samples, the fact remains that large samples in time series context do call, in practice, for (relatively) high frequency data sets which are not typically fit as MA(1). The question we are specifically addressing here is small sample performance.

Four values of θ were considered namely, $\theta = .10, .50, .90, .99$. As mentioned above, the simulation-based estimators were implemented with an

TABLE 1
Bias and RMSE of the G-ZW and the ALS Estimators

T	θ	$\hat{\theta}^{G-ZW}_{(T,p)}$		$\hat{\theta}^{ALS}_{(T,p)}$ WEIGHT $= I$		$\hat{\theta}^{ALS}_{(T,p)}$ WEIGHT $= \hat{\Omega}_{(T)}$	
		RMSE	Bias	RMSE	Bias	RMSE	Bias
50	.10	.1664	.0029	.1940	.0080	.2375	.0068
200	.10	.0730	– .0004	.0746	.0002	.0845	.0029
50	.50	.1666	– .0069	.1525	– .0040	.1656	– .0088
200	.50	.0730	– .0024	.0677	.0002	.0732	.0000
50	.90	.1778	– .0526	.1725	.0055	1.9179	.2092
200	.90	.0854	– .0411	.0750	– .0644	.1266	– .0623
50	.99	.2073	– .1157	.3521	– .1160	1.8281	.0107
200	.99	.1296	– .1048	.1384	– .1326	.1299	– .1228

Note: $\hat{\theta}^{G-ZW}_{(T,p)}$ *denotes the G-ZW estimators defined in (22), and* $\hat{\theta}^{ALS}_{(T,p)}$ *refers to the ALS estimator defined in (23).*

TABLE 2
Bias and RMSE of the Simulation-Based Estimators

T	θ	$\hat{\theta}^{SW}_{(T,1,p)} \equiv \hat{\theta}^{SW}_{(T1,p)}$		$\hat{\theta}^{SS}_{(T,1,p)} \equiv \hat{\theta}^{SS}_{(T1,p)}$		$\hat{\theta}^{SS}_{(T3,p)}$	
		RMSE	Bias	RMSE	Bias	RMSE	Bias
50	.10	.2455	– .0184	.2455	– .0085	.2451	– .0122
200	.10	.1042	.0002	.1128	– .0039	.1128	– .0024
50	.50	.2000	– .0714	.2379	– .0857	.2408	– .0865
200	.50	.0874	– .0111	.1001	– .0270	.1041	– .0245
50	.90	.2056	– .0389	.2409	– .1814	.2377	– .1718
200	.90	.0899	.0008	.1019	– .0702	.1012	– .0641
50	.99	.1681	.0442	.2906	– .2472	.2867	– .2379
200	.99	.0836	– .0374	.1488	– .1296	.1481	– .1279

Note: $\hat{\theta}^{SW}_{(T,1,p)}$ *and* $\hat{\theta}^{SW}_{(T1,p)}$ *are the GMR estimators defined in (15) and (18) with* $H = 1$. $\hat{\theta}^{SS}_{(T,1,p)}$ *refers to the EMM estimator defined in (21) with* $H = 1$. $\hat{\theta}^{SS}_{(TH,p)}$ *for* $H = 1,3$ *are the EMM estimators corresponding to (19).*

identity weighting matrix, following the suggestion of GMR.[9] The theory on ALS estimation recommends optimal weighting, yet given our choice for the simulation-based estimators, we considered both identity and optimal weights for the ALS estimators. To conform with GMR's experiment, we examined $H = 1,3$. The results dealing with the GMR estimator with $H = 3$ are qualitatively similar to the $H = 1$ case. For brevity, we report only the $H = 1$ case. Yet in the case of the EMM estimator, given its performance was somewhat disconcerting, we also report the $H = 3$ experiment. More simulated paths were also considered (up to 10 following the recommendation in GMR). Our results may be summarized as follows.

For low values of θ, we see that the relative performance of the estimators studied is in line with the asymptotic efficiency ranking. As there is little dependence in the series, this is not surprising particularly with $T = 200$. For θ at .5 this ranking is no longer upheld. Specifically, the (identity-weighted) ALS estimator outperforms the G-ZW estimator, but the simulation-based estimators still lag behind.

As θ approaches the noninvertibility region, like for instance $\theta = .99$, the effects of simulations become clear. For instance, the RMSE of the G-ZW estimator (for $T = 200$) is .1296 while it is only .0836 in the case of the GMR estimator. The latter estimator's bias is also much lower (the biases are $-.1048$ and $-.0374$ respectively). However, the score-based estimator does not seem to do as well, neither in RMSE nor in bias terms. As mentioned above, we examined whether this is due to the fact that H was set to one. We thus increased H. The results in Table 2 (with $H = 3$) show that this had little effect on the performance of the EMM estimator.[10] In fact, the non-sensitivity to H is worth noting in this case, although this is somewhat implicit in the results of GMR. The performance of the ALS estimator with weight $= \hat{\Omega}_{(T)}$ is dramatically poor in smaller samples for $\theta \geqslant .9$.

Finally, a few comments about computational advantages. In terms of execution ease, no significant differences were noted in convergence time. Specifically, the ALS and simulation based estimation algorithms converged with comparable computer time (with a slight advantage in favor of the GMR estimator). With respect to ALS, this suggests that simulation-based estimates of the *binding function* (or the estimating equations, see section 3) improve numerical stability, which in a way compensates for the H effect. Furthermore, although the score-based estimator is in principle less expensive that the *Wald*-type (does not demand to estimate β), in this case, no notable advantage was observed, which is expected given that the auxiliary estimation uses OLS.

Our results suggest that indirect estimators of the *Wald*-type seem to perform best at the boundary. In view of the results on the ALS estimators (which, along with the G-ZW estimator, does not require normality), it is interesting to envisage new estimation techniques which combine ALS and

9. We have tried optimally-weighted versions of these estimators with qualitatively similar results. Our trials were motivated by the unfavorable performance of the EMM estimator, as will be discussed later. Our results are in the direction of GMR's suggestion in this regard (at least for the example at hand).

10. A minor improvement is noted for smaller samples. With $H > 3$, (results not reported here), the latter pattern was not sustained.

simulation. Unlike the GMR estimator which depends importantly on drawings from a correctly specified structural model, there is an emerging literature (see, for example, DRIDI [2000], DRIDI and RENAULT [2000]) which focuses on procedures more robust to the structural model specification.

6 Conclusion

In this paper, we have shown, through a simple Monte Carlo experiment, that *Wald*-type indirect inference methods appear to perform relatively better than EMM, ALS and the G-ZW estimator, at the parameter space boundary. Unless explicit unit-root modelling is envisaged (a technique not studied in this paper), the GMR estimator seems the best choice available at the boundary. One further advantage of this procedure is that extensions to higher order MA and ARMA models are indeed straightforward, in the sense that no methodological nor statistical modifications are called for. Given the present day technology, we can safely say that the computational efforts involved are practically minor. For illustrative purposes, the reader may refer to GALBRAITH and ZINDE-WALSH [1997] where the extension of their simple estimator to the general ARMA model is discussed. Although the ARMA estimator they propose is still (relatively) computationally attractive, it is evident that more involved derivations had to be dealt with to extend their procedure to the ARMA context. Furthermore, indirect inference may be generalized to multivariate models. Such extensions are a bit more involved, but are conceptually uncomplicated. ▼

• References

ANSLEY C.F. (1979). – « An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process », *Biometrika*, 66, p. 59-65.

ANSLEY C.F., NEWBOLD P. (1980). – « Finite Sample Properties of Estimators for Autoregressive Moving Average Models », *Journal of Econometrics*, 13, p. 159-85.

BOX G.E.P., JENKINS G.M. (1976). – *Time Series Analysis: Forecasting and Control*, (Revised ed.) Holden-Day, San Francisco.

DAVIS R.A., CHEN M., DUNSMUIR W.T.M. (1994). – « Inference for MA(1) Processes with a Unit Root on or Near the Unit Circle », *Probability and Mathematical Statistics*, 15, p. 227-242.

DAVIS R.A., DUNSMUIR W.T.M. (1996). – « Maximum-Likelihood Estimation for MA(1) Processes with a Unit Root on or Near the Unit Circle », *Econometric Theory*, 12, p. 1-9.

DRIDI R. (2000). – « Simulated Asymptotic Least Squares Theory », Department of Economics, London School of Economics, *Working Paper*.

DRIDI R., RENAULT E. (2000). – « Semiparametric Indirect Inference », Department of Economics, London School of Economics, *Working Paper*.

DUFFIE D., SINGLETON K. (1993). – « Simulated Moments Estimation of Markov Models of Asset Prices », *Econometrica*, 61, p. 929-952.

FULLER W.A. (1976). – *Introduction to Statistical Time Series*, Wiley, New York.

GALBRAITH J.W., ZINDE-WALSH V. (1994). – « A simple, Non-Iterative Estimator for Moving Average Models », *Biometrika*, 81, p. 143-156.

GALBRAITH J.W., ZINDE-WALSH V. (1997). – « On Simple, Auto-Regressive-Based Estimation and Identification Techniques for ARMA Models », *Biometrika*, 84, p. 685-696.

GALBRAITH J.W., ZINDE-WALSH V. (2001). – « Analytical Indirect Inference », *Working Paper* presented at the 2001 meetings of the European Economic Association.

GALLANT A.R., TAUCHEN G. (1996). – « Which Moments to Match », *Econometric Theory*, 12, p. 657-681.

GODOLPHIN E.J. (1977). – « A Direct Representation for the Maximum Likelihood Estimation of the Parameters of a Gaussian Moving Average Process », *Biometrika*, 64, p. 375-84.

GOURIÉROUX C., MONFORT A. (1996). – « Simulation-Based Econometric Methods », Core *lectures*, Oxford University Press.

GOURIÉROUX C., MONFORT A., RENAULT E. (1993). – « Indirect Inference », *Journal of Applied Econometrics*, 8, p. S85-S118.

GOURIÉROUX C., MONFORT A., TROGNON A. (1985). – « Moindres carrés asymptotiques », *Annales de l'INSEE*, 58, p. 91-122.

HANNAN E.J., RISSANEN J. (1982). – « Recursive Estimation of Mixed Autoregressive Moving Average Order », *Biometrika*, 69, p. 81-94.

HARRIS D. (1999). – « GMM Estimation of Time Series Models », in Matyas, L. (Ed.), « *Generalized method of moments estimation* », Cambridge University Press.

INGRAM B., LEE B.S. (1991). – « Simulation Estimation of Time Series Models », *Journal of Econometrics*, 47, p. 197-205.

KOREISHA S., PUKKILA T. (1990). – « A Generalized Least Squares Approach for Estimation of Autoregressive Moving Average Models », *Journal of Time Series Analysis*, 11, p. 139-151.

LAROQUE G., SALANIÉ B. (1989). – « Estimation of Multi-Market Fix-Price Models: An Application of Pseudo Maximum Likelihood Methods », *Econometrica*, 57, p. 831-860.

MCCABE B.P.M., LEYBOURNE S.J. (1998). – « On Estimating an ARMA Model with an MA Unit Root », *Econometric Theory*, 14, p. 326-338.

MCCLAVE E.J. (1973). – « On the Bias of AR Approximation to Moving Averages », *Biometrika*, 60, p. 599-605.

MENTZ R.P. (1977). – « Estimation in the First Order Moving Average Model Through the Finite Autoregressive Approximation: Some Asymptotic Results », *Journal of Econometrics*, 6, p. 225-236.

OSBORN D.R. (1977). – « Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes », *Journal of the Royal Statistical Society*, B, 39, p. 144-118.

SHEPPARD N. (1993). – « Distribution of the ML Estimator of an MA(1) and a Local Level Model », *Econometric Theory*, 9, p. 377-401.

SMITH A. (1993). – « Estimating Nonlinear Time Series Models Using Simulated Vector Autoregresions », *Journal of Applied Econometrics*, 8, p. S63-S84.

TAUCHEN G. (1998). – « The Objective Function of Simulation Estimators near the Boundary of the Unstable Region of the Parameter Space », *Review of Economics and Statistics*, LXXX, p. 389-398.