CONSTRAINED EMM AND INDIRECT INFERENCE ESTIMATION

Giorgio Calzolari, Gabriele Fiorentini and Enrique Sentana

WP-AD 2000-26

Correspondence to G. Fiorentini: University of Alicante. Departamento de Fundamentos del Análisis Económico. Carretera San Vicente del Raspeig, s/n. 03071 Alicante. Tel.: 34 96 590 35 63 / E-mail: ga…@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, s.a.

First Edition November 2000.

Depósito Legal: V-4602-2000

IVIE working papers o¤er in advance the results of economic research under way in order to encourage a discussion process before sending them to scienti…c journals for their …nal publication.

* We are grateful to Torben Andersen, Manuel Arellano, Frank Diebold, Russell Davidson, Ramdan Dridi, Lars Hansen, Jim Heckman, Hidehiko Ichimura, Adrian Pagan, Eric Renault and Neil Shephard, as well as seminar audiences at CEMFI, Chicago, UCL and the 1999 European Meeting of the Econometric Society for very helpful comments and suggestions. Special thanks are due to George Tauchen for his invaluable advice. Of course, the usual caveat applies. Financial support from CICYT, IVIE and MURST through the project "Statistical Models for Time Series Analysis" is gratefully acknowledged.

** G. Calzolari: University of Florence; G. Fiorentini: University of Alicante, ISIS-JRC and Ivie; E. Sentana: CEMFI.

CONSTRAINED EMM AND INDIRECT INFERENCE ESTIMATION

Giorgio Calzolari, Gabriele Fiorentini and Enrique Sentana

A B S T R A C T

We develop generalised indirect inference procedures that handle equality and inequality constraints on the auxiliary model parameters. We obtain expressions for the optimal weighting matrices, and discuss as examples an ma(1) estimated as ar(1), an ar(1) estimated as ma(1), and a log-normal stochastic volatility process estimated as a garch(1,1) with Gaussian or t distributed errors. In the first example, the constraints have no effect, while in the second, they allow us to achieve full efficiency. As for the third, neither procedure systematically outperforms the other, but equality restricted estimators are better when the additional parameter is poorly estimated.

Keywords: Simulation estimators, GMM, Minimum distance, ARCH, stochastic volatility:

JEL: C13, C15

1 INTRODUCTION

Consider a stochastic process, x_t, characterised by the sequence of parametric conditional densities $p(x_t|x_{t-1}, x_{t-2}, \ldots; \rho)$, where ρ denotes the d parameters of interest. Consider also a possibly misspecified, auxiliary model, described by the sequence of conditional densities $f(x_t|x_{t-1}, x_{t-2}, \ldots; \theta)$, where θ is a q dimensional vector of parameters, with $d \leq q$. In those situations in which no closed-form expression for $p(x_t|x_{t-1}, x_{t-2}, \ldots; \rho)$ exists, but at the same time it is easy to compute expectations of possibly nonlinear functions of x_t, either analytically, or by simulation or quadrature, the so-called efficient method of moments (EMM) of Gallant and Tauchen (1996) (GT) is a computationally convenient indirect inference (II) procedure, which uses the score of the auxiliary model to derive a generalised method of moments (GMM) estimator of ρ (see Hansen, 1982).

Existing EMM procedures, though, assume that the parameters of the auxiliary model are unrestricted, and consequently, that their pseudo maximum likelihood (ML) estimators have asymptotically normal distribution with a full rank covariance matrix under standard regularity conditions (see e.g. Gourieroux, Monfort and Trognon (1984) or White (1982) for a discussion of unconstrained pseudo ML estimation). Nevertheless, in many situations of interest, some inequality restrictions on θ are usually taken into account in the estimation of the auxiliary model because (i) they lead to more efficient estimates under correct specification, (ii) the pseudo log-likelihood function may not be well defined when the restrictions are violated, or (iii) some of the auxiliary parameters may become underidentified in certain regions of the parameter space. Importantly, such parameter restrictions are often binding in empirical applications.

In this paper, we show how EMM procedures can be generalised to handle such situations. In particular, we propose an alternative set of moment restrictions based on the Kuhn-Tucker first order conditions, which nest the usual ones when the inequality constraints are not binding, but which remain valid even if they are. We also derive the corresponding optimal GMM weighting matrix, and explain how it can be consistently estimated in practice. In this respect, we consider not only the usual two-step GMM method proposed by GT, but also a continuously updated one (à la Hansen, Heaton and Yaaron, 1996). In addition, we combine the constrained parameter estimators and Kuhn-Tucker multipliers to extend the original class of minimum distance (MD) II estimators introduced by Smith (1993) and Gourieroux, Monfort and Renault (1993) (GMR) to the inequality restricted case. It turns out that like in the unconstrained case (see Gourieroux and Monfort, 1996) (GM96), one can …nd inequality restricted II estimators that are asymptotically equivalent to the inequality constrained EMM estimators by an appropriate choice of weighting matrix.

It is important to bear in mind that our results in no way require that the restrictions are correct, in the sense that they are satis…ed by the unrestricted pseudo-true values of the auxiliary parameters. Of course, if we knew that this was indeed the case, we might be able to obtain more e¢cient estimators of the parameters of interest (see Dridi, 2000). It is also worth mentioning that although we concentrate on pseudo log-likelihood estimation of the auxiliary model for expositional purposes, our procedures can be extended to cover any other extremum estimators of just identi…ed auxiliary models, such as M-estimators or method of moments (see section 4.1.3 of GM96).

We also discuss EMM and II procedures based on equality constrained pseudo ML estimators of μ, as well as on those that combine equality and inequality constraints. Equality restricted procedures may be particularly useful in practice from a computational point of view, because in many situations of signi…cant empirical interest, it is considerably simpler to estimate a special restricted case of the auxiliary model than to maximise the unrestricted log-likelihood function.

For the same reason, we also consider II procedures based on partially optimised unconstrained estimators that do not satisfy the standard first order conditions for extrema of the pseudo log-likelihood function, as well as those that impose the constraints depending on the significance of some preliminary specification test.

For illustrative purposes, we apply our modified procedures to three time series models. The first two are (i) an ma(1) process, estimated either as an ar(1) with a non-negativity constraint on the autoregressive coefficient, or as white noise, and (ii) an ar(1) process, estimated either as an ma(1) with a non-positivity constraint on the moving average coefficient, or as white noise. The third model that we study is the popular discrete time version of the log-normal stochastic volatility process, which we estimate via a garch(1,1) model with either t distributed errors, or Gaussian ones. This model is important in its own right, and has become the acid test of any simulation-based estimation method. In addition, it also helps to illustrate the implementation of our proposed procedures in some non-standard situations. In particular, the pseudo log-likelihood function based on the t distribution cannot be defined in part of the neighbourhood of the parameter values that correspond to the Gaussian case, and moreover, some of the auxiliary model parameters become underidentified under conditional homoskedasticity.

The paper is organized as follows. In section 2, we include a thorough discussion of EMM and II procedures with either equality or inequality constraints on the auxiliary model parameters. Since it is often impossible to obtain some of the required expressions in closed form, we also discuss how they can be evaluated by simulation. Detailed applications of such procedures to the three examples can be found in section 3. Finally, our conclusions are presented in section 4. Proofs and auxiliary results are gathered in the appendix.

2 THEORETICAL SET UP

2.1 Inequality constrained EMM and II estimators

Let $l_t(\theta) = \ln f(x_t|x_{t-1}, x_{t-2}, \ldots, \theta)$. The pseudo log-likelihood function for a sample of size T on x_t based on the auxiliary model (ignoring initial conditions) will be given by $L_T(\theta) = \sum_t l_t(\theta)$. Let's now define the (scaled) Lagrangian function

$$Q_T(\beta) = \frac{1}{T}L_T(\theta) + h'(\theta)\gamma \tag{1}$$

where $\beta = (\theta', \gamma')'$, and γ are the s multipliers associated with s mutually consistent inequality constraints implicitly characterized by $h(\theta) \geq \mathbf{0}$. Assuming that both the pseudo-log likelihood function $L_T(\theta)$, and the vector of functions $h(\theta)$ are twice continuously differentiable with respect to θ, the latter with a full column rank Jacobian matrix $\partial h'(\theta)/\partial\theta$, the first-order conditions that take into account the inequality constraints will be given by the usual Kuhn-Tucker conditions:

$$\frac{\partial Q_T(\tilde{\beta}_T)}{\partial\theta} = \frac{1}{T}\frac{\partial L_T(\tilde{\theta}_T)}{\partial\theta} + \frac{\partial h'(\tilde{\theta}_T)}{\partial\theta}\tilde{\gamma}_T = \mathbf{0} \tag{2}$$

together with the sign and exclusion restrictions:

$$h(\tilde{\theta}_T) \geq \mathbf{0} \qquad \tilde{\gamma}_T \geq 0 \qquad h(\tilde{\theta}_T) \odot \tilde{\gamma}_T = \mathbf{0} \tag{3}$$

where $\sim$ indicates inequality restricted pseudo-ML estimators, the subscript T refers to the sample size of the observed series, and the symbol $\odot$ denotes the Hadamard (or element by element) product of two matrices of the same dimensions.

Standard EMM procedures cannot be used in this context because, as we shall see below, the expected value of the score of the auxiliary model is no longer necessarily zero when some of the restrictions of the auxiliary model are binding. Nevertheless, a modified procedure can be derived from (2). Specifically, we can base our estimation of ρ on the following moments:

$$m_T(\rho;\beta) = E\left[\left.\frac{\partial Q_T(\beta)}{\partial\theta}\right|\rho\right] = E\left[\left.\frac{1}{T}\sum_t \frac{\partial l_t(\theta)}{\partial\theta} + \frac{\partial h'(\theta)}{\partial\theta}\mu\right|\rho\right] \tag{4}$$

where the symbol $E(.|\rho)$ refers to an expected value computed with respect to the distribution of the model of interest evaluated at ρ. The main difference with the unrestricted case is that $m_T(\rho;\beta)$ not only depends on the q auxiliary model parameters θ, but also on the s Kuhn-Tucker multipliers μ associated with the inequality restrictions. In this respect, note that if we define

$$\mathcal{L}_T(\rho;\theta) = E\left[\left.\frac{1}{T}L_T(\theta)\right|\rho\right] \tag{5}$$

we can interpret $m_T(\rho;\beta) = \mathbf{0}$ as the first-order conditions of the population program

$$\max_\theta \mathcal{L}_T(\rho;\theta) \quad s.t. \quad h(\theta) \geq \mathbf{0} \tag{6}$$

as long as the differentiation and expectation operators can be interchanged, which we assume henceforth. We also assume that $\mathcal{L}_T(\rho;\theta)$ is twice continuously differentiable with respect to both θ and ρ. Importantly, in those time series situations in which the functional form of $l_t(\theta)$ is time-invariant, and x_t strictly stationary, the dependence of the moments on T disappears, and expressions (5) and (4) simplify to

$$\mathcal{L}(\rho;\theta) = E\left[l_t(\theta)|\rho\right]$$

$$m(\rho;\beta) = E\left[\left.\frac{\partial l_t(\theta)}{\partial\theta}\right|\rho\right] + \frac{\partial h'(\theta)}{\partial\theta}\mu$$

For each value of ρ, we can define a deterministic sequence of binding functions for the inequality constrained auxiliary parameters θ and associated Kuhn-Tucker multipliers μ, $\beta_T^i(\rho) = \left[\theta_T^{i\prime}(\rho); \mu_T^{i\prime}(\rho)\right]'$ say, such that they solve the population program (6). As a result, these functions must satisfy the first order conditions:

$$
\begin{aligned}
& m_T\left[\rho;\beta_T^{i}(\rho)\right] = 0 \\
& \left[\theta_T^{i}(\rho)\right] \geq \mathbf{0} \qquad \mu_T^{i}(\rho) \geq \mathbf{0} \qquad \left[\theta_T^{i}(\rho)\right] \odot \mu_T^{i}(\rho) = \mathbf{0}
\end{aligned} \tag{7}
$$

and obviously become time-invariant under strict stationarity. To guarantee the identification of ρ, we assume that for all T larger than a given value, $\beta_T^{i}(\rho)$ is the only such solution, and that the equation $\beta_T^{i}(\rho) = \beta$ admits a unique solution in ρ (cf. GM96).

Let ρ^0 denote the true value of the parameters of interest, and let $\theta_T^{i}(\rho^0)$ and $\mu_T^{i}(\rho^0)$ denote the inequality constrained pseudo-true values for θ and μ. If we knew these values, we could recover ρ^0 by either inverting the binding functions, or solving the possibly non-linear system of equations $m_T\left[\rho^0;\beta_T^{i}(\rho^0)\right] = 0$. In practice, though, we do not know the pseudo true values, but if they are consistently estimated by the auxiliary model, we can obtain consistent estimators of ρ^0 by choosing the parameter values that minimize either some appropriately defined distance between $\beta_T^{i}(\rho)$ and $\tilde{\beta}_T$, or a given norm of the sample moments $m_T(\rho;\tilde{\beta}_T)$. In particular, we can minimise with respect to ρ the following quadratic forms:

$$
D_T^{i}(\rho;\mathbf{\Omega}) = \left[\beta_T^{i}(\rho) - \tilde{\beta}_T\right]' \cdot \mathbf{\Omega} \cdot \left[\beta_T^{i}(\rho) - \tilde{\beta}_T\right]
$$

or

$$
G_T(\rho;\mathbf{\Psi}) = m_T'(\rho;\tilde{\beta}_T) \cdot \mathbf{\Psi} \cdot m_T(\rho;\tilde{\beta}_T)
$$

where $\mathbf{\Omega}$ and $\mathbf{\Psi}$ are positive semi-definite (p.s.d.) weighting matrices of orders $q+s$ and q respectively, and the letters D and G are a reminder that these objective functions correspond to MD and GMM estimation criteria respectively. In what follows, we shall refer to the resulting estimators

$$
\begin{aligned}
\tilde{\rho}_T^{D}(\mathbf{\Omega}) &= \arg\min_{\rho} D_T^{i}(\rho;\mathbf{\Omega}) \\
\tilde{\rho}_T^{G}(\mathbf{\Psi}) &= \arg\min_{\rho} G_T(\rho;\mathbf{\Psi})
\end{aligned}
$$

as the inequality restricted II and EMM estimators of ρ. Obviously, without a judicious choice of metric that accounts for sample variation in the estimators of the inequality restricted auxiliary parameters and multipliers in $\tilde{\beta}_T$, the asymptotic covariance matrix of $\tilde{\rho}_T^D(\mathbf{\Omega})$ and $\tilde{\rho}_T^G(\mathbf{\Psi})$ is likely to be unnecessarily large.

Let's start by analysing the second criterion function. It is well known that if the sample moments $m_T(\rho;\tilde{\beta}_T)$ have a limiting normal distribution, the optimal GMM weighting matrix (in the sense that the difference between the covariance matrices of the resulting estimator and an estimator based in any other norm is p.s.d.) is given by the inverse of the asymptotic variance of $\sqrt{T}m_T(\rho;\tilde{\beta}_T)$ (see e.g. Hansen, 1982). In order to derive the required asymptotic distribution, we follow GT in assuming the necessary regularity conditions for $\tilde{\beta}_T$ to converge uniformly almost surely to $\beta_T^i(\rho^0)$, and for a strong law of large numbers and a central limit theorem to apply to the Hessian and modified score of the log-likelihood of the auxiliary model respectively. More formally,

Assumption 1

$$P\left[\lim_{T\to\infty}\sup\left\|\begin{matrix}\tilde{\theta}_T-\theta_T^i(\rho^0)\\ \tilde{\pi}_T-\pi_T^i(\rho^0)\end{matrix}\right\|=0\right]=1$$

$$P\left\{\lim_{T\to\infty}\left\|\frac{1}{T}\sum_t\frac{\partial^2 l_t(\theta_T^*)}{\partial\theta\partial\theta'}-\mathcal{J}_{0T}^i\right\|=0\right\}=1$$

$$\sqrt{T}\frac{1}{T}\sum_t\left\{\frac{\partial l_t\left[\theta_T^i(\rho^0)\right]}{\partial\theta}+\frac{\partial h'\left[\theta_T^i(\rho^0)\right]}{\partial\theta}\pi_T^i(\rho^0)\right\}\to N(\mathbf{0};\mathcal{I}_{0T}^i)$$

where $\mathcal{J}_{0T}^i$ and $\mathcal{I}_{0T}^i$ are non-stochastic, $q\times q$ matrices, with $\mathcal{I}_{0T}^i$ p.d., and θ_T^* is any sequence that converges in probability to $\theta_T^i(\rho^0)$.

In this respect, it is important to note that there are many situations in which the pseudo log-likelihood function is not well-defined outside the restricted parameter space, and yet the (possibly directional) score and Hessian behave regularly at its boundary (see e.g. the score of the Student's t GARCH model in section 3.3 below under conditional Gaussianity).

However, we cannot directly rely on the results in GT to derive the asymptotic distribution of these sample moments, since the inequality restricted estimator $\tilde{\theta}_T$ may not be asymptotically normal in large samples (see Andrews (1999) and the references therein). In addition, the asymptotic distribution of $\tilde{\beta}_T$ is singular, in the sense that there are s linear combinations of the elements of $\sqrt{T}\left[\tilde{\beta}_T - \beta_T^i(\rho_0)\right]$ that converge in probability to 0. Specifically:

Proposition 1 Under Assumption 1,

$$\mu_T^i(\rho_0) \odot \frac{\partial h\left[\theta_T^i(\rho_0)\right]}{\partial \theta'} \sqrt{T}\left[\tilde{\theta}_T - \theta_T^i(\rho_0)\right] + h\left[\theta_T^i(\rho_0)\right] \odot \sqrt{T}\left[\tau_T - \mu_T^i(\rho_0)\right] = o_p(1)$$

In contrast, there are q linear combinations that are asymptotically well behaved:

Proposition 2 Under Assumption 1,

$$\left[\mathcal{J}_{0T}^i + \left[\mu_T^i(\rho_0) \otimes I_q\right] \frac{\partial vec\left\{\partial h'\left[\theta_T^i(\rho_0)\right]/\partial \theta\right\}}{\partial \theta'}\right] \sqrt{T}\left[\tilde{\theta}_T - \theta_T^i(\rho_0)\right]$$
$$+ \frac{\partial h'\left[\theta_T^i(\rho_0)\right]}{\partial \theta} \sqrt{T}\left[\tau_T - \mu_T^i(\rho_0)\right]$$
$$+ \sqrt{T}\frac{1}{T}\sum_t \left\{ \frac{\partial l_t\left[\theta_T^i(\rho_0)\right]}{\partial \theta} + \frac{\partial h'\left[\theta_T^i(\rho_0)\right]}{\partial \theta} \mu_T^i(\rho_0) \right\} = o_p(1)$$

Hence, even though $\tilde{\theta}_T$ and τ_T have a singular and possibly non-Gaussian asymptotic distribution, Proposition 2 shows that under our regularity conditions, there are always q linear combinations that are asymptotically normally distributed, irrespectively of the exact nature of the inequality restrictions, and irrespectively of whether the sign restrictions on $h\left[\theta_T^i(\rho_0)\right]$ and $\mu_T^i(\rho_0)$ in (7) are satisfied with equality, or strictly so. It turns out that those q linear combinations are implicitly contained in the expected value of the modified score:

Proposition 3 Under Assumption 1,

$$\sqrt{T}\, m_T(\rho^0;\tilde{\beta}_T)+\sqrt{T}\frac{1}{T}\sum_t\left\{\frac{\partial l_t\left[\theta_T^-(\rho^0)\right]}{\partial\theta}+\frac{\partial h'\left[\theta_T^-(\rho^0)\right]}{\partial\theta}\mu_T^-(\rho^0)\right\}=o_p(1)$$

Therefore, $\sqrt{T}\, m_T(\rho^0;\tilde{\beta}_T)$ has indeed a limiting Gaussian distribution, and the optimal GMM weighting matrix is precisely the inverse of $\mathcal{I}_{0T}^-$.

The following proposition specifies the asymptotic distribution of the (infeasible) optimal GMM estimator of ρ based on the inequality restricted auxiliary model:

Proposition 4

$$\sqrt{T}\left[\tilde{\rho}_T^G(\mathcal{I}_{0T}^-)-\rho^0\right]\rightarrow N\left[\mathbf{0};\left\{\frac{\partial m_T'\left[\rho^0;\beta_T^-(\rho^0)\right]}{\partial\rho}\cdot(\mathcal{I}_{0T}^-)^{-1}\cdot\frac{\partial m_T\left[\rho^0;\beta_T^-(\rho^0)\right]}{\partial\rho'}\right\}^{-1}\right]$$

Given that this expression is completely analogous to the one derived by GT for the optimal EMM estimator in the absence of constraints, the required matrices can also be consistently estimated using their suggested procedures. In particular, since in those cases in which

$$E\left[\partial l_t\left[\theta_T^-(\rho^0)\right]/\partial\theta+\left(\partial h'\left[\theta_T^-(\rho^0)\right]/\partial\theta\cdot\mu_T^-(\rho^0)\right|\rho^0\right]=0\quad\forall t;$$

such as strictly stationary and ergodic time series processes with absolutely summable autocovariance matrices, $\mathcal{I}_{0T}^-$ converges to $\mathcal{I}_0^-=\sum_{\tau=-\infty}^{\infty}S_\tau\left[\rho^0;\beta_T^-(\rho^0)\right]$, where

$$S_\tau(\rho;\beta)=E\left\{\left[\frac{\partial l_t(\theta)}{\partial\theta}+\frac{\partial h'(\theta)}{\partial\theta}\mu\right]\left[\frac{\partial l_{t-\tau}(\theta)}{\partial\theta}+\frac{\partial h'(\theta)}{\partial\theta}\mu\right]'\Big|\beta=\rho\right\}$$

for $\tau\geq 0$ and $S_\tau(\rho;\beta)=S'_{-\tau}(\rho;\beta)$ for $\tau<0$ (see e.g. Hansen, 1982), we could obtain a consistent estimate of the matrix $\mathcal{I}_{0T}^-$ as

$$\hat{\mathcal{I}}_T^-=\sum_{\tau=-T^*}^{T^*}w(\tau)\hat{S}_{\tau T}\tag{8}$$

with

$$\tilde{S}_{\tau T} = \frac{1}{T}\sum_{t=\tau+1} \left\{ \left[\frac{\partial l_t(\tilde{\theta}_T)}{\partial \theta} + \frac{\partial h'(\tilde{\theta}_T)}{\partial \theta}\iota_T \right] \left[\frac{\partial l_{t-\tau}(\tilde{\theta}_T)}{\partial \theta} + \frac{\partial h'(\tilde{\theta}_T)}{\partial \theta}\iota_T \right]' \right\}$$

where $w(\tau)$ are weights suggested by a standard heteroskedasticity and autocorrelation consistent (HAC) covariance estimation procedure, and ι the corresponding rate (see e.g. de Jong and Davidson (2000) and the references therein). Then, a feasible optimal GMM estimator will be given by $\tilde{\rho}^D_T(\tilde{\Sigma}^{-}_T)$. Alternatively, we could consider continuously updated GMM estimators à la Hansen, Heaton and Yaaron (1996), by replacing $\tilde{S}_{\tau T}$ in the above expressions with $S_\tau(\rho;\tilde{\beta}_T)$.

Let's now turn to the II estimators of ρ based on the MD function $D^{-}_T(\rho;\mathbf{\Omega}_T)$. Unfortunately, we cannot directly rely on standard MD theory, because as we saw before, the limiting distribution of $\sqrt{T}\left[\tilde{\beta}'_T - \beta^{-\prime}_T(\rho^0)\right]$ is singular and possibly non-normal. To overcome this difficulty, it is convenient to write down the linear transformations in Propositions 1 and 2 together in terms of the following square matrix of order $q+s$:

$$\mathcal{K}^{-}_{0T} = \begin{bmatrix} \mathcal{J}^{-}_{0T} + [\mu^{-}_T(\rho^0)\otimes I_q]\,\partial vec\left\{\partial h'\left[\theta^{-}_T(\rho^0)\right]/\partial\theta\right\}/\partial\theta' & \partial h'\left[\theta^{-}_T(\rho^0)\right]/\partial\theta \\ diag\,[\mu^{-}_T(\rho^0)]\,\partial h\left[\theta^{-}_T(\rho^0)\right]/\partial\theta' & diag\left\{h\left[\theta^{-}_T(\rho^0)\right]\right\} \end{bmatrix}$$

$$= \begin{bmatrix} \mathcal{K}^{-}_{11;0T} & \mathcal{K}^{-}_{12;0T} \\ \mathcal{K}^{-}_{21;0T} & \mathcal{K}^{-}_{22;0T} \end{bmatrix}$$

where $diag\,(.)$ is the operator that transforms a vector into a diagonal matrix of the same order by placing its elements along the main diagonal. Then, if we transform the MD conditions by premultiplying them by $\mathcal{K}^{-}_{0T}$, we will have that the asymptotic distribution of $\mathcal{K}^{-}_{0T}\sqrt{T}\left[\tilde{\beta}'_T - \beta^{-\prime}_T(\rho^0)\right]$ will be normal, with the singularity confined to the last s elements. In this framework, we can prove the following result, which can be regarded as the inequality restricted version of Proposition 4.3 in GM96:

Proposition 5

$$\sqrt{T}\left[\hat{\rho}^G_T(\mathbf{\Psi}_T) - \hat{\rho}^D_T(\tilde{\mathcal{K}}_T'\mathbf{\Psi}^*_T\tilde{\mathcal{K}}_T)\right] = o_p(1)$$

where

$$\mathbf{\Psi}^*_T = \begin{pmatrix} \mathbf{\Psi}_T & 0 \\ 0 & 0 \end{pmatrix}$$

There are some cases of practical relevance in which $\hat{\rho}^D_T\left[\left(\tilde{\mathcal{K}}_T'\mathcal{I}^*_{0T}\right)^+\tilde{\mathcal{K}}_T\right]$, where $+$ denotes the Moore-Penrose generalised inverse, is relatively easy to compute. For instance, suppose that all the restrictions are of the simple "bounds" form, i.e. $\theta_{j\,min} \leq \theta_j \leq \theta_{j\,max}$ $(j = 1, \ldots, q)$, with $|\theta_{j\,min}|, |\theta_{j\,max}|$ possibly infinity, and define $\mu_{j\,min}, \mu_{j\,max}$ as the matching pair of Kuhn-Tucker multipliers (which are set to zero by definition if the corresponding bound is $\pm\infty$). In addition, assume for simplicity that we knew that only one restriction, say the lower limit on the first parameter, is strictly binding in the limit, in the sense that $\lim_{T\to\infty}\tilde{\mu}_{1\,min,T}(\rho^0) > 0$, while all the other parameters are asymptotically strictly unconstrained (i.e. $\theta_{j\,min} < \lim_{T\to\infty}\tilde{\theta}_{jT}(\rho^0) < \theta_{j\,max}$ for $j = 2, \ldots, q$). Then, it is easy to see from Proposition 2 that the $q \times 1$ vector $(\hat{\mu}_{1\,min,T}, \hat{\theta}_{2T}, \ldots, \hat{\theta}_{qT})$ will have an asymptotically normal distribution with a full rank covariance matrix, which can be used to compute the optimal MD estimator of ρ. However, the EMM procedure generally has the advantage that the optimal weighting matrix can be readily computed as the variance of the limiting normal distribution of the modified score (4), irrespectively of the exact nature of the inequality restrictions, and irrespectively of whether the sign restrictions on $h\left[\tilde{\theta}_T(\rho^0)\right]$ and $\tilde{\mu}_T(\rho^0)$ in (7) are satisfied with equality, or strictly so.

Nevertheless, there is one instance in which both our proposed procedures are numerically identical. In particular, suppose that $d = q$, so that the auxiliary model just identifies the parameters of interest, and that all the restrictions are of the simple bounds form. Then, the value of ρ that for $j = 1, \ldots, q$ produces estimates of the triplets $\left[\theta^i_j(\rho), \mu_{q\,min}(\rho), \mu_{q\,max}(\rho)\right]$ that are equal to (i) $(\tilde{\theta}_{jT}, 0, 0)$ if $\theta_{j\,min} < \tilde{\theta}_{jT} < \theta_{j\,max}$, (ii) $(\theta_{j\,min}, \tilde{\mu}_{q\,min}, 0)$ if $\tilde{\theta}_{jT} = \theta_{j\,min}$, or (iii) $(\theta_{j\,max}, 0, \tilde{\mu}_{q\,max})$ if $\tilde{\theta}_{jT} = \theta_{j\,max}$, will also set to zero the sample moments $m_T(\rho, \tilde{\beta}_T)$, and therefore, will be numerically identical to $\tilde{\rho}^{G}_{T}(\Psi)$ for all Ψ.

2.2 Relationship with the existing unrestricted procedures

Let $\hat{\theta}_T$ denote the unconstrained pseudo-ML estimator of the auxiliary model parameters θ, and define $\theta^u_T(\rho^0)$ as the corresponding pseudo-true values, where $\theta^u_T(\rho)$ are the usual binding functions that solve the unrestricted population program $\max_\theta \mathcal{L}_T(\rho, \theta)$, with $\hat{\mu}_T = \mathbf{0} = \mu^u_T(\rho)$. If the auxiliary model is asymptotically strictly unconstrained, in the sense that $\lim_{T\to\infty}\left[\theta^i_T(\rho^0) - \theta^u_T(\rho^0)\right] = 0$, $\lim_{T\to\infty} \mu^i_T(\rho^0) = 0$ and $\lim_{T\to\infty} h\left[\theta^i_T(\rho^0)\right] > \mathbf{0}$, our proposed inequality constrained EMM and II procedures converge to the standard unconstrained EMM and II approaches of GT and GMR, because $\sqrt{T}\tilde{\mu}_T$ and $\sqrt{T}(\tilde{\theta}_T - \hat{\theta}_T)$ converge in probability to 0 from Propositions 1 and 2 respectively. In fact, the inequality constrained and unconstrained procedures will yield numerically identical results if none of the inequality restrictions is binding in a given sample, since in that case $\tilde{\theta}_T$ coincides with the unconstrained pseudo-ML estimator, $\hat{\theta}_T$ (and $\tilde{\mu}_T$ with $\hat{\mu}_T = \mathbf{0}$). Moreover, if the auxiliary model exactly identifies the parameters of interest, all the different procedures will be the same for T sufficiently large (see Proposition 4.1 in GM96).

It may seem at first sight that one can handle inequality restrictions on the parameters of the auxiliary model with the existing unconstrained EMM or II procedures, by simply reparametrising the constraints appropriately. For instance, a non-negativity constraint on θ_j can be formally avoided by replacing θ_j with $\left(\theta_j^*\right)^2$, where $-\infty < \theta_j^* < \infty$. Unfortunately, the regularity conditions in Assumption 1 are no longer satisfied in terms of the new parameter when the pseudo-true value of the original parameter $\theta^i_{jT}(\rho^0)$ converges to its lower bound asymptotically, as the Jacobian of the transformation is 0 at $\theta^i_{jT}(\rho^0) = 0$.

2.3 Equality constrained EMM and mixed procedures

It is easy to see that if we replace the Kuhn-Tucker multipliers by the usual Lagrange multipliers, the theoretical derivations in section 2.1 also apply to EMM procedures based on equality constrained pseudo-ML estimators of the auxiliary model parameters, provided that the set of moments used for GMM estimation include the first order conditions corresponding to all the elements of θ. In particular, if we call $\bar{\theta}_T$ the pseudo-ML estimates of θ that satisfy with equality all the restrictions implicit in $h(\theta)$, and denote by δ_T the associated (ordinary) Lagrange multipliers, the first-order conditions will be given by:

$$\frac{\partial Q_T(\bar{\beta}_T)}{\partial \theta} = \frac{1}{T}\frac{\partial L_T(\bar{\theta}_T)}{\partial \theta} + \frac{\partial h'(\bar{\theta}_T)}{\partial \theta}\delta_T = \mathbf{0} \tag{9}$$

together with $h(\bar{\theta}_T) = \mathbf{0}$. In this context, we can again define the population moments $m_T(\rho; \beta)$ as in (4). Similarly, we can define a deterministic sequence of binding functions for the equality constrained auxiliary parameters θ and associated Lagrange multipliers μ, $\beta^e_T(\rho) = [\theta^{e\prime}_T(\rho); \mu^{e\prime}_T(\rho)]'$ say, such that for each value of ρ, they solve the population program, $\max_\theta \mathcal{L}_T(\theta)$ subject to $h(\theta) = \mathbf{0}$. As a result, these functions must satisfy the first order conditions

$$\begin{aligned} m_T[\rho; \beta^e_T(\rho)] &= 0 \\ h[\theta^e_T(\rho)] &= \mathbf{0} \end{aligned} \tag{10}$$

and again become time-invariant under strict stationarity. Of course, the binding functions $\beta^e_T(\rho)$ will generally be different from $\beta^i_T(\rho)$, which result from imposing the same constraints as inequalities. Similarly, the equality restricted pseudo-true values and limiting matrices $\mathcal{J}^e_{0T}$ and $\mathcal{I}^e_{0T}$ will often differ from $\bar{\beta}^i_T(\rho^0)$ and $\mathcal{J}^i_{0T}$ and $\mathcal{I}^i_{0T}$. Nevertheless, note that the nature of the regularity conditions is the same.

In addition, it is also possible to consider equality restricted II procedures that generalise the GMR approach, by choosing ρ so as to minimise a well-defined distance between the expanded vector of equality constrained parameter estimators and multipliers in the original sample, $\bar{\beta}_T$, and $\bar{\beta}^e_T(\rho)$. The main difference with respect to the inequality constrained case discussed in section 2.1 is that the joint asymptotic distribution of $\bar{\beta}_T$ will be normal (albeit singular) under regularity conditions analogous to the ones in Assumption 1, with ^{i} replaced by ^{e}, and $\sim$ by $-$. In any case, Propositions 1 to 5 continue to hold if we replace inequality restricted estimators and Kuhn-Tucker multipliers by equality restricted estimators and Lagrange multipliers.

Once more, the EMM procedure has the advantage that the optimal weighting matrix can be readily computed as the variance of the limiting normal distribution of the modified score, regardless of the exact nature of the equality restrictions. There are some simple cases, though, in which the asymptotically equivalent II estimators can be easily obtained. For instance, suppose that all s restrictions are of the simple form, $\theta_j = \theta^\dagger_j$ for $j = 1, \ldots, s \leq q$. Then, it is easy to see from Proposition 2 that the $q \times 1$ vector $(\mu_{1,T}, \ldots, \mu_{s,T}, \bar{\theta}_{s+1T}, \ldots, \bar{\theta}_{qT})$ will have an asymptotically normal distribution with a full rank covariance matrix, which can be used to compute the "optimal" equality constrained II estimator of ρ, $\hat{\rho}^D_T\left[\mathcal{K}^{e\prime}_T\left(\mathcal{I}^{e}_{0T}\right)^+\mathcal{K}^e_T\right]$. If, in addition, $p = q$, so that the auxiliary model just identifies the parameters of interest, then the value of ρ that produces values of $\left[\theta^e_j(\rho), \mu^e_j(\rho)\right]$ that are equal to (i) $(\theta^\dagger_j, \mu_{jT})$ for $j = 1, \ldots, s$ and (ii) $(\bar{\theta}_{jT}, 0)$ for $j = s+1, \ldots, q$ will also set to zero the sample moments $m_T\left[\rho, \tilde{\beta}_T\right]$, and therefore, will be numerically identical to $\hat{\rho}^G_T(\mathbf{\Psi})$ for all $\mathbf{\Psi}$.

Equality restricted EMM and II procedures may be particularly useful from a computational point of view, because in many situations of interest, it is considerably simpler to estimate a special restricted case of the auxiliary model than the unrestricted model itself. The extensive literature on LM (or score) tests provides many such examples (see e.g. Godfrey, 1988). For instance, the estimation of a var(p) model is much easier than the estimation of any varma(p,q) model that nests it.

Again, it may seem again at first sight that one can handle equality restrictions on the parameters with the existing unconstrained procedures by re-writing the constraints in explicit form (see e.g. chapter 10 of Gourieroux and Monfort (1995) (GM95) for a thorough discussion). For instance, a simple linear equality constraint of the form $\mu_j + \mu_k = 0$ can be formally avoided by eliminating μ_k (or μ_j) from the active set of parameters, and replacing it with $-\mu_j$ (or $-\mu_k$). However, it is very important to emphasise that in doing so, we would be reducing the number of moments used in the GMM estimation of the parameters of interest, ½, and therefore, incurring in an efficiency loss relative to our proposed procedure. As an extreme example, suppose that $p = q = s$, and that $h(\mu) = \mu - \mu^y$, so that the only admissible value for the equality restricted estimator $\bar{\mu}_T$ is precisely μ^y. In this case, there is no need for any extra parameters in order to re-write the implicit restrictions in explicit form. But then, no unconstrained EMM or II estimator based on those inexistent parameters can be defined. In contrast, our equality constrained II procedure will work by simply matching the q equality restricted binding functions $^1\tilde{e}_T$(½) with the sample estimates of the q Lagrange multipliers.

Our proposed constrained EMM procedures can be trivially extended to handle a mix of equality and inequality constraints, since in all cases the relevant moments adopt the form of (4). Similarly, II procedures that match parameters and a mix of Kuhn-Tucker and Lagrange multipliers can also be entertained.

Finally, it would certainly be desirable to compare the efficiency of the different possible versions of the EMM and asymptotically equivalent II estimators. Unfortunately, it is very difficult to say anything in general terms, even for a given set of implicit constraints $h(\theta)$. The problem is that different types of "constrained" estimators (i.e. unconstrained, equality, inequality or mixed) lead to different sets of moments, which despite their common form, cannot usually be written as a one-to-one function of each other, either in finite samples or asymptotically (but see section 3.1 below). Nevertheless, we can establish the relationship between some of them. In particular, since the inequality estimators of the auxiliary model parameters $\tilde{\theta}_T$, and the associated Kuhn-Tucker multipliers λ_T, will be a mixture of the unrestricted estimators $\hat{\theta}_T$, and every possible restricted estimator that satisfies with equality a subset of the s constraints, then the inequality restricted EMM estimator based on them will also be a mixture (with the same weights) of the unconstrained EMM estimator $\tilde{\rho}_T^G(\mathbf{\Psi})$, and every possible equality restricted EMM estimator. Therefore, the asymptotic distribution of $\sqrt{T}\left[\tilde{\rho}_T^G(\mathbf{\Psi}) - \rho_T^0\right]$ will often coincide with the asymptotic distribution of one of those estimators. The exception is when one (or several) of the constraints is just binding in the limit, in the sense that the pseudo-true value of the corresponding Kuhn-Tucker multiplier converges to zero, but the constraint is satisfied with equality by the unconstrained pseudo-true value. In that case, the inequality constrained EMM estimator will continue to be in large samples a mixture with positive weights of the corresponding equality constrained and unconstrained EMM estimators, but since they are asymptotically equivalent, so will be the inequality constrained one (see sections 3.1 and 3.2 for examples).

In addition, it is worth mentioning that any unconstrained EMM estimator is asymptotically equivalent to an equality constrained EMM estimator that sets all the parameters of the auxiliary model to their unconstrained pseudo-true values, $\mu_T^u(½^0)$. The intuition is that from (9), the associated Lagrange multipliers will coincide with the (minus) score of the unconstrained pseudo-log likelihood function. Therefore, if the true model is "smoothly embedded" within the auxiliary model (see De…nition 1 in GT), and ½ is unconstrained, Theorem 2 in GT show that such an equality constrained EMM estimator will be as e¢cient as the (possibly infeasible) maximum likelihood estimator of ½.

Unfortunately, it is often the case that the auxiliary model does not nest the true model, as the examples in section 3 illustrate. Therefore, we may have situations in which it makes no di¤erence whether or not we impose constraints on µ as far as the estimation of ½ is concerned (see section 3.1), and others in which a constrained estimator is more e¢cient than an unconstrained one (see section 3.2).

2.4 Partially optimised unconstrained EMM and pre-test procedures

It is often the case that an empirical researcher tries to estimate a reasonably complex auxiliary model, in the hope of capturing the most distinctive features of the data, and in this way, coming close to the idealised situation covered by Theorem 2 in GT. Unfortunately, such attempts often encounter numerical optimisation problems. It turns out that our results can be easily adapted to cover such a situation as well, at the cost of increasing the complexity of the notation. For simplicity of exposition, we concentrate on EMM procedures, and assume that the auxiliary model is unconstrained, that the numerical procedure used to maximise the pseudo log-likelihood function $L_T(\mu)$ is a standard gradient method (such as Newton-Raphson, scoring, BHHH, steepest ascent, or any Quasi-Newton procedure),[1] that the step size is computed by quadratic approximation, and that the researcher abandons her attempts to maximise the pseudo-log likelihood function after k_{max} steps, with $k_{max} \geq 0$. More specifically, if $\hat{\theta}_T^{(k)}$ denotes the value of the parameters after iteration k ($1 \leq k \leq k_{max}$), we assume that the recursive formula employed is

$$\hat{\theta}_T^{(k)} = \hat{\theta}_T^{(k-1)} + \gamma\left(\hat{\theta}_T^{(k-1)}\right) \mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right) \frac{1}{T} \sum_{t=1}^{T} \frac{\partial l_t\left(\hat{\theta}_T^{(k-1)}\right)}{\partial \theta}$$

where $\mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right)$ is the $q \times q$ symmetric matrix associated with the particular first-derivative algorithm used, and

$$\gamma\left(\hat{\theta}_T^{(k-1)}\right) = -\frac{\left[\sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta' \cdot \mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right) \cdot \sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta\right]}{\begin{bmatrix} \sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta' \cdot \mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right) \cdot \\ T^{-1} \sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta\partial\theta' \cdot \\ \mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right) \cdot \sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta \end{bmatrix}},$$

is the chosen value of the step length.

Let's initially consider the case of $k_{max} = 0$, so that no optimisation whatsoever takes place. Nevertheless, we assume that the initial value $\hat{\theta}_T^{(0)}$ is stochastic, for otherwise, we would simply have a special case of the equality constrained EMM estimator, with the restrictions $\theta = \theta^{(0)}$. If the regularity conditions in Assumption 1 (with $\partial h'(\theta)/\partial\theta = \mathbf{I}_q$) remain valid when (i) $\tilde{\theta}_T$ is replaced by $\hat{\theta}_T^{(0)}$, (ii) $\theta_T^{-}(\rho^0)$ by the limiting pseudo-true value of $\hat{\theta}_T^{(0)}$, $\theta_T^{(0)}(\rho^0)$ say, (iii) $\tilde{\lambda}_T$ by $\tilde{\lambda}_T^{(0)}$, which are the Lagrange multipliers required to satisfy the sample first-order conditions (9) at $\theta = \hat{\theta}_T^{(0)}$, and (iv) $\lambda_T^{-}(\rho^0)$ by the corresponding pseudo-true value, $\lambda_T^{(0)}(\rho^0)$ say, then it follows from the arguments made in section 2.1 that the fully non-optimised EMM estimator of ρ based on $\hat{\theta}_T^{(0)}$ and $\tilde{\lambda}_T^{(0)}$, $\tilde{\rho}_T^{(0)}$ say, will

[1] See e.g. chapter 13 of GM95 for a review of numerical optimization methods.

be consistent and asymptotically normal. Typically, $\hat{\theta}_T^{(0)}$ would be the result of an earlier optimisation procedure, during which some of the parameters were fixed at constant values as part of a step-by-step computational strategy. If that is the case, the previous sentence is just a re-statement of the results in sections 2.1 and 2.3.

Let's now consider the more interesting case of $k_{max} = 1$, but for the sake of brevity, let's concentrate on the Newton-Raphson method, so that $\mathbf{P}\left(\hat{\theta}_T^{(k-1)}\right) = \left[T^{-1}\sum_{t=1}^{T} \partial l_t\left(\hat{\theta}_T^{(k-1)}\right)/\partial\theta\partial\theta'\right]^{-1}$, and consequently, $\nu\left(\hat{\theta}_T^{(k-1)}\right) = 1$. It is then clear that $\hat{\theta}_T^{(1)}$ and $\hat{\mu}_T^{(1)}$ will also be stochastic, with pseudo-true values given by

$$
\begin{aligned}
\theta_T^{(1)}(\rho^0) &= \theta_T^{(0)}(\rho^0) - \mu_T^{(0)}(\rho^0)\mathcal{J}_{0T}^{(0)} \\
\mu_T^{(1)}(\rho^0) &= -E\left\{\left[\partial l_t\left(\theta_T^{(1)}(\rho^0)\right)/\partial\theta\right]\Big|\rho^0\right\}
\end{aligned}
$$

If, mutatis mutandi, the regularity conditions in Assumption 1 remain valid, then the one-step optimised EMM estimator of ρ based on $\hat{\theta}_T^{(1)}$ and $\hat{\mu}_T^{(1)}$, $\hat{\rho}_T^{(1)}$ say, will also be consistent and asymptotically normal. But since the above argument does not really depend on k_{max} being 1, or the way in which $\hat{\theta}_T^{(0)}$ was obtained, it remains valid for any k_{max}.

If k_{max} itself is not fixed a priori, but rather the result of "sampling" variation highly correlated with the impatience of the empirical researcher, then the resulting EMM estimator will still be consistent, but its limiting distribution (in the usual classical sense) will be a mixture of multivariate normals, whose asymptotic variances generally depend on the number of iterations. Of course, in practice the resulting EMM estimator would be numerically identical to the one obtained by another researcher who happened to choose a priori exactly the same number of iterations as her stopping rule. But in any case, the important conclusion from the analysis in this section is that an unsuccessful attempt to optimise the pseudo-log likelihood function can still be successfully used to obtain a consistent EMM estimator of the parameters of interest ρ, as long as the moment conditions used include Lagrange multipliers to reflect the lack of convergence of the algorithm.

For reasons analogous to the ones discussed at the beginning of this section, an empirical researcher may alternatively decide to conduct some specification test in order to assess if there is any evidence in the sample for an additional feature of the data that she has not yet incorporated in her auxiliary model, which merits the optimisation of an even more complex pseudo log-likelihood function. Since most existing specification tests are of the LM form, they can often be written in terms of zero parameter restrictions. Therefore, a numerically sensible strategy could be to base the EMM estimator on the unrestricted estimator of the more complex model if the specification test rejects the null hypothesis, or on the equality restricted version if does not. If the specification test is consistent (in the sense that it rejects the null hypothesis with probability one when the limiting unrestricted pseudo-true value of the relevant parameter is different from zero), then the limiting distribution of the pre-test EMM estimator of ρ is the same as the limiting distribution of the fully optimised unconstrained EMM estimator. In contrast, if the limiting unrestricted pseudo-true value is exactly zero, then the limiting distribution of the pre-test EMM estimator of ρ will be a mixture of the equality restricted estimator, and the unconstrained EMM estimator. But since equality restricted and unconstrained estimators would have the same distribution under the (pseudo) null, then they will all share the same asymptotically normal distribution.

2.5 Simulation-based estimators

For the sake of clarity, we have assumed so far that analytical expressions for (4) and (5) can be readily obtained, as in sections 3.1 and 3.2 below. However, in many other cases, such expressions may be very difficult, or simply impossible to find, and yet they can often be easily obtained by numerical simulation (see e.g. GM96). In particular, we can compute the required expectations as ensemble averages of the levels and derivatives of the Lagrangian function (1) across H realizations of size T of the true process simulated with parameter values equal to ρ. Specifically,

$$\mathcal{L}_T(\rho;\theta) \simeq \mathcal{L}_{HT}(\rho;\theta) = \frac{1}{H}\sum_h \frac{1}{T}\sum_t \frac{l_t(\theta)}{\partial\theta}$$

$$m_T(\rho;\beta) \simeq m_{HT}(\rho;\beta) = \frac{1}{H}\sum_h \frac{1}{T}\sum_t \frac{\partial l_t(\theta)}{\partial\theta} + \frac{\partial h'(\theta)}{\partial\theta}$$

where we can make the last terms arbitrarily close in a numerical sense to the first ones as $H \to \infty$. In those models in which x_t is strictly stationary and ergodic, there is, in fact, an alternative simulation scheme, which computes the required expectations by their sample analogues in a single but very large realization of the process. In particular, we will have:

$$\mathcal{L}(\rho;\theta) \simeq \mathcal{L}_{TH}(\rho;\theta) = \frac{1}{T\cdot H}\sum_{n=1}^{TxH} l_n(\theta)$$

$$m(\rho;\beta) \simeq m_{TH}(\rho;\beta) = \frac{1}{T\cdot H}\sum_{n=1}^{TxH} \frac{\partial l_n(\theta)}{\partial\theta} + \frac{\partial h'(\theta)}{\partial\theta}$$

In this case, we can again make left and right hand sides arbitrarily close in a numerical sense as $H \to \infty$. Similarly, we can approximate the different binding functions $\beta_T(\rho)$ by means of appropriately constrained pseudo ML estimators computed on the basis of a single simulated realization of size $T \times H$ of the true process generated with the parameters of interest set at ρ, or by the average across H simulations of size T of estimators obtained from each simulated sample. From a numerical point of view, the main advantage of EMM estimators is that they avoid the computation of the possibly constrained estimators for each simulation of the process. Finally, note that the autocovariance matrices $S_\tau(\rho;\beta_T)$ used in the computation of the optimal weighting matrix for the continuously updated

EMM and II estimators can also be arbitrarily approximated by replacing the required expected values by their sample counterparts in a long simulation of length $T \cdot H$. Nevertheless, it is important to bear in mind that since H is finite in practice, the asymptotic covariance matrix of the EMM and II estimators in Proposition 4 must be multiplied by the scalar quantity $(1 + H^{-1})$ (see GMR).

3 EXAMPLES

3.1 MA(1) estimated as AR(1)

3.1.1 True and auxiliary models

Consider the following Gaussian ma(1) process:

$$x_t = u_t - \delta u_{t-1}; \quad u_t | x_{t-1}; \ldots \sim N(0; \psi); \quad |\delta| \leq 1; \quad 0 < \psi < \infty \tag{11}$$

where the parameters of interest are $\rho = (\delta; \psi)'$. It is well known that $E(x_t) = 0$, and that its autocovariance structure is given by

$$\begin{aligned} \gamma_0(\rho) &= (1 + \delta^2)\psi \\ \gamma_1(\rho) &= -\delta\psi \\ \gamma_j(\rho) &= 0; \ j > 1. \end{aligned} \tag{12}$$

In order to estimate ρ by II and EMM, we are going to consider initially the following inequality restricted first order autoregression:

$$x_t = \phi x_{t-1} + v_t; \quad v_t | x_{t-1}; \ldots \sim N(0; \omega); \quad \phi \geq 0; \ \omega \geq 0$$

where $\theta = (\phi; \omega)'$. Since the autovariances of an ar(1) process are given by

$$\begin{aligned} Var(x_t) &= \frac{\omega}{1 - \phi^2} \\ cov(x_t; x_{t-1}) &= \phi Var(x_t) \\ cov(x_t; x_{t-j}) &= \phi cov(x_{t-1}; x_{t-j}); \ j > 1 \end{aligned}$$

the non-negativity constraint on ϕ implies that the signs of the first autocorrelations of the true and auxiliary models coincide when $\delta < 0$, and differ when $\delta > 0$. Note, however, that the auxiliary model only nests the true model when $\delta = 0$.

3.1.2 Pseudo-ML estimators

The log-likelihood function of the auxiliary AR(1) model for a sample of size T (ignoring initial conditions) will be given by:

$$L_T(\theta) = \sum_t l_t(\theta) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln \omega - \frac{1}{2\omega}\sum_t (x_t - \phi x_{t-1})^2$$

and the (scaled) Lagrangian function by

$$Q_T(\beta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \omega - \frac{1}{2\omega}\frac{1}{T}\sum_t (x_t - \phi x_{t-1})^2 + \phi\mu_1 + \omega\mu_2$$

where $\mu = (\mu_1, \mu_2)'$ are the multipliers associated with the inequality restrictions $\phi \geq 0$ and $\omega \geq 0$ respectively. Therefore, the sample first-order conditions that take into account the inequality constraints will be given by the Kuhn-Tucker conditions:

$$\begin{aligned} \frac{1}{\hat{\omega}_T}\frac{1}{T}\sum_t (x_t - \hat{\phi}_T x_{t-1})x_{1t-1} + \hat{\mu}_{1T} &= 0 \\ \frac{1}{2\hat{\omega}_T}\frac{1}{T}\sum_t \left[\frac{(x_t - \hat{\phi}_T x_{t-1})^2}{\hat{\omega}_T} - 1\right] + \hat{\mu}_{2T} &= 0 \end{aligned} \qquad (13)$$

together with the sign and complementary slackness constraints:

$$\begin{aligned} \hat{\phi}_T \geq 0 \quad \hat{\mu}_{1T} \geq 0 \quad \hat{\phi}_T \cdot \hat{\mu}_{1T} = 0 \\ \hat{\omega}_T \geq 0 \quad \hat{\mu}_{2T} \geq 0 \quad \hat{\omega}_T \cdot \hat{\mu}_{2T} = 0 \end{aligned}$$

But since

$$\hat{\omega}_T = \frac{1}{T}\sum_t (x_t - \hat{\phi}_T x_{t-1})^2 \geq 0;$$

we can safely take μ_{2T} as 0 in what follows. Also note that since

$$\bar{\mu}_{1T} = -\frac{1}{\bar{\omega}_T}\frac{1}{T}\sum_t (x_t - \bar{\phi}_T x_{t-1})x_{t-1}$$

we can interpret the other multiplier as (minus) the coefficient in the OLS regression of x_{t-1} on the inequality restricted residuals $(x_t - \bar{\phi}_T x_{t-1})$ (see Gourieroux, Holly and Monfort, 1982). Therefore, this Kuhn-Tucker multiplier will be 0 if the inequality restriction is not binding in the sample, or the usual Lagrange multiplier associated with the equality constraint $\phi = 0$ otherwise.

Let $\hat{\phi}_T$; $\hat{\omega}_T$ and $\hat{\mu}_{1T}(=0)$ denote the unrestricted OLS estimators of ϕ; ω and μ_1. Similarly, let $\tilde{\phi}_T(=0)$; $\tilde{\omega}_T$ and $\tilde{\mu}_{1T}$ denote the corresponding equality restricted estimators, and define the sample second moment matrix as follows:

$$\hat{\Sigma}_T = \begin{pmatrix} \sigma_{00T} & \sigma_{01T} \\ \sigma_{01T} & \sigma_{11T} \end{pmatrix} = \frac{1}{T}\sum_t \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} \begin{pmatrix} x_t & x_{t-1} \end{pmatrix}$$

Then we can show that,

$$\begin{array}{lll}
\hat{\phi}_T = \sigma_{01T}/\sigma_{11T} & \tilde{\phi}_T = 0 & \bar{\phi}_T = I(\sigma_{01T} \geq 0)\sigma_{01T}/\sigma_{11T} \\
\hat{\omega}_T = \sigma_{00T} - \sigma_{01T}^2/\sigma_{11T} & \tilde{\omega}_T = \sigma_{00T} & \bar{\omega}_T = \sigma_{00} - I(\sigma_{01T} \geq 0)\sigma_{01T}^2/\sigma_{11T} \\
\hat{\mu}_{1T} = 0 & \tilde{\mu}_{1T} = -\sigma_{01}/\sigma_{00} & \bar{\mu}_{1T} = -I(\sigma_{01T} \leq 0)\sigma_{01T}/\sigma_{00T}
\end{array} \quad (14)$$

where $I(.)$ is the usual indicator function. Therefore, the inequality restricted OLS estimators of ϕ and ω take two different forms depending on whether the sign of σ_{01T} (and $\hat{\phi}_T$) is positive or negative.

3.1.3 Population moments and binding functions

In view of the discussion in section 2, we can base the different EMM estimators of ρ on the following population moments

$$m_{1T}(\rho;\beta) = E\left[\frac{1}{\omega}\frac{1}{T}\sum_t (x_t - \phi x_{t-1})x_{1t-1} + \mu_1 \right|_{\beta} \Bigg]_{\rho}$$

$$m_{2T}(\rho;\beta) = E\left[\frac{1}{2\omega}\frac{1}{T}\sum_t \left(\frac{(x_t-\phi x_{t-1})^2}{\omega}-1\right)+\mu_2 \,\Big|\, \rho\right]$$

which, due to the covariance stationarity of the true model, reduce to the following time-invariant expressions

$$\begin{aligned} m_1(\rho;\beta) &= \frac{1}{\omega}\left[\gamma_1(\rho)-\phi\gamma_0(\rho)\right]+\mu_1 \\ m_2(\rho;\beta) &= \frac{1}{2\omega}\left(\frac{-2\phi\gamma_1(\rho)+(1+\phi^2)\gamma_0(\rho)}{\omega}-1\right)+\mu_2 \end{aligned} \qquad (15)$$

where the dependence of γ_j on ρ comes from (12).

If we define $\theta^i(\rho)$ and $\mu^i(\rho)$ as the values of the parameters and multipliers of the auxiliary model that for each value of ρ solve the population program

$$\max_{\theta} \mathcal{L}_T(\rho;\theta) \quad s.t. \quad \phi \geq 0;\ \omega \geq 0$$

where

$$\mathcal{L}_T(\rho;\theta) = E\left[l_t(\theta)|\rho\right] = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\omega - \frac{(1+\phi^2)\gamma_0(\rho)-2\phi\gamma_1(\rho)}{2\omega},$$

it is clear that the inequality restricted binding functions $\beta^i(\rho)$ satisfy the moment conditions

$$m\left[\rho;\beta^i(\rho)\right] = \mathbf{0}$$

together with the sign and exclusion restrictions

$$\begin{aligned} \phi^i(\rho) &\geq 0; \quad \mu_1^i(\rho)\geq 0; \quad \phi^i(\rho)\cdot\mu_1^i(\rho) = 0 \\ \omega^i(\rho) &\geq 0; \quad \mu_2^i(\rho)\geq 0; \quad \omega^i(\rho)\cdot\mu_2^i(\rho) = 0 \end{aligned}$$

From here, it is easy to see that

$$\omega^i(\rho) = E\left\{\left[x_t-\phi^i(\rho)x_{t-1}\right]^2\Big|\rho\right\} = \left\{1+\left[\phi^i(\rho)\right]^2\right\}\gamma_0(\rho)-2\phi^i(\rho)\gamma_1(\rho)\geq 0$$

so that $\mu_2^i(\rho) = 0$, as expected. As for the other elements, in principle there may be two different situations depending on whether or not $\delta \leq 0$. Specifically:

$$
\begin{aligned}
\psi^u(\rho) &= \frac{\gamma_1(\rho)}{\gamma_0(\rho)} & \psi^e(\rho) &= 0 & \psi^i(\rho) &= I(\delta \geq 0)\frac{\gamma_1(\rho)}{\gamma_0(\rho)} \\
\omega^u(\rho) &= \gamma_0(\rho) - \frac{\gamma_1^2(\rho)}{\gamma_0(\rho)} & \omega^e(\rho) &= \gamma_0(\rho) & \omega^i(\rho) &= \gamma_0(\rho) - I(\delta \geq 0)\frac{\gamma_1^2(\rho)}{\gamma_0(\rho)} \\
\mu_1^u(\rho) &= 0 & \mu_1^e(\rho) &= -\frac{\gamma_1(\rho)}{\gamma_0(\rho)} & \mu_1^i(\rho) &= -I(\delta \leq 0)\frac{\gamma_1(\rho)}{\gamma_0(\rho)}
\end{aligned}
$$

where $\beta^u(\rho)$ denotes the usual unrestricted binding functions, and $\beta^e(\rho)$ the equality restricted ones associated with the constraint $\psi = 0$. Obviously, they all coincide for $\delta = 0$, in which case

$$
\begin{aligned}
\psi^u(0;\chi) &= \psi^e(0;\chi) = \psi^i(0;\chi) = 0 (= \delta) \\
\omega^u(0;\chi) &= \omega^e(0;\chi) = \omega^i(0;\chi) = \chi \\
\mu_1^u(0;\chi) &= \mu_1^e(0;\chi) = \mu_1^i(0;\chi) = 0
\end{aligned}
$$

Figure 1 plots the binding functions $\psi^u(\rho)$ and $\mu_1^e(\rho)$ for $-1 \leq \delta \leq 1$. Note that in this framework, $\psi^i(\rho) = \max[\psi^u(\rho); 0]$ and $\mu_1^i(\rho) = \max[\mu_1^e(\rho); 0]$.

3.1.4 Asymptotic distributions of pseudo-ML estimators and sample moments

Given the different expressions for the inequality restricted pseudo-ML estimators of θ and μ discussed previously, the sample counterparts to (15) will be given by either:

$$
\begin{aligned}
m_1(\rho; \hat{\beta}_T) &= \frac{[\gamma_1(\rho) - (\sigma_{01T}/\sigma_{11T})\gamma_0(\rho)]}{\sigma_{00T} - \sigma_{01T}^2/\sigma_{11T}} \\
m_2(\rho; \hat{\beta}_T) &= \frac{\left[-2(\sigma_{01T}/\sigma_{11T})\gamma_1(\rho) + (1 + \sigma_{01T}^2/\sigma_{11T}^2)\gamma_0(\rho) - \left(\sigma_{00T} - \sigma_{01T}^2/\sigma_{11T}\right)\right]}{2\left(\sigma_{00T} - \sigma_{01T}^2/\sigma_{11T}\right)^2}
\end{aligned}
$$

when $\sigma_{01T} \geq 0$, or

$$
m_1(\rho; \bar{\beta}_T) = \frac{[\gamma_1(\rho) - \sigma_{01T}]}{\sigma_{00T}}
$$

$$m_2(\rho; \bar{\beta}_T) = \frac{[\gamma_0(\rho) - \sigma_{00T}]}{2\sigma^2_{00T}}$$

when $\sigma_{01T} \leq 0$. In this respect, note that $m(\rho; \hat{\beta}_T)$ are precisely the sample moments that we would use in a standard unrestricted EMM procedure, while $m(\rho; \bar{\beta}_T)$ are the ones that correspond to the equality constrained EMM procedure based on the constraint $\varphi = 0$.

Let's now derive the asymptotic distribution of the pseudo-ML estimators of the auxiliary parameters, multipliers and moments in the three different relevant situations that may occur: (i) $\delta^0 < 0$, (ii) $\delta^0 > 0$, and (iii) $\delta^0 = 0$. To do so, we shall use the following lemma, which can be proved as a straightforward application of Theorem 5.7.1 in Anderson (1971):

Lemma 1 When x_t is given by the Gaussian ma(1) model (11), the first sample autocorrelation $\hat{\varphi}_T$ is $T^{1/2}$-consistent for the first population autocorrelation $\varphi^u(\rho^0)$, with the following limiting distribution

$$\sqrt{T}\left[\hat{\varphi}_T - \varphi^u(\rho^0)\right] \stackrel{d}{\to} N\left\{0; \frac{1 + (\delta^0)^2 + 4(\delta^0)^4 + (\delta^0)^6 + (\delta^0)^8}{\left[1 + (\delta^0)^2\right]^4}\right\}$$

Note that the asymptotic variance of $\hat{\varphi}_T$, which not surprisingly is the same for a non-invertible ma(1) process with parameter $1/\delta$, achieves its maximum (=1) for $\delta^0 = 0$ and its minimum (=1/2) for $\delta^0 = \pm 1$. In addition, it is easy to see that $\sqrt{T}\left(\hat{\varphi}_T + \mu_{1T}\right) = o_p(1)$ because $\sigma_{00} - \sigma_{11} = (x_T^2 - x_0^2)/T = O_p(T^{-1})$. As a result, we will have that

$$\lim_{T\to\infty} P(\sqrt{T}\hat{\varphi}_T > 0) = \lim_{T\to\infty} P(\sqrt{T}\mu_T < 0) = \begin{cases} 1 & \text{if } \delta^0 < 0 \\ 1/2 & \text{if } \delta^0 = 0 \\ 0 & \text{if } \delta^0 > 0 \end{cases}$$

Hence, when $\delta^0 < 0$, $\sqrt{T}\left(\tilde{\varphi}_T - \hat{\varphi}_T\right)$ and $\sqrt{T}\mu_{1T}$ are both $o_p(1)$, and the inequality restricted EMM and II estimators of ρ are asymptotically equivalent to the usual unrestricted EMM and II estimators. In contrast, when $\delta^0 > 0$, $\sqrt{T}\tilde{\phi}_T$ and $\sqrt{T}(\tilde{\mu}_{1T} - \mu_{1T})$ are $o_p(1)$, and the inequality restricted EMM and II estimators of ρ will then coincide in large samples with the equality restricted ones. The most interesting situation arises when $\delta^0 = 0$. In this case, $\tilde{\beta}_T$ has a non-normal asymptotic distribution, as it will be equal to either $(\hat{\phi}_T; \hat{\omega}_T; 0)'$ or $(0; \tilde{\omega}_T; \mu_{1T})'$ with probability approximately one half each. As a consequence, the sample moment conditions will also be $m(\rho; \hat{\beta}_T)$ fifty per cent of the time, and $m(\rho; \bar{\beta}_T)$ the other fifty. Nevertheless, given that when $\delta^0 = 0$ we can write

$$\sqrt{T}m_1(\rho^0; \hat{\beta}_T) = -\frac{\gamma_0(\rho^0)}{\sigma_{00} - \sigma_{01}^2/\sigma_{11}}\sqrt{T}\hat{\phi}_T$$

and

$$\sqrt{T}m_1(\rho^0; \bar{\beta}_T) = -\sqrt{T}\mu_{1T}$$

it is clear that $\sqrt{T}\left[m_1(\rho^0; \hat{\beta}_T) - m_1(\rho^0; \bar{\beta}_T)\right] = o_p(1)$, so that the limiting distribution of $\sqrt{T}m_1(\rho^0; \tilde{\beta}_T)$ will also be normal, with an analogous result for the other moment. The reason is that despite the fact that both $\sqrt{T}\tilde{\phi}_T$ and $\sqrt{T}\tilde{\mu}_{1T}$ have half normal distributions, asymptotically $\sqrt{T}(\tilde{\phi}_T - \tilde{\mu}_{1T})$ has the same $N(0; 1)$ distribution as either $\sqrt{T}(\hat{\phi}_T - \hat{\mu}_{1T}) = \sqrt{T}\hat{\phi}_T$ or $\sqrt{T}(\mathring{\phi}_T - \mu_{1T}) = -\sqrt{T}\mu_{1T}$. In fact, this last statement is true irrespectively of $\delta^0 \gtrless 0$, and simply constitutes an example of Proposition 2. As for Proposition 1, we trivially have that

$$\tilde{\mu}_{1T}\sqrt{T}\left[\tilde{\phi}_T - \tilde{\phi}_T^i(\rho^0)\right] + \tilde{\phi}_T\sqrt{T}\left[\tilde{\mu}_{1T} - \mu_1^i(\rho^0)\right] = 0$$

and the same applies to the unrestricted and equality restricted pseudo-ML estimators and multipliers.

3.1.5 Indirect inference estimators

If the parameters of interest of the true model were $\gamma = (\gamma_0; \gamma_1)'$ rather than ρ, the solution of the linear system of equations $m[\tilde{\gamma}_T; \tilde{\beta}_T] = 0$ with respect to $\tilde{\gamma}_T$

would give us the inequality restricted EMM estimator of these autocovariances. More explicitly, since the system above could be re-written as

$$\begin{pmatrix} -\tilde{\psi}_T & 1 \\ 1+\tilde{\psi}_T^2 & -2\tilde{\psi}_T \end{pmatrix}\begin{pmatrix} \tilde{\gamma}_{0T} \\ \tilde{\gamma}_{1T} \end{pmatrix} = \begin{pmatrix} -\tilde{\omega}_T \tilde{\xi}_{1T} \\ \tilde{\omega}_T \end{pmatrix} \tag{16}$$

we would have that the inequality constrained EMM estimators of γ would be given by

$$\begin{pmatrix} \tilde{\gamma}_{0T} \\ \tilde{\gamma}_{1T} \end{pmatrix} = I(\hat{\psi}_T \geq 0)\begin{pmatrix} \hat{\gamma}_{0T} \\ \hat{\gamma}_{1T} \end{pmatrix} + I(\hat{\psi}_T < 0)\begin{pmatrix} \bar{\gamma}_{0T} \\ \bar{\gamma}_{1T} \end{pmatrix}$$

where

$$\begin{pmatrix} \hat{\gamma}_{0T} \\ \hat{\gamma}_{1T} \end{pmatrix} = \begin{pmatrix} \sigma_{00T} \\ \sigma_{01T} \end{pmatrix} \cdot \left(\frac{\sigma_{00T}^2 - \sigma_{01T}^2}{\sigma_{11T}^2 - \sigma_{01T}^2} \right) \tag{17}$$

are the EMM estimators of γ that use as score generator an unrestricted AR(1) model, and

$$\begin{pmatrix} \bar{\gamma}_{0T} \\ \bar{\gamma}_{1T} \end{pmatrix} = \begin{pmatrix} \sigma_{00T} \\ \sigma_{01T} \end{pmatrix}$$

the ones based on a white noise process, provided that in the latter case we include in the set of moments the Lagrange first order condition of the autoregressive parameter with the corresponding multiplier.[2] But given that $\sigma_{00T} - \sigma_{11T} = O_p(T^{-1})$ for any value of δ^0, it is easy to see that the EMM estimator of γ based on an inequality restricted AR(1) process, $\tilde{\gamma}_T$, is always asymptotically equivalent to both $\hat{\gamma}_T$ and $\bar{\gamma}_T$.

This result is not totally surprising if we note that the two sets of sample moments satisfy the following relationships:

$$\begin{aligned} \left(\sigma_{00T} - \frac{\sigma_{01T}^2}{\sigma_{11T}}\right) m_1(\rho; \hat{\ }_T) &= \sigma_{00T}\, m_1(\rho; \bar{\ }_T) - \sigma_{00T}^2 \frac{\sigma_{01T}}{\sigma_{11T}} m_2(\rho; \bar{\ }) \\ &\quad + \frac{\sigma_{01T}}{\sigma_{11T}}(\sigma_{11T} - \sigma_{00T}) \end{aligned}$$

[2] Note that the implied estimate of the first autocorrelation is the same in both cases.

$$\left(\sigma_{00T} - \frac{\sigma_{01T}^2}{\sigma_{11T}}\right)^2 m_2(\rho;\hat\beta_T) = -\sigma_{00T}\frac{\sigma_{01T}}{\sigma_{11T}} m_1(\rho;\bar\beta) + \sigma_{00T}^2\left(1+\frac{\sigma_{01T}^2}{\sigma_{11T}^2}\right) m_2(\rho;\bar\beta) + \frac{\sigma_{01T}^2}{\sigma_{11T}^2}(\sigma_{11T}-\sigma_{00T})$$

Hence, $\sqrt{T}m(\rho;\hat\beta_T)$ and $\sqrt{T}m(\rho;\bar\beta_T)$ are almost an exact linear combination of each other for large T irrespectively of ρ^0.

On the other hand, the unconstrained II estimators of γ_0 and γ_1 would be obtained by minimising the following MD criterion function:

$$D_T^u(\gamma;\mathbf{I}_2) = \left(\hat\chi_T - \frac{\gamma_1}{\gamma_0}\right)^2 + \left[\hat\psi_T - \left(\gamma_0 - \frac{\gamma_1^2}{\gamma_0}\right)\right]^2$$

while the equality constrained II estimators would minimise

$$D_T^e(\gamma;\mathbf{I}_2) = \left(\mu_{1T} + \frac{\gamma_1}{\gamma_0}\right)^2 + (\omega_T - \gamma_0)^2$$

instead. But in view of the expressions for $\hat\chi_T$, $\hat\psi_T$, μ_{1T} and ω_T in (14), it is obvious that such II estimators will numerically coincide with $\hat\gamma_T$ and $\tilde\gamma_T$ respectively. Moreover, since the inequality constrained II estimator would minimise the objective function

$$D_T^i(\gamma;\mathbf{I}_3) = \left[\acute\chi_T - \frac{\gamma_1}{\gamma_0}I(\gamma_1 \geq 0)\right]^2 + \left[\mu_{1T} + \frac{\gamma_1}{\gamma_0}I(\gamma_1 \leq 0)\right]^2 + \left\{\tilde\omega_T - \left[\gamma_0 - \frac{\gamma_1^2}{\gamma_0}I(\gamma_1 \geq 0)\right]\right\}^2$$

it is clear that it will be given by $\check\gamma_T$, as expected. The reason is that since the auxiliary model exactly identifies the first two autocovariances, and there are no binding constraints of γ, then II and EMM yield the same estimators.

The common asymptotic distribution of $\tilde\gamma_T; \hat\gamma_T$ and $\check\gamma_T$ can be directly obtained as a special case of Theorem 8.4.2 in Anderson (1971):

Lemma 2 When x_t is given by the Gaussian ma(1) model (11), $\hat\nu_{0T}$ and $\hat\nu_{1T}$ are $T^{1/2}$-consistent for $\gamma_0(\rho^0)$ and $\gamma_1(\rho^0)$ respectively, with the following limiting distribution

$$\sqrt{T}\left[\hat{\gamma}_T - \gamma(\rho^0)\right] \xrightarrow{d} N\left[0; \mathbf{V}(\rho^0)\right]$$

where

$$\mathbf{V}(\rho) = \psi^2 \begin{pmatrix} 2 + 8\delta^2 + 2\delta^4 & -4\delta - 4\delta^3 \\ -4\delta - 4\delta^3 & 1 + 5\delta^2 + \delta^4 \end{pmatrix}$$

But even though γ are not really the parameters of interest, we can regard their EMM estimators as "sufficient statistics" from which we can estimate ρ. At first sight, it may seem that we could recover ρ by solving numerically the nonlinear system of equations (12). Unfortunately, there is no solution if $\left|\hat{\chi}_T\right| > .5$. One attractive possibility involves the minimisation of the optimal (continuously updated) MD criterion:

$$\begin{pmatrix} \hat{\gamma}_{0T} - \gamma_0(\rho) & \hat{\gamma}_{1T} - \gamma_1(\rho) \end{pmatrix} \mathbf{V}^{-1}(\rho) \begin{pmatrix} \hat{\gamma}_{0T} - \gamma_0(\rho) \\ \hat{\gamma}_{1T} - \gamma_1(\rho) \end{pmatrix}$$

subject to the inequality constraints $-1 \leq \delta \leq 1$ and $\psi \geq 0$. Tedious but otherwise straightforward algebra shows that the resulting estimators of δ and ψ will be given by the following expressions:

$$\left.\begin{array}{l} \tilde{\delta}_T = 0 \\ \tilde{\psi}_T = \hat{\sigma}_{00} \end{array}\right\} \quad \text{if } \hat{\chi}_T = 0$$

$$\left.\begin{array}{l} \tilde{\delta}_T = \left(-1 + \sqrt{1 - 4\hat{\chi}_T^2}\right) / (2\hat{\chi}_T) \\ \tilde{\psi}_T = \hat{\sigma}_{00} / (1 + \tilde{\delta}_T^2) \end{array}\right\} \quad \text{if } 0 < \hat{\chi}_T^2 \leq .25 \tag{18}$$

$$\left.\begin{array}{l} \tilde{\delta}_T = -\text{sign}(\hat{\chi}_T) \\ \tilde{\psi}_T = \hat{\sigma}_{00}\left(7 + 12\hat{\chi}_T^2 - 16\left|\hat{\chi}_T\right|\right) / \left(6 - 4\left|\hat{\chi}_T\right|\right) \end{array}\right\} \quad \text{if } \hat{\chi}_T^2 > .25$$

In fact, given that the above MD criterion would numerically coincide with the optimal (continuously updated) GMM criterion based on the restrictions

$$E[x_t^2 - \gamma_0(\rho)] = 0$$
$$E[x_t x_{t-1} - \gamma_1(\rho)] = 0$$

if $\hat{\sigma}_{00T} = \hat{\sigma}_{11T}$, and that the estimating equations used in (un)restricted EMM and II procedures would be a linear combination of these ones, it is clear that the different estimators of ρ are asymptotically equivalent.[3]

An analogous line of reasoning applies to pretest EMM and II estimators that use either the equality restricted estimators when a standard LM test for first order serial correlation does not reject the null of white noise, or the unrestricted estimators when it does. Since as we have just seen, $\tilde{\rho}_T$ and $\hat{\rho}_T$ have the same asymptotic distribution regardless of the value of δ^0, such a common distribution will be inherited by the pretest estimators.

Finally, note that since the auxiliary model "smoothly embeds" the true model when $\delta^0 = 0$, Theorem 2 in GT implies that in this particular case, the unrestricted estimator $\hat{\rho}_T$ is asymptotically equivalent to maximum likelihood, and the same obviously applies to all the other estimators. However, the asymptotic efficiency of $\tilde{\rho}_T$ relative to the ML estimator decreases as $|\delta^0|$ increases. In particular, the asymptotic distribution of $\sqrt{T}(\hat{\delta}_T - \delta^0)$ when $\delta^0 = 1$ is half normal by virtue of Lemma 1 and expression (18), while the ML estimator of δ is superconsistent (i.e. consistent at the rate T; see Sargan and Bhargava, 1983).

In principle, it may seem that the imposition of the correct restriction $|\phi| \leq .5$ in the estimation of the auxiliary $ar(1)$ model should produce more efficient estimators of the parameters of interest. However, it turns out that exactly the same II estimator of ρ is obtained when we replace the non-negativity restriction on ϕ by a general restriction of the form $\phi_{min} \leq \phi \leq \phi_{max}$, for any ϕ_{min}, ϕ_{max}. Moreover, the equivalence between the different EMM and II estimators of ρ in the ma(1) via $ar(1)$ example does not really depend on the nature of the true

[3]Nevertheless, when $|\hat{\phi}_T| > .5$, the different estimators of ψ will differ in finite samples, not only because $\hat{\sigma}_{00T} - \hat{\sigma}_{11T}$ is only approximately zero, but also because unless one uses the analytical expression for $V(\rho)$ above, there will be estimation error in the HAC calculation of the optimal weighting matrices.

model, whose parameters only enter through $\gamma_0(\rho)$ and $\gamma_1(\rho)$, but rather on the particular form of the auxiliary model used. As we mentioned above, the reason is that from the point of II and EMM estimation, $\hat{\gamma}_T; \hat{\delta}_T$ and $\hat{\psi}_T$ play the role of "sufficient statistics" from which we infer ρ. In this respect, it is possible to prove that the same result is true whenever the auxiliary model is given by a conditionally homoskedastic Gaussian AR(p) process, with p finite, and the restrictions are linear in the autoregressive parameters.

3.2 AR(1) estimated as MA(1)

3.2.1 True and auxiliary models

Consider now the following stationary AR(1) process:

$$x_t = \phi x_{t-1} + v_t; \quad v_t|x_{t-1}; \ldots \sim N(0; \omega); \quad |\phi| < 1; 0 < \omega < \infty \tag{19}$$

where the parameters of interest are $\rho = (\phi; \omega)'$. It is well known that $E(x_t) = 0$, and that its autocovariance structure is given by

$$\gamma_j(\rho) = \phi^j \frac{\omega}{1-\phi^2}; \; j \geq 0 \tag{20}$$

In order to estimate ρ by indirect inference, we are going to use initially the following inequality restricted MA(1) model:

$$x_t = u_t - \delta u_{t-1}; \quad u_t|x_{t-1}; \ldots \sim N(0; \psi); \quad \delta \leq 0; \psi \geq 0$$

where $\theta = (\delta; \psi)'$. Since its autocovariance structure is given by:

$$\begin{aligned} Var(x_t) &= (1+\delta^2)\psi \\ cov(x_t; x_{t-1}) &= -\delta\psi \\ cov(x_t; x_{t-j}) &= 0; \; j > 1 \end{aligned}$$

the non-positivity constraint on δ implies that the signs of the first autocorrelations of the auxiliary and true models coincide when $\phi > 0$, and differ when $\phi < 0$. Note, however, that the auxiliary model only nests the true model when $\phi = 0$.

3.2.2 Pseudo-ML estimators

The log-likelihood function of the MA(1) model for a sample of size T will be given by:

$$L_T(\theta) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\psi - \frac{1}{2\psi}\sum_t [x_t - \gamma_t(\delta)]^2$$

with

$$\gamma_t(\delta) = -\sum_{j=1}^{\infty} \delta^j x_{t-j};$$

and the (scaled) Lagrangian function by

$$Q_T(\beta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\psi - \frac{1}{2\psi}\frac{1}{T}\sum_t [x_t - \gamma_t(\delta)]^2 + \delta\mu_1 + \psi\mu_2$$

where $\mu = (\mu_1; \mu_2)'$ are the multipliers associated with the inequality restrictions $\delta \leq 0$ and $\psi \geq 0$ respectively. Therefore, the first-order conditions that take into account the inequality constraints will be given by the Kuhn-Tucker conditions:

$$\begin{aligned}
\frac{1}{\hat{\psi}_T}\frac{1}{T}\sum_t u_t(\hat{\delta}_T)\frac{\partial\gamma_t(\hat{\delta}_T)}{\partial\delta} + \mu_{1T} &= 0 \\
\frac{1}{2\hat{\psi}_T}\frac{1}{T}\sum_t \left[\frac{u_t^2(\hat{\delta}_T)}{\hat{\psi}_T} - 1\right] + \mu_{2T} &= 0
\end{aligned}$$

where

$$u_t(\delta) = \sum_{j=0}^{\infty} \delta^j x_{t-j} \tag{21}$$

$$\frac{\partial\gamma_t(\delta)}{\partial\delta} = -\sum_{j=1}^{\infty} j\delta^{j-1}x_{t-j} \tag{22}$$

together with sign and exclusion constraints

$$\begin{aligned} \tilde{\delta}_T &\leq 0;\ \tilde{\mu}_{1T} \geq 0;\ \tilde{\delta}_T \cdot \tilde{\mu}_{1T} = 0 \\ \tilde{\psi}_T &\geq 0;\ \tilde{\mu}_{2T} \geq 0;\ \tilde{\psi}_T \cdot \tilde{\mu}_{2T} = 0 \end{aligned}$$

But as

$$\tilde{\psi}_T = \frac{1}{T}\sum_t u_t^2(\tilde{\delta}_T) \geq 0$$

we can safely take $\tilde{\mu}_{2T} = 0$ in what follows. Also since

$$\tilde{\mu}_{1T} = -\frac{1}{\tilde{\psi}_T}\frac{1}{T}\sum_t u_t(\tilde{\delta}_T)\frac{\partial \xi_t(\tilde{\delta}_T)}{\partial \delta}$$

we can interpret this Kuhn-Tucker multiplier as (minus) the coefficient in the OLS regression of $\partial \xi_t(\tilde{\delta}_T)/\partial \delta$ on the inequality restricted residuals $u_t(\tilde{\delta}_T)$ (see Gourieroux, Holly and Monfort, 1980). Therefore, $\tilde{\mu}_{1T}$ will be 0 if the inequality restriction is satisfied, or the usual Lagrange multiplier associated with the equality constraint $\delta = 0$ otherwise. Not surprisingly, the Lagrange multiplier is simply

$$\hat{\mu}_{1T} = -\frac{T^{-1}\sum x_t x_{t-1}}{T^{-1}\sum_t x_t^2} = -\frac{\hat{\gamma}_{10T}}{\hat{\gamma}_{00T}}$$

which, as in the previous example, is approximately the same as the (opposite of the) first sample autocorrelation in large samples. Similarly,

$$\hat{\psi}_T = \frac{1}{T}\sum_t x_t^2 = \hat{\gamma}_{00T}$$

i.e. the sample variance with denominator T.

3.2.3 Population moments and binding functions

Given the covariance stationarity of the true model, we can base our estimation of ρ on the following time-invariant expressions

$$\begin{aligned} m_1(\rho;\beta) &= E\left[\frac{1}{\psi}u_t(\delta)\frac{\partial \xi_t(\delta)}{\partial \delta} + \mu_1\right]\Bigg|_{\rho} \\ m_2(\rho;\beta) &= E\left\{\frac{1}{2\psi}\left[\frac{u_t^2(\delta)}{\psi} - 1\right] + \mu_2\right\}\Bigg|_{\rho} \end{aligned} \tag{23}$$

which using the results in the appendix, can be written as

$$m_1(\rho;\beta) = -\frac{\gamma_0(\rho)}{\psi(1-\delta^2)^2}\left\{\delta + \sum_{l=1}^{\infty}\left[2\delta^2 + \left(1-\delta^2\right)l\right]\delta^{l-1}\frac{\gamma_l(\rho)}{\gamma_0(\rho)}\right\} + \mu_1$$

$$= \left(\frac{\omega}{\psi(1-\phi^2)\left(1-\delta^2\right)^2(1-\delta\phi)^2}\right)\left(\phi^2\delta^3 + \phi\delta^2 - \delta - \phi\right) + \mu_1$$

$$m_2(\rho;\beta) = \frac{1}{2\psi^2}\left[\frac{\gamma_0(\rho)}{1-\delta^2}\left(1 + 2\sum_{l=1}^{\infty}\delta^l\frac{\gamma_l(\rho)}{\gamma_0(\rho)}\right) - \psi\right] + \mu_2$$

$$= \frac{1}{2\psi^2}\left[\frac{\omega}{\left(1-\delta^2\right)\left(1-\phi^2\right)}\cdot\frac{1-\delta\phi}{1+\delta\phi} - \psi\right] + \mu_2$$

where the intermediate expressions only depend on the auxiliary model, while the final expressions are obtained by replacing (20) in the intermediate ones.

If we define $\theta^{-}(\rho)$ and $\mu^{-}(\rho)$ as the values of the parameters and multipliers of the auxiliary model that for each value of ρ solve the population program

$$\max_{\theta}\mathcal{L}_T(\rho;\theta) \quad \text{s.t.} \quad \delta \leq 0; \psi \geq 0$$

where

$$\mathcal{L}_T(\rho;\theta) = E\left[l_t(\theta)|\rho\right] = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\psi - \frac{1}{2\psi}E\left\{u_t^2(\delta)\middle|\rho\right\}$$

it is clear that the inequality restricted binding functions $\beta^{-}(\rho)$ satisfy the moment conditions

$$\mathbf{m}\left[\rho;\beta^{-}(\rho)\right] = \mathbf{0}$$

together with the sign and exclusion restrictions

$$\delta^{-}(\rho) \leq 0; \quad \mu_1^{-}(\rho) \geq 0; \quad \delta^{-}(\rho)\cdot\mu_1^{-}(\rho) = 0$$

$$\psi^{-}(\rho) \geq 0; \quad \mu_2^{-}(\rho) \geq 0; \quad \psi^{-}(\rho)\cdot\mu_2^{-}(\rho) = 0$$

From here, it is easy to see that

$$\psi^{-}(\rho) = E\left\{u_t^2\left[\delta^{-}(\rho)\right]\middle|\rho\right\} = \frac{\gamma_0(\rho)}{1-\left[\delta^{-}(\rho)\right]^2}\left(1 + 2\sum_{l=1}^{\infty}\left[\delta^{-}(\rho)\right]^l\frac{\gamma_l(\rho)}{\gamma_0(\rho)}\right) \geq 0$$

and consequently, that $\mu_2^i(\rho) = 0$, as expected.

>From the above moment expressions, we also have that the usual unconstrained binding function for δ, $\delta^u(\rho)$ will be the real root of the following third order equation

$$\phi^2 [\delta^u(\rho)]^3 + \phi [\delta^u(\rho)]^2 - \delta^u(\rho) - \phi = 0$$

whose modulus is less than or equal to 1. [4]

As a result, if $\delta^u(\rho) \leq 0$, then $\beta^i(\rho) = \beta^u(\rho)$, where

$$\begin{aligned} \psi^u(\rho) &= \frac{\omega'}{\left(1 - [\delta^u(\rho)]^2\right)\left(1 - \phi^2\right)} \cdot \frac{1 - \delta^u(\rho)\phi}{1 + \delta^u(\rho)\phi'} \\ \mu^u(\rho) &= 0 \end{aligned}$$

are the remaining unconstrained binding functions, while if $\delta^u(\rho) > 0$, then $\beta^i(\rho) = \beta^e(\rho)$, where

$$\begin{aligned} \delta^e(\rho) &= 0 \\ \psi^e(\rho) &= \gamma_0(\rho) = \frac{\omega}{1 - \phi^2} \\ \mu_1^e(\rho) &= -\frac{\gamma_1(\rho)}{\gamma_0(\rho)} = -\phi \geq 0 \end{aligned} \quad (24)$$

are the binding functions associated with the equality constraint $\delta = 0$. Since the first theoretical autocorrelation has the same sign as ϕ, the first solution applies when $\phi \geq 0$, while the second solution when $\phi \leq 0$. Obviously, they all coincide when $\phi = 0$, in which case

$$\begin{aligned} \delta^u(0; \omega) &= \delta^e(0; \omega) = \delta^i(0; \omega) = 0 (= \phi) \\ \psi^u(0; \omega) &= \psi^e(0; \omega) = \psi^i(0; \omega) = \omega \\ \mu_1^u(0; \omega) &= \mu_1^e(0; \omega) = \mu_1^i(0; \omega) = 0 \end{aligned}$$

[4] It is important to mention that $\delta^u(\rho)$ is different from the first inverse autocorrelation of the AR(1) model, which is given by $\phi/(1 + \phi^2)$, since the range of $\delta^u(\rho)$ is -1 to 1, rather than -1/2 to 1/2 (see e.g. Bhansali, 1980).

Figure 2 plots the binding functions $\delta^u(\rho)$ and $\mu_1^e(\rho)$ for $-1 < \phi < 1$. Note that in this framework, $\delta^i(\rho) = \min[\delta^u(\rho); 0]$ while $\mu_1^i(\rho) = \max[\mu_1^e(\rho); 0]$.

3.2.4 Asymptotic distributions of pseudo-ML estimators and sample moments

First of all, let's state the AR(1) version of Lemma 2 above, which can again be obtained from theorem 8.4.2 in Anderson (1971):

Lemma 3 When x_t is given by the Gaussian AR(1) model (19), σ_{00T} and σ_{01T} are $T^{1/2}$-consistent for $\gamma_0(\rho^0)$ and $\gamma_1(\rho^0)$ in (20) respectively, with the following limiting distribution

$$\sqrt{T}\left[\hat{\gamma}_T - \gamma(\rho^0)\right] \xrightarrow{d} N\left[0; \mathbf{V}(\rho^0)\right]$$

where

$$\mathbf{V}(\rho) = \frac{\omega^2}{(1-\phi^2)^3}\begin{pmatrix} 2+2\phi^2 & 4\phi \\ & 1+4\phi^2-\phi^4 \end{pmatrix}$$

Given that the population moments evaluated at the equality restricted pseudo-ML estimators are given by:

$$\begin{aligned} m_1(\rho;\beta) &= -\frac{\omega}{\sigma_{00T}(1-\phi^2)}\phi - \frac{\sigma_{10T}}{\sigma_{00T}} \\ m_2(\rho;\beta) &= \frac{1}{2\sigma_{00T}}\left[\frac{\omega}{(1-\phi^2)} - \sigma_{00T}\right] \end{aligned}$$

it is straightforward to derive their asymptotic distribution by means of the delta method. Similarly, we can use the same technique to derive the asymptotic distribution of $\mu_{1T} = -\sigma_{10T}/\sigma_{00T}$ and $\hat{\phi}_T = \sigma_{00T}$. Alternatively, the asymptotic distribution of the estimator of the Lagrange multiplier can be directly obtained from the Mann and Wald theorem.

In contrast, the asymptotic distribution of the unrestricted estimators $\hat{\delta}_T$ and $\hat{\psi}_T$ is rather more laborious to obtain, as we need to derive closed form expressions for the matrices $\mathcal{I}^u_{0T}$ and $\mathcal{J}^u_{0T}$. For simplicity, we shall only do it for the case of $\phi^0 = 0$, which as we saw before, corresponds to $\delta^u(\rho^0) = 0$ and $\psi^u(\rho^0) = \omega^0$. In this case, the score of the ma(1) log-likelihood function evaluated at the pseudo-true parameter values will be given by the following expressions:

$$\frac{1}{\omega^0}\frac{1}{T}\sum_t x_t x_{t-1} = \frac{1}{\omega^0}\sigma_{01T}$$

$$\frac{1}{2\omega^0}\frac{1}{T}\sum_t \left(\frac{x_t^2}{\omega^0} - 1\right) = \frac{1}{2(\omega^0)^2}\left[\sigma_{00T} - \omega^0\right]$$

Hence, we can use Lemma 3 directly with $\delta^0 = 0$ to show that

$$\mathcal{I}^u_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1/2(\omega^0)^2 \end{bmatrix}$$

Similarly, it is also easy to prove that for $\rho^0 = (0; \omega^0)'$

$$\mathcal{J}^u_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1/2(\omega^0)^2 \end{bmatrix}$$

so that

$$\sqrt{T}\begin{pmatrix} \hat{\delta}_T \\ \hat{\psi}_T - \omega^0 \end{pmatrix} \xrightarrow{d} N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{bmatrix} 1 & 0 \\ 0 & 2(\omega^0)^2 \end{bmatrix}\right\}$$

as expected, since the true process is white noise, and the ma and ar log-likelihood functions are locally equivalent.

As for the inequality restricted pseudo-ML estimators of δ, ψ, and μ_1, there may be three different situations, according to whether $\phi^0 < 0$, $\phi^0 > 0$ or $\phi^0 = 0$. In the first case, it is easy to see from Propositions 1 and 2 that $\sqrt{T}(\tilde{\delta}_T - \hat{\delta}_T)$, $\sqrt{T}(\tilde{\psi}_T - \hat{\psi}_T)$ and $\sqrt{T}\tilde{\mu}_{1T}$ are all $o_p(1)$, while in the second case the same applies to $\sqrt{T}\tilde{\delta}_T$, $\sqrt{T}(\tilde{\psi}_T - \hat{\psi}_T)$ and $\sqrt{T}(\tilde{\mu}_{1T} - \mu_{1T})$. Once more, the interesting case arises when $\phi^0 = 0$, because $\sqrt{T}\tilde{\delta}_T$ and $\sqrt{T}\tilde{\mu}_{1T}$ have half normal asymptotic distributions. Nevertheless, from Proposition 2 we will again have that $\sqrt{T}(\tilde{\delta}_T - \tilde{\mu}_{1T})$ will share an asymptotic $N(0;1)$ distribution with $\sqrt{T}(\hat{\delta}_T - \hat{\mu}_{1T}) = \sqrt{T}\hat{\delta}_T$ and $\sqrt{T}(\bar{\delta}_T - \bar{\mu}_{1T}) = -\sqrt{T}\bar{\mu}_{1T}$.

3.2.5 Indirect inference estimators

Given the two different expressions for the inequality restricted pseudo-ML estimates of θ and μ discussed previously, the sample counterparts to the population moments (23) will be given by either $m(\rho;\hat{\beta}_T)$, which correspond to the sample moments used by an unrestricted EMM procedure, or $m(\rho;\bar{\beta}_T)$, which will be the moments used by the equality constrained one. But since when we solve $m(\rho;\bar{\beta}_T) = 0$ we get

$$
\begin{aligned}
\bar{\delta}_T &= -\bar{\mu}_{1T} = \frac{\hat{\sigma}_{10T}}{\hat{\sigma}_{00T}} \\
\bar{\omega}_T &= \bar{\phi}_T(1-\bar{\delta}_T^2) = \hat{\sigma}_{00T} - \frac{\hat{\sigma}_{10T}^2}{\hat{\sigma}_{00T}}
\end{aligned}
$$

it is clear that the equality constrained EMM estimator converges in probability to the first order sample autocorrelation, which is the maximum likelihood estimator of the parameter of interest. Hence, it is always at least as efficient as the unrestricted EMM estimator. Note that this is true regardless of the sign of $\delta^u(\rho^0)$, and therefore independently of whether or not $\phi^0 = 0$. Of course, if we knew that $\delta^u(\rho^0) = 0$, or any other value for that matter, we could recover ϕ^0 from the binding function directly without estimation error (cf. Dridi, 2000). The same result applies to the corresponding equality constrained II estimators, which minimise the MD objective function

$$
D_T^i(\rho;\mathbf{I}_2) = \left[\bar{\mu}_{1T} + \frac{\gamma_1(\rho)}{\gamma_0(\rho)}\right]^2 + \left[\bar{\phi}_T - \gamma_0(\rho)\right]^2
$$

As for the inequality restricted estimators, it depends on whether or not the pseudo-true value $\delta^i(\rho^0)$ is 0 or strictly negative (or the associated Kuhn-Tucker multiplier $\mu_1^{-1}(\rho^0)$ is 0 or strictly positive). If $\phi^0 > 0$, then $\check{\rho}_T$ will be asymptotically equivalent to the unrestricted estimator $\hat{\rho}_T$ because the sign restriction on ψ_T is not binding in large samples. As a result, the inequality restricted estimators will be less efficient than the equality constrained ones. If on the other hand, $\phi^0 < 0$, the restriction is almost surely binding in the limit, and therefore $\check{\rho}_T$ will be asymptotically equivalent to the equality restricted estimator $\tilde{\rho}_T$. Finally, the most interesting situation arises when $\phi^0 = 0$. In this case, since the unrestricted pseudo log-likelihood nests the true log-likelihood, the unrestricted estimators will also be as efficient as maximum likelihood by virtue of Theorem 2 in GT. But since the inequality restricted estimators will be a 50:50 mixture of $\hat{\rho}_T$ and $\tilde{\rho}_T$ in large samples, it will share their common asymptotic distribution.

A similar line of reasoning can be applied to a pre-test estimator that uses either $\tilde{\rho}_T$ when a standard LM test for first order serial correlation does not reject the null hypothesis of white noise, or $\hat{\rho}_T$ when it does. Since such an LM test is consistent in the context of the AR(1) model (19), then the pretest EMM estimator will always be asymptotically equivalent to $\hat{\rho}_T$, and therefore inefficient relative to $\check{\rho}_T$ except when $\phi^0 = 0$.

3.3 Stochastic volatility estimated as GARCH(1,1) with Gaussian and Student's t distributed errors

3.3.1 True and auxiliary models

Consider the following log-normal stochastic volatility process

$$\begin{aligned} x_t &= \sqrt{h_t}\,u_t \\ \ln h_t &= \alpha + \delta \ln h_{t-1} + \sigma_v v_t \end{aligned} \qquad (25)$$

where $|\delta| < 1; 0 < \sigma_v < \infty$, and $(u_t; v_t)'|x_{t-1}; \ldots \sim N(0; I_2)$. This model was originally proposed as an alternative to the ARCH class, and can be regarded as the discrete time analogue of the continuous time Orstein-Uhlenbeck stochastic processes for instantaneous log volatility frequently used in the theoretical finance literature. Unfortunately, it is impossible to find analytical expressions for the conditional distribution of x_t based on its own past values alone, despite the fact that its distribution conditional on $h_t; x_{t-1}; \ldots$ is Gaussian, with zero mean and variance h_t. Given its importance, though, it is not surprising that a voluminous collection of research papers has been devoted to the estimation of the parameters of interest $\rho = (\alpha; \delta; \sigma_v)'$ (see Shephard (1996) for a survey).

In an influential such paper, Kim, Shephard and Chib (1998) consider likelihood-based estimators of (25), and analyse its goodness of fit relative to some popular ARCH-type competitors. In particular, they find that the log-normal stochastic model above and a GARCH(1,1) model with (standardised) Student's t distributed errors fit the data equally well, as long as the additional tail-thickness parameter is not set to its limiting value under Gaussianity. Therefore, since the latter has a conditional density that can be written in closed form, it looks like the ideal candidate for auxiliary model. On this basis, the model we estimate is given by

$$x_t = \sqrt{\lambda_t}\varepsilon_t$$
$$\lambda_t = \psi + \varphi x_{t-1}^2 + \pi \lambda_{t-1}$$

where $\varepsilon_t|x_{t-1} \ldots$ follows a standardised Student's t distribution with η^{-1} degrees of freedom,[5] so that $\theta = (\psi; \varphi; \pi; \eta)'$. Note that by having an extra parameter, the auxiliary model (seemingly) overidentifies ρ. As is well known, the standardised t distribution nests the standard normal for $\eta = 0$, but otherwise has fatter tails. Also note that like in the previous two examples, the auxiliary and true models are non-nested except in the trivial case in which x_t is Gaussian white noise.

[5] Since the implied degrees of freedom parameter can take any real value above 2, in fact the errors have a distribution that is $\sqrt{(1-2\eta)/\eta}$ times the ratio of a standard normal to the square root of an independent gamma variate with parameters $1/2\eta$ and 2.

The parameters of the auxiliary model are usually estimated subject to several inequality restrictions for the following reasons:

1. As discussed by e.g. Nelson and Cao (1991), when ε_t^2 has infinite support, the conditional variance λ_t will be nonnegative with probability one if $\psi \geq 0$, $\varphi \geq 0$ and $\pi \geq 0$.

2. The pseudo-ML estimators of θ may not be well behaved when $\varphi + \pi > 1$ (see Lumsdaine, 1996).

3. The pseudo log-likelihood function based on the standardised Student's t distribution cannot be defined when the inverse of the degrees of freedom parameter is either negative, or exceeds 1/2.

4. When $\varphi = 0$, π becomes asymptotically underidentified, which may also happen in finite samples depending on the treatment of the initial observations (see e.g. Andrews, 1999).

As a consequence, we estimate the auxiliary model subject to the following set of inequality constraints:

$$\psi \geq 0; \quad \varphi \geq \varphi_{min}; \quad \pi \geq 0; \quad \varphi + \pi \leq 1; \quad 0 \leq \eta \leq \eta_{max} \tag{26}$$

where γ_{min}, and $1/2 - \eta_{max}$ are arbitrarily chosen small values.[6]

In addition, the Student's t-based log-likelihood function often becomes rather flat for very small values of η, because it is very difficult to numerically distinguish a standardised t with 2,000 degrees of freedom from another one with 5,000 degrees of freedom, or indeed from their Gaussian limit. In fact, we effectively set $\eta = 0$ whenever $\eta < \eta_{min}$ to avoid large numerical errors in the computation

[6] After some experimentation, we chose $\gamma_{min} = .025$, and $\eta_{max} = .499$, which corresponds to 2.04 degrees of freedom.

of the derivatives.[7] For that reason, we also consider a mixed equality/inequality estimator that sets η to 0 to obtain a Gaussian pseudo log-likelihood function, but which computes the value of the corresponding multiplier from the relevant first order condition. For the sake of brevity, we refer to the estimator that allows η to vary freely within its bounds as the "inequality restricted" estimator, and to the other as the "equality restricted" one. Nevertheless, the remaining auxiliary parameters are always estimated subject to the other bounds in (26).

3.3.2 Monte Carlo study

We assess the performance of our proposed procedures by means of an extended Monte Carlo analysis, with the same experimental design as Jacquier, Polson and Rossi (1994) (JPR). In this respect, the results in JPR suggest that the most important determinant of the performance of the different estimators will be the unconditional coefficient of variation of the unobserved volatility level h_t, κ say, where

$$\kappa^2 = \frac{V(h_t)}{E^2(h_t)} = \exp\left(\frac{\sigma_v^2}{1-\delta^2}\right) - 1$$

Intuitively, the reason is that when κ^2 is low, the observed process is close to Gaussian white noise, and the estimation of the stochastic volatility parameters is difficult. Unfortunately, the existing empirical evidence suggests that low κ^2's are the rule, rather than the exception (see JPR and the references therein).

The Monte Carlo designs considered by JPR in their tables 5, 6 and 7, have nine entries, arranged in three rows and columns. The rows are defined in terms of κ^2, and the columns by the autocorrelation coefficient for log volatility, δ. Finally, the remaining parameter ψ is chosen so that the unconditional mean of the volatility level equals .0009. Although most of their reported results correspond to a sample size of $T = 500$ observations, we have also considered $T = 1{,}000$ and $2{,}000$.

[7] We chose $\eta_{min} = .0005$, which corresponds to 2,000 degrees of freedom.

For convenience, we first optimise the pseudo log-likelihood function in terms of some unrestricted parameters θ^*, where $\psi = \theta_1^{*2}$, $\phi = \phi_{min} + (1 - \phi_{min}) \sin^2(\theta_2^*)$, $\pi = (1 - \phi) \sin^2(\theta_3^*)$ and $\eta = \left[1 - \sin^2(\theta_4^*)\right] \eta_{min} + \sin^2(\theta_4^*) \eta_{max}$. Then, we compute the score in terms of the original parameters $\theta = (\psi; \phi; \pi; \eta)'$ using the analytical expressions derived by Calzolari, Fiorentini and Sentana (2000), and introduce one multiplier for each of the four first order conditions in order to take away any slack left. Since there are no closed-form expressions for the expected value of the modified score, we compute them on the basis of single simulations of length TH, with $H = 10$, as explained in section 2.5. A larger value of H should in theory reduce the Monte Carlo variability of the EMM estimators according to the relation $(1 + H^{-1})$, but at the cost of a significant increase in the computational burden. Finally, we minimise numerically the GMM criterion function in terms of some unrestricted parameters ρ^*, with $\alpha = \rho_1^*$, $\delta = \delta_{max} \sin(\rho_2^*)$ and $\sigma_v = \rho_3^{*2}$, where $\delta_{max} = .9999$, so as to ensure that $|\delta| < 1$ and $\sigma_v \geq 0$.

Tables 1, 2 and 3 contain the proportion of inequality and equality restricted pseudo-ML estimators of θ that satisfy with equality the different restrictions in (26). When κ^2 is 1, such restrictions are hardly ever binding, especially for $T = 2{,}000$. However, when κ^2 is large (=10), most of the estimated garch models are of the igarch variety. This is particularly true when η is free, but it also happens when the conditional distribution is assumed Gaussian. Somewhat surprisingly, such a finding does not seem to constitute a finite sample problem, because the proportion of boundary cases actually increases with the sample size. In contrast, in those empirically relevant situations in which κ^2 is small (=.1), igarch parameter configurations are hardly ever estimated, but the estimates of the arch and garch coefficients, and the reciprocal of the degrees of freedom parameter, reach their lower bounds fairly often, especially for the smaller sample sizes. For instance, when $T = 500$ and $\delta = .98$, almost 60% of the simulations have inequality constrained pseudo-ML estimators for which at least one of those restrictions is binding. As pointed out by Shephard (1996), part of the empirical success of the stochastic volatility and t-garch models simply lies on their ability to capture the fat-tailed behaviour of asset returns. Therefore, when one tries to fit a t-distributed garch(1,1) auxiliary model to artificial data that shows little volatility clustering, and only a small degree of leptokurtosis, it is not totally surprising that one ends up with parameter estimates that correspond to Gaussian white noise. In any case, the results clearly show that our proposed generalisations of EMM and II procedures are not only of theoretical interest, but also highly relevant in practice.

Tables 4 to 9 present the means, root mean square errors, mean biases and standard deviations of the inequality and equality restricted EMM estimators of the parameters of interest ρ for the case in which the optimal GMM weighting matrix is estimated as the variance in the original data of the modified score of the auxiliary model evaluated at the pseudo-ML parameter estimates. In this respect, note that by including a multiplier in each first order condition, we automatically centre the scores around their sample mean. Given that the auxiliary model tends to fit the simulated data rather well, we have not included any correction for serial correlation (cf. GT).

As expected, the estimates of the autoregressive parameter δ are downward biased. This is particularly so when δ^0 is high, and/or σ_v^0 low, which mimics the behaviour of a pseudo-ML estimator of the autoregressive parameter of an ar(1) process observed subject to measurement error. And exactly like in that situation, the downward bias in the estimator of δ is transmitted into an upward bias in the absolute value of the estimates of the mean constant, α, and the standard deviation of the log-volatility innovations σ_v. Therefore, it is not surprising that the most important determinant of the performance of the different estimators is precisely κ^2, which effectively plays the role of a signal to noise ratio.

But perhaps more importantly for our purposes, neither of the two restricted versions of the EMM estimator seems to dominate the other across all Monte Carlo designs. When κ^2 is 10, the inequality restricted EMM estimator systematically outperforms the equality restricted one in terms of root mean square error, although not necessarily in terms of mean bias for $T = 500$. In contrast, when κ^2 is .1, the equality restricted EMM estimator tends to outperform the inequality restricted one, except perhaps as far as σ_v is concerned. The reason is that when the behaviour of the data is close to Gaussian white noise, our attempts to estimate simultaneously the reciprocal of the degrees of freedom, η, result in a deterioration of the estimators of the garch parameters relative to the Gaussian case. At the same time, since the first order condition for η is the most directly related to the degree of leptokurtosis of the observed data, the equality restricted EMM estimator of σ_v is somewhat less precise than its inequality restricted counterpart. As for the middle row, the results are mixed, at least for $T = 500$. As T increases, the inequality restricted EMM estimator tends to have a smaller root mean square error than the equality restricted one, at the cost of a slightly higher mean bias.

Finally, a comparison of our results with the ones reported by JPR suggests that our EMM procedures tend to outperform the QML and MM estimators considered by these authors, except in those instances in which, according to JPR, the performance of the latter is exceptionally good. In contrast, the EMM estimators are dominated by the empirical Bayesian estimators proposed by JPR, which is not very surprising given that our auxiliary model does not nest the model of interest, and we do not use any prior information. In this respect, it is important to mention that the relatively poor performance of the EMM estimators is partly due to those simulations in which δ is estimated as being negative. For instance, the root mean square error of the equality restricted estimator of ± in row 2, column 3 of Table 5 decreases from .0765 to .0524 if we exclude the only two negative estimates of ± found in 1,000 replications.

4 CONCLUSIONS

In this paper, we generalise the II approaches of GT and GMR to those empirically relevant situations in which there are constraints on the parameters of the auxiliary model. In the EMM case, speci…cally, we derive the moments used in GMM estimation from either the Kuhn-Tucker …rst order conditions for inequality constraints, or the usual Lagrange …rst order conditions for equality restrictions. Similarly, in the II case, we minimise the distance between an extended vector that includes both pseudo-ML parameter estimates and multipliers, and the corresponding binding functions. Equality constrained estimators may be particularly useful from a computational point of view, since in many situations of interest, it is considerably simpler to estimate a special restricted case of the auxiliary model. We also obtain expressions for the optimal GMM weighting matrix, and the MD one that yields asymptotically equivalent II estimators. In addition, we also consider EMM and II procedures based on partially optimised unconstrained estimators, as well as those that impose the constraints depending on the signi…cance of some preliminary speci…cation test.

For illustrative purposes, we discuss the usual example of ma(1) estimated as ar(1), and show that the inequality restricted EMM and II estimators are asymptotically equivalent to the unrestricted estimators, and indeed, to equality restricted EMM and II estimators that set the autoregressive parameter to 0 in the auxiliary model, but include either the corresponding …rst order condition in the set of moments, or the Lagrange multiplier in the distance function. Importantly, the equivalence of the different EMM and II estimators in this example does not really depend on the specific inequality restriction imposed, or the nature of the true model, but rather on the particular form of the auxiliary model used. In this respect, the same result continues to hold if the auxiliary model is given by a conditionally homoskedastic Gaussian ar(p) process with linear restrictions on the autoregressive parameters. We also discuss the reverse example in which an ar(1) model is estimated via ma(1). It turns out that the equality restricted EMM and II estimators that impose the white noise restriction not only dominate the unrestricted estimators, but also become as efficient as maximum likelihood, even though the auxiliary model does not nest the true one. Finally, we compare the performance of our proposed procedures for a log-normal stochastic volatility process estimated as a garch(1,1) model with either Gaussian or t-distributed errors. In this case, we find that the pseudo-ML estimators are quite often at the boundary of the parameter space. We also find that although neither estimator systematically outperforms the other, the equality restricted estimator dominates the inequality restricted one in those situations in which there is little information in the data about the additional tail-thickness parameter.

Appendix

1 PROOFS OF RESULTS

1.1 Proposition 1

If we linearise the complementary slackness conditions

$$h(\tilde{\theta}_T) \odot \tilde{\gamma}_T = 0$$

around $\psi_T(\rho_0)$, taking into account that $h\left[\theta_T(\rho_0)\right] \odot \gamma_T(\rho_0) = 0$, and that Hadamard products are commutative, we obtain:

$$\gamma_T^* \odot \frac{\partial h'(\theta_T^*)}{\partial \theta} \sqrt{T}\left[\tilde{\theta}_T - \theta_T(\rho_0)\right] + h(\theta_T^*) \odot \sqrt{T}\left[\tilde{\gamma}_T - \gamma_T(\rho_0)\right] = 0$$

where $\psi_T^* = (\theta_T^{*\prime}, \gamma_T^{*\prime})'$ is an "intermediate" value (in fact, a different one for each row). Then, given that in view of our high level assumptions, $\gamma_T^* - \gamma_T(\rho_0) = o_p(1)$, $h(\theta_T^*) - h\left[\theta_T(\rho_0)\right] = o_p(1)$, and $\partial h(\theta_T^*)/\partial\theta - \partial h\left[\theta_T(\rho_0)\right]/\partial\theta = o_p(1)$, the result follows. ∎

1.2 Proposition 2

If we linearise the first-order Kuhn-Tucker conditions

$$\frac{\sqrt{T}}{T} \sum_t \left[\frac{\partial l_t(\tilde{\theta}_T)}{\partial \theta} + \frac{\partial h'(\tilde{\theta}_T)}{\partial \theta} \tilde{\gamma}_T \right] = 0$$

around $\psi_T(\rho_0)$, we obtain:

$$\frac{\sqrt{T}}{T} \sum_t \left[\frac{\partial l_t\left[\theta_T(\rho_0)\right]}{\partial \theta} + \frac{\partial h'\left[\theta_T(\rho_0)\right]}{\partial \theta} \gamma_T(\rho_0) \right]$$

$$+ \frac{1}{T} \sum_t \left\{ \frac{\partial l_t^2(\bar{\theta}_T)}{\partial \theta \partial \theta'} + (\bar{\gamma}_T \otimes I_q) \frac{\partial \operatorname{vec}\left[\partial h'(\bar{\theta}_T)/\partial \theta\right]}{\partial \theta'} \right\} \sqrt{T}\left[\tilde{\theta}_T - \theta_T(\rho_0)\right]$$

$$+ \frac{\partial h'(\bar{\theta}_T)}{\partial \theta} \sqrt{T}\left[\tilde{\gamma}_T - \gamma_T(\rho_0)\right]$$

where $\bar{\beta}^\star_T = (\theta^{\star\prime}_T, \gamma^{\star\prime}_T)'$ is another "intermediate" value. Then, since in view of Assumption 1

$$\frac{1}{T}\sum_t \left\{\frac{\partial l_t^2(\theta_T^\star)}{\partial\theta\partial\theta'}\right\} = \mathcal{J}^+_{0T} + o_p(1)$$

$$(\gamma^\star_T \otimes I_q)\frac{\partial vec[\partial h'(\theta^\star_T)/\partial\theta]}{\partial\theta'} = \left[\gamma^+_T(\rho^0)\otimes I_q\right]\frac{\partial vec\left\{\partial h'\left[\theta^+_T(\rho^0)\right]/\partial\theta\right\}}{\partial\theta'} + o_p(1)$$

$$\frac{\partial h'(\theta^\star_T)}{\partial\theta} = \frac{\partial h'\left[\theta^+_T(\rho^0)\right]}{\partial\theta} + o_p(1)$$

a straightforward application of Crámer's theorem completes the proof. □

1.3 Proposition 3

Let's now linearise the sample moments $m_T(\rho^0;\tilde{\beta}_T)$ around $\beta^+_T(\rho^0)$ to obtain

$$\sqrt{T}m_T(\rho^0;\tilde{\beta}_T) = \sqrt{T}m_T\left[\rho^0;\beta^+_T(\rho^0)\right] + \frac{\partial m_T(\rho^0;\bar{\beta}^+_T)}{\partial\theta'}\sqrt{T}\left\langle\tilde{\theta}_T-\theta^+_T(\rho^0)\right\rangle + \frac{\partial m_T(\rho^0;\bar{\beta}^+_T)}{\partial\gamma'}\sqrt{T}\left[\tilde{\gamma}_T-\gamma^+_T(\rho^0)\right]$$

where $\bar{\beta}^+_T$ is yet another "intermediate" value. This implies that under Assumption 1, $\sqrt{T}m_T(\rho^0;\tilde{\beta}_T)$ has the same asymptotic distribution as

$$\frac{\partial m^0_T\left[\rho^0;\beta^+_T(\rho^0)\right]}{\partial\theta}\sqrt{T}\left\langle\tilde{\theta}_T-\theta^+_T(\rho^0)\right\rangle + \frac{\partial m^0_T\left[\rho^0;\beta^+_T(\rho^0)\right]}{\partial\gamma}\sqrt{T}\left[\tilde{\gamma}_T-\gamma^+_T(\rho^0)\right]$$

where

$$\frac{\partial m_T\left[\rho^0;\beta^+_T(\rho^0)\right]}{\partial\theta'} = \mathcal{J}^+_{0T} + \left[\gamma^+_T(\rho^0)\otimes I_q\right]\frac{\partial vec\left\{\partial h'\left[\theta^+_T(\rho^0)\right]/\partial\theta\right\}}{\partial\theta'} = \mathcal{K}^+_{11;0T}$$

$$\frac{\partial m_T\left[\rho^0;\beta^+_T(\rho^0)\right]}{\partial\gamma'} = \frac{\partial h'\left[\theta^+_T(\rho^0)\right]}{\partial\theta} = \mathcal{K}^+_{12;0T}$$

But then, Proposition 2 directly yields the required result □

1.4 Proposition 4

The first order conditions associated with $\tilde{\rho}^G_T[(\mathcal{I}^{-}_{0T})^{-1}]$ can be written as

$$\frac{\partial m'_T\left\{\tilde{\rho}^G_T[(\mathcal{I}^{-}_{0T})^{-1}];\tilde{\beta}_T\right\}}{\partial\rho}\cdot(\mathcal{I}^{-}_{0T})^{-1}\cdot\sqrt{T}\,m_T\left\{\tilde{\rho}^G_T\left[(\mathcal{I}^{-}_{0T})^{-1}\right];\tilde{\beta}_T\right\}=\mathbf{0}$$

Expanding around ρ^0 yields

$$\begin{aligned}&\frac{\partial m'_T\left(\rho^0;\tilde{\beta}_T\right)}{\partial\rho}\cdot(\mathcal{I}^{-}_{0T})^{-1}\cdot\sqrt{T}\,m_T\left(\rho^0;\tilde{\beta}_T\right)\\&+\frac{\partial m'_T\left(\rho^*_T;\tilde{\beta}_T\right)}{\partial\rho}\cdot(\mathcal{I}^{-}_{0T})^{-1}\cdot\frac{\partial m_T\left(\rho^*_T;\tilde{\beta}_T\right)}{\partial\rho}\sqrt{T}\left[\tilde{\rho}^G_T(\mathcal{I}^{-}_{0T})-\rho^0\right]\\&+\left[(\mathcal{I}^{-}_{0T})^{-1}\cdot m_T\left(\rho^*_T;\tilde{\beta}_T\right)\otimes I_d\right]\frac{\partial vec\left[\partial m'_T(\rho^*_T;\tilde{\beta}_T)/\partial\rho\right]}{\partial\rho'}\sqrt{T}\left[\tilde{\rho}^G_T(\mathcal{I}^{-}_{0T})-\rho^0\right]\end{aligned}$$

where ρ^*_T is some "intermediate" value. But since $m_T\left(\rho^*_T;\tilde{\beta}_T\right)$ is $o_p(1)$, we finally have that

$$\begin{aligned}\sqrt{T}\left[\tilde{\rho}^G_T(\mathcal{I}^{-}_{0T})-\rho^0\right]&=\left\{\frac{\partial m'_T\left[\rho^0;\beta^{-}_T(\rho^0)\right]}{\partial\rho}\cdot(\mathcal{I}^{-}_{0T})^{-1}\cdot\frac{\partial m_T\left[\rho^0;\beta^{-}_T(\rho^0)\right]}{\partial\rho'}\right\}^{-1}\\&\times\frac{\partial m'_T\left[\rho^0;\beta^{-}_T(\rho^0)\right]}{\partial\rho}(\mathcal{I}^{-}_{0T})^{-1}\sqrt{T}\,m_T\left(\rho^0;\tilde{\beta}_T\right)+o_p(1)\end{aligned}$$

as required. ¤

1.5 Proposition 5

The result follows directly if we combine the proofs of Propositions 2 and 3 to show that

$$\sqrt{T}\,m_T(\rho^0;\tilde{\beta}_T)-\left\{\mathcal{K}^{-}_{11;0T}\sqrt{T}\left[\tilde{\mu}_T-\mu^{-}_T(\rho^0)\right]+\mathcal{K}^{-}_{12;0T}\sqrt{T}\left[\tilde{\pi}_T-\pi^{-}_T(\rho^0)\right]\right\}=o_p(1)$$

2 THE EXPECTED VALUE OF THE SCORE OF AN MA(1) MODEL

In order to find

$$
\begin{aligned}
m_1(\rho;\beta) &= E\left[\frac{1}{\psi}u_t(\delta)\frac{\partial \nu_t(\delta)}{\partial\delta} + \mu_1 \,\Big|\, \rho\right] \\
m_2(\rho;\beta) &= E\left[\frac{1}{\psi}\left(\frac{u_t^2(\delta)}{\psi} - 1\right) + \mu_2 \,\Big|\, \rho\right]
\end{aligned}
$$

it is convenient to write

$$
u_t(\delta) = \sum_{j=0} \delta^j x_{t-j} = \frac{1}{1-\delta L}x_t
$$

and

$$
\frac{\partial \nu_t(\delta)}{\partial\delta} = -\sum_{j=1} j\delta^{j-1}x_{t-j} = \frac{-L}{(1-\delta L)^2}x_t
$$

so that we can understand both $u_t(\delta)$ and $\partial\nu_t(\delta)/\partial\delta$ as the output of linear filters applied to the original series x_t. In this light, we can obtain the required expectations as the constant terms in the autocovariance generating function of $u_t^2(\delta)$ and $u_t(\delta)\cdot\partial\nu_t(\delta)/\partial\delta$. In particular, $\Gamma_{u_t(\delta);u_t(\delta)}(z)$ will be given by

$$
\begin{aligned}
&\frac{1}{1-\delta z}\cdot\Gamma_{x_t}(z)\cdot\frac{1}{1-\delta z^{-1}} \\
&= \frac{1}{1-\delta^2}\left(\begin{array}{l}
\gamma_0(\rho)\left[1+\sum_{j=1}^{\infty}\delta^j(z^j+z^{-j})\right] \\
+\sum_{l=1}^{\infty}\gamma_l(\rho)z^l\left[1+\sum_{j=1}^{\infty}\delta^j(z^j+z^{-j})\right] \\
+\sum_{l=1}^{\infty}\gamma_l(\rho)z^{-l}\left[1+\sum_{j=1}^{\infty}\delta^j(z^j+z^{-j})\right]
\end{array}\right)
\end{aligned}
$$

Hence,

$$
E\left[u_t^2(\delta)\,\big|\,\rho\right] = \frac{\gamma_0(\rho)}{1-\delta^2}\left(1+2\sum_{l=1}\delta^l\frac{\gamma_l(\rho)}{\gamma_0(\rho)}\right)
$$

which for the special case of the true process being a stationary AR(1) reduces to

$$
E\left[u_t^2(\delta)\,\big|\,\rho\right] = \frac{1}{\left(1-\delta^2\right)\left(1-\phi^2\right)}\cdot\frac{1-\delta\phi}{1+\delta\phi}
$$

In fact, given that we can write

$$u_t(\delta) = \frac{1}{1-\delta L}x_t = \frac{1}{(1-\delta L)(1-\varphi L)}v_t;$$

it is not surprising that $E\left[u_t^2(\delta)|\rho\right]$ coincides with the unconditional variance of an AR(2) process with autoregressive roots δ and φ, and innovation variance ω.

Similarly, the cross-covariance generating function of $\partial\gamma_t(\delta)/\partial\delta$ and $u_t(\delta)$, $\Gamma_{\partial\gamma_t(\delta)/\partial\delta; u_t(\delta)}(z)$, will be given by (minus) the following expression

$$\begin{aligned}
&\frac{z}{(1-\delta z)^2}\cdot\Gamma_{x_t}(z)\cdot\frac{1}{1-\delta z^{-1}} \\
=& \sum_{j=1}^{\infty} j\delta^{j-1}z^j \times \left(\gamma_0(\rho) + \sum_{l=1}^{\infty}\gamma_l(\rho)z^l + \sum_{l=1}^{\infty}\gamma_l(\rho)z^{-l}\right) \times \left(1 + \sum_{k=1}^{\infty}\delta^k z^{-k}\right) \\
=& \gamma_0(\rho)\sum_{j=1}^{\infty} j\delta^{j-1}z^j + \sum_{j=1}^{\infty}\sum_{l=1}^{\infty} j\delta^{j-1}\gamma_l(\rho)(z^l+z^{-l})z^j \\
&+\gamma_0(\rho)\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} j\delta^{j-1}\delta^k z^j z^{-k} + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} j\delta^{j-1}\delta^k\gamma_l(\rho)(z^l+z^{-l})z^{-k}z^j
\end{aligned}$$

Therefore, the coefficient associated with the constant term will be

$$\gamma_0(\rho)\sum_{j=1}^{\infty} j\delta^{2j-1} + \sum_{l=1}^{\infty} l\delta^{l-1}\gamma_l(\rho) + 2\sum_{l=1}^{\infty}\delta^l\gamma_l(\rho)\sum_{j=1}^{\infty} j\delta^{2j-1} + \sum_{l=1}^{\infty} l\delta^l\gamma_l(\rho)\sum_{j=1}^{\infty}\delta^{2j-1}$$

But since for $|\delta| < 1$

$$\begin{aligned}
\sum_{j=1}^{\infty}\delta^{2j-1} &= \delta\sum_{j=0}^{\infty}\delta^{2j} = \frac{\delta}{1-\delta^2} \\
\sum_{j=1}^{\infty} j\delta^{2j-1} &= \delta\sum_{j=0}^{\infty}(j+1)\delta^{2j} = \frac{\delta}{\left(1-\delta^2\right)^2}
\end{aligned}$$

we will have that

$$E\left[u_t(\delta)\frac{\partial\gamma_t(\delta)}{\partial\delta}\Big|\rho\right] = -\frac{\gamma_0(\rho)}{(1-\delta^2)^2}\left\{\delta + \sum_{l=1}^{\infty}\left[2\delta^2 + \left(1-\delta^2\right)l\right]\delta^{l-1}\frac{\gamma_l(\rho)}{\gamma_0(\rho)}\right\}$$

For the special case of a stationary AR(1) process, this expression reduces to:

$$E\left[u_t(\delta)\frac{\partial\gamma_t(\delta)}{\partial\delta}\Big|\rho\right] = \frac{\omega}{(1-\varphi^2)\left(1-\delta^2\right)^2(1-\delta\varphi)^2}\left(\delta^3\varphi^2 + \delta^2\varphi - \delta - \varphi\right)$$

Figure 1: Binding Functions for MA(1) estimated as AR(1)

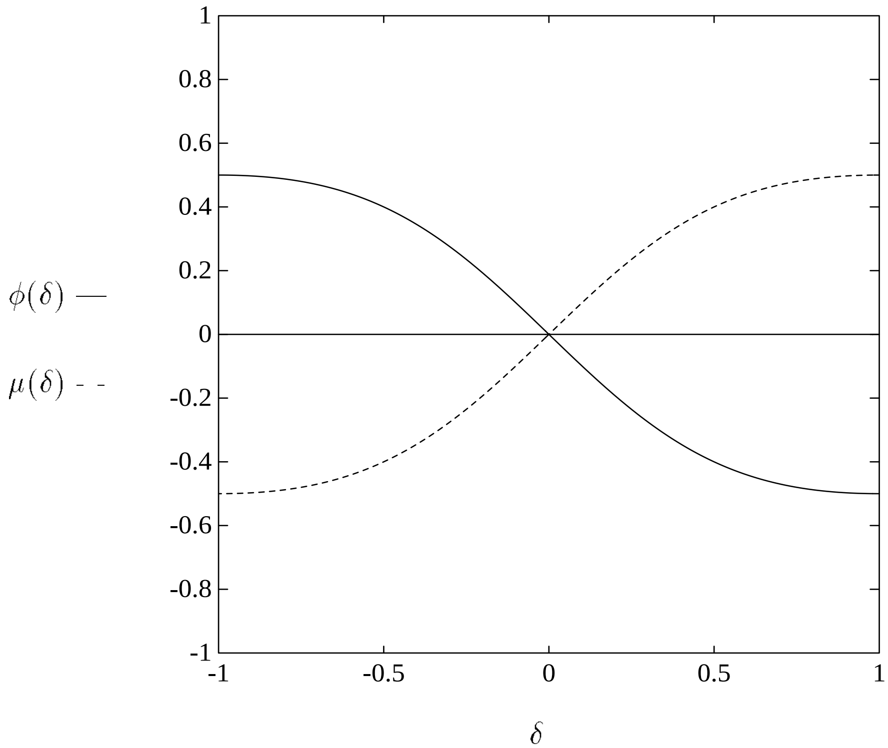

Figure 2: Binding Functions for AR(1) estimated as MA(1)

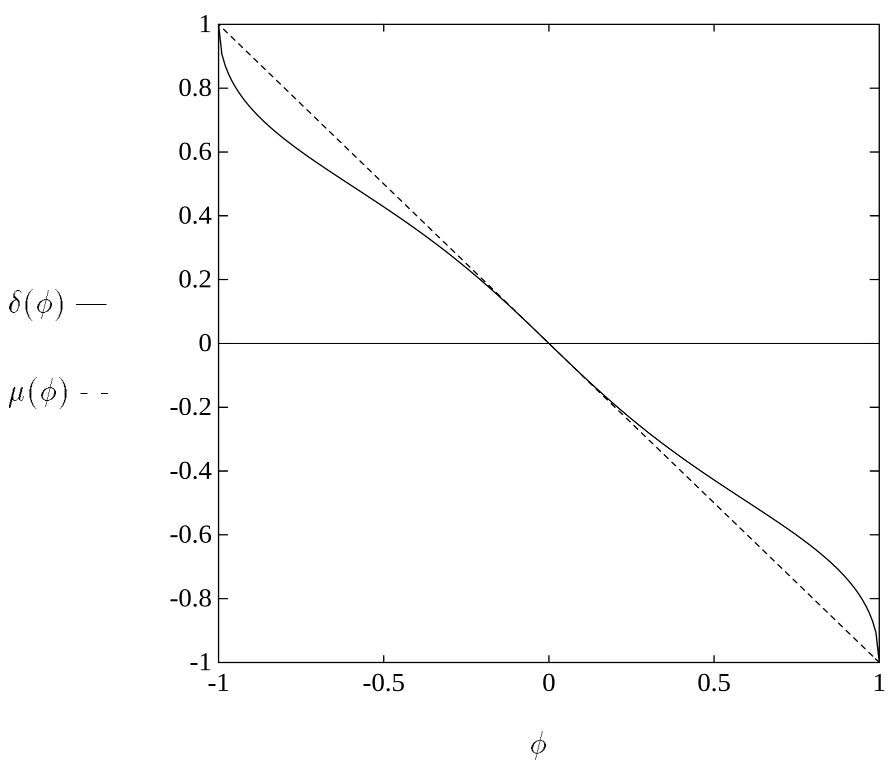

Table 1

Proportion of auxiliary model parameter estimates at the boundary (Inequality/Equality)

T=500, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
$\varphi = \varphi_{\min}$		0/0			0/0			0/0	
$\pi = 0$		0/0			0/0			0/0	
$\varphi + \pi = 1$		.967/.815			.949/.867			.816/.762	
$\eta = \eta_{\min}$		0/*1*			0/*1*			0/*1*	
total		.967/.815			.949/.867			.816/.762	
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
$\varphi = \varphi_{\min}$		.002/.004			.003/.002			.006/.003	
$\pi = 0$		.003/.003			.001/0			.003/.004	
$\varphi + \pi = 1$		.012/.010			.063/.047			.111/.076	
$\eta = \eta_{\min}$		0/*1*			0/*1*			.014/*1*	
total		.015/.016			.066/.049			.132/.082	
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
$\varphi = \varphi_{\min}$		.291/.287			.260/.280			.306/.328	
$\pi = 0$		.169/.177			.133/.162			.149/.115	
$\varphi + \pi = 1$		0/.004			.001/.004			0/.001	
$\eta = \eta_{\min}$		.215/*1*			.264/*1*			.299/*1*	
total		.533/.383			.526/.363			.577/.393	

Table 2

Proportion of auxiliary model parameter estimates at the boundary (Inequality/Equality)

T=1,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
$\varphi = \varphi_{\min}$		0/0			0/0			0/0	
$\pi = 0$		0/0			0/0			0/0	
$\varphi + \pi = 1$		.995/.894			.989/.954			.960/.918	
$\eta = \eta_{\min}$		0/*1*			0/*1*			0/*1*	
total		.995/.894			.989/.954			.960/.918	
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
$\varphi = \varphi_{\min}$		0/0			0/0			0/0	
$\pi = 0$		0/0			0/0			0/.001	
$\varphi + \pi = 1$		.001/.001			.030/.020			.112/.081	
$\eta = \eta_{\min}$		0/*1*			0/*1*			.002/*1*	
total		.001/.001			.030/.020			.114/.082	
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
$\varphi = \varphi_{\min}$		.215/.228			.188/.213			.239/.241	
$\pi = 0$		.082/.100			.059/.059			.051/.035	
$\varphi + \pi = 1$		0/.003			0/0			0/0	
$\eta = \eta_{\min}$		.113/*1*			.126/*1*			.169/*1*	
total		.352/.277			.320/.239			.386/.260	

Table 3

Proportion of auxiliary model parameter estimates at the boundary (Inequality/Equality)

T=2,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
$\varphi = \varphi_{\min}$		0/0			0/0			0/0	
$\pi = 0$		0/0			0/0			0/0	
$\varphi + \pi = 1$		1/.973			1/.995			.998/.988	
$\eta = \eta_{\min}$		0/*1*			0/*1*			0/*1*	
total		1/.973			1/.995			.998/.988	
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
$\varphi = \varphi_{\min}$		0/0			0/0			0/0	
$\pi = 0$		0/0			0/0			0/0	
$\varphi + \pi = 1$		0/0			.009/.002			.089/.069	
$\eta = \eta_{\min}$		0/*1*			0/*1*			0/*1*	
total		0/0			.009/.002			.089/.069	
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
$\varphi = \varphi_{\min}$		.147/.153			.130/.128			.198/.192	
$\pi = 0$		.027/.034			.015/.012			.008/.006	
$\varphi + \pi = 1$		0/.001			0/0			0/0	
$\eta = \eta_{\min}$		.034/*1*			.056/*1*			.096/*1*	
total		.197/.169			.186/.133			.281/.194	

Table 4

Mean, root mean square error, mean bias and standard deviation of the inequality restricted EMM estimator

T=500, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.9613	.8834	.6804	-.5549	.9325	.4959	-.3342	.9595	.3290
rmse	.3902	.0468	.1003	.3124	.0381	.0844	.3181	.0389	.0742
mean bias	-.1403	-.0166	.0054	-.1443	-.0175	.0124	-.1700	-.0205	.0210
std. dev.	.3641	.0438	.1001	.2771	.0339	.0835	.2688	.0330	.0712
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-1.0108	.8628	.3840	-.6402	.9130	.2907	-.4013	.9452	.1997
rmse	.6969	.0929	.1064	.6284	.0876	.0973	.5431	.0741	.0913
mean bias	-.2748	-.0372	.0210	-.2722	-.0370	.0307	-.2541	-.0348	.0337
std. dev.	.6404	.0851	.1044	.5663	.0794	.0923	.4800	.0654	.0848
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-2.3819	.6642	.1712	-1.9527	.7247	.1526	-1.5684	.7792	.1227
rmse	3.3865	.4773	.1418	3.1102	.4383	.1420	2.8693	.4025	.1343
mean bias	-1.6759	-.2358	.0362	-1.5997	-.2253	.0562	-1.4274	-.2008	.0613
std. dev.	2.9428	.4150	.1371	2.6673	.3759	.1304	2.4891	.3488	.1196

Table 5

Mean, root mean square error, mean bias and standard deviation of the equality restricted EMM estimator

T=500, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.9245	.8884	.5836	-.5249	.9365	.4386	-.3084	.9629	.2958
rmse	.5582	.0657	.1926	.3759	.0438	.1305	.2772	.0332	.0914
mean bias	-.1035	-.0116	-.0914	-.1143	-.0135	-.0449	-.1442	-.0171	-.0122
std. dev.	.5485	.0647	.1695	.3581	.0417	.1225	.2367	.0284	.0906
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.9546	.8711	.3423	-.5671	.9237	.2577	-.3620	.9515	.1781
rmse	.7267	.0983	.1213	.5590	.0739	.1023	.6130	.0765	.0845
mean bias	-.2186	-.0289	-.0207	-.1991	-.0263	-.0023	-.2148	-.0285	.0121
std. dev.	.6930	.0940	.1195	.5523	.0690	.1023	.5741	.0710	.0837
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-2.2013	.6892	.1636	-1.8327	.7416	.1449	-1.4194	.7999	.1131
rmse	3.0879	.4363	.1380	2.8495	.4019	.1340	2.6460	.3724	.1251
mean bias	-1.4953	-.2108	.0286	-1.4797	-.2083	.0485	-1.2784	-.1801	.0517
std. dev.	2.7017	.3820	.1350	2.4352	.3437	.1249	2.3167	.3260	.1139

Table 6

Mean, root mean square error, mean bias and standard deviation of the inequality restricted EMM estimator

T=1,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.8752	.8940	.6726	-.4726	.9427	.4870	-.2395	.9706	.3188
rmse	.2388	.0281	.0712	.1836	.0225	.0568	.1498	.0183	.0445
mean bias	-.0542	-.0060	-.0024	-.0620	-.0073	.0035	-.0753	-.0094	.0108
std. dev.	.2326	.0275	.0711	.1728	.0213	.0567	.1295	.0157	.0432
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.8426	.8859	.3703	-.4679	.9366	.2725	-.2521	.9658	.1840
rmse	.3343	.0447	.0677	.2396	.0326	.0541	.2023	.0277	.0485
mean bias	-.1066	-.0141	.0073	-.0999	-.0134	.0125	-.1049	-.0142	.0180
std. dev.	.3168	.0425	.0673	.2178	.0297	.0527	.1730	.0237	.0450
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.7295	.7559	.1684	-1.2584	.8226	.1381	-.9353	.8680	.1062
rmse	2.3196	.3270	.1085	2.1294	.2995	.1010	1.9963	.2813	.0955
mean bias	-1.0235	-.1441	.0334	-.9054	-.1274	.0417	-.7943	-.1120	.0449
std. dev.	2.0815	.2936	.1033	1.9273	.2711	.0920	1.8314	.2581	.0843

Table 7

Mean, root mean square error, mean bias and standard deviation of the equality restricted EMM estimator

T=1,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.8228	.9000	.6032	-.4610	.9443	.4524	-.2373	.9714	.3020
rmse	.3243	.0378	.1416	.2337	.0268	.0953	.1734	.0199	.0622
mean bias	-.0018	.0000	-.0718	-.0504	-.0057	-.0311	-.0731	-.0086	-.0060
std. dev.	.3243	.0378	.1220	.2282	.0262	.0901	.1572	.0179	.0619
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.8249	.8885	.3497	-.4557	.9385	.2580	-.2480	.9668	.1753
rmse	.3677	.0490	.0869	.2518	.0336	.0653	.3091	.0388	.0530
mean bias	-.0889	-.0115	-.0133	-.0877	-.0115	-.0020	-.1001	-.0132	.0093
std. dev.	.3568	.0476	.0859	.2360	.0315	.0653	.2922	.0364	.0521
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.6778	.7631	.1580	-1.2176	.8282	.1343	-.8975	.8733	.1015
rmse	2.3131	.3264	.1062	2.0173	.2847	.1015	1.8984	.2676	.0923
mean bias	-.9718	-.1368	.0223	-.8646	-.1218	.0380	-.7565	-.1067	.0401
std. dev.	2.0991	.2963	.1037	1.8226	.2573	.0941	1.7412	.2455	.0831

Table 8

Mean, root mean square error, mean bias and standard deviation of the inequality restricted EMM estimator

T=2,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.8590	.8957	.6777	-.4463	.9458	.4872	-.2074	.9747	.3150
rmse	.1603	.0190	.0510	.1157	.0138	.0407	.0932	.0117	.0319
mean bias	-.0380	-.0043	.0027	-.0357	-.0042	.0037	-.0432	-.0053	.0070
std. dev.	.1558	.0185	.0509	.1101	.0132	.0406	.0826	.0104	.0312
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.7960	.8921	.3696	-.4184	.9432	.2676	-.2009	.9727	.1752
rmse	.1955	.0261	.0465	.1323	.0180	.0366	.1109	.0156	.0342
mean bias	-.0600	-.0079	.0066	-.0504	-.0068	.0076	-.0537	-.0073	.0092
std. dev.	.1860	.0249	.0460	.1223	.0167	.0358	.0970	.0137	.0330
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.2351	.8254	.1604	-.7558	.8933	.1256	-.4874	.9313	.0934
rmse	1.4174	.2001	.0791	1.0600	.1493	.0703	1.0634	.1423	.0654
mean bias	-.5291	-.0746	.0254	-.4028	-.0567	.0292	-.3464	-.0487	.0320
std. dev.	1.3150	.1856	.0750	.9805	.1382	.0640	.9555	.1337	.0570

Table 9

Mean, root mean square error, mean bias and standard deviation of the equality restricted EMM estimator

T=2,000, H=10, Fixed GMM weighting matrix, 1,000 replications

κ^2	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*-.821*	*.9*	*.675*	*-.4106*	*.95*	*.4835*	*-.1642*	*.98*	*.308*
mean	-.8229	.8999	.6288	-.4249	.9485	.4576	-.2012	.9758	.3027
rmse	.2397	.0279	.1083	.1351	.0159	.0693	.0952	.0110	.0435
mean bias	-.0019	-.0001	-.0462	-.0143	-.0015	-.0259	-.0370	-.0042	-.0053
std. dev.	.2396	.0279	.0980	.1343	.0158	.0643	.0877	.0102	.0432
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.7798	.8943	.3549	-.4149	.9439	.2593	-.1954	.9736	.1723
rmse	.2212	.0295	.0596	.1498	.0199	.0449	.1043	.0138	.0349
mean bias	-.0438	-.0057	-.0081	-.0469	-.0062	-.0007	-.0482	-.0064	.0063
std. dev.	.2168	.0290	.0590	.1422	.0190	.0449	.0925	.0123	.0343
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.1955	.8310	.1549	-.7130	.8994	.1212	-.4184	.9410	.0885
rmse	1.3864	.1956	.0783	.9254	.1301	.0687	.6934	.0969	.0601
mean bias	-.4895	-.0690	.0199	-.3600	-.0506	.0248	-.2774	-.0390	.0271
std. dev.	1.2971	.1830	.0757	.8525	.1198	.0641	.6355	.0887	.0537

REFERENCES

Anderson, T.W. (1971): The statistical analysis of time series, Wiley.

Andrews, D.W.K. (1999): "Estimation when a parameter in on the boundary", Econometrica 67, 1341-1383.

Bhansali, R.J. (1980): "Autoregressive and window estimates of the inverse correlation function", Biometrika 67, 551-566.

Calzolari, G., Fiorentini, G. and Sentana, E. (2000): "Analytical derivatives of the log-likelihood function of garch models with Student t innovations", mimeo, CEMFI.

Dridi, R. (2000): "Simulated asymptotic least squares theory", mimeo, LSE.

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

Godfrey, L.G. (1988): Misspeci…cation tests in econometrics: The Lagrange multiplier principle and other approaches, Cambridge University Press.

Gourieroux, C. and Monfort, A. (1995): Statistics and econometric models, vol. I and II, Cambridge University Press.

Gourieroux, C. and Monfort, A. (1996): Simulation-based econometric methods, Oxford University Press.

Gourieroux C., Holly A. and Monfort A. (1980): "Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Harvard Institute of Economic Research Discussion Paper 770.

Gourieroux C., Holly A. and Monfort A. (1982): "Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters", Econometrica 50, 63-80.

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

Gourieroux, C., Monfort A. and Trognon, A. (1984): "Pseudo maximum likelihood methods: Theory", Econometrica 52, 681-700.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, 1029-1054.

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

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994): "Bayesian analysis of stochastic volatility models" (with discussion), Journal of Business and Economic Statistics 12, 371-417.

de Jong, R.M. and Davidson, J. (2000): "Consistency of kernel estimators of heteroskedasticity and autocorrelation covariance matrices", Econometrica 68, 407-423.

Kim, S., Shephard, N. and Chib, S. (1998): "Stochastic volatility: likelihood inference and comparison with arch models", Review of Economic Studies 65, 361-393.

Lumsdaine, R.L. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch(1,1) and covariance stationary garch(1,1) models", Econometrica 64, 575-596.

Nelson, D.B. and Cao, C.Q. (1992): "Inequality constraints in the univariate garch model", Journal of Business and Economic Statistics 10, 229-235.

Sargan, J.D. and Bhargava, A. (1983): "Maximum likelihood estimation of regression models with …rst order moving average errors when the root lies on the unit circle", Econometrica 51, 799-820.

Shephard, N. (1996). "Statistical aspects of arch and stochastic volatility". In D.R. Cox, D.V. Hinkley and O.E. Barndor¤-Nielsen (eds.), Time series models with econometric, …nance and other applications, pp. 1-67. Chapman and Hall.

Smith, A.A. (1993): "Estimating nonlinear time series models using simulated vector autoregressions", Journal of Applied Econometrics 8, S63-S84.

White, H. (1982): "Maximum likelihood estimation of misspeci…ed models", Econometrica 50, 1-28.