Constrained Indirect Inference Estimation

By

Giorgio Calzolari
Gabriele Fiorentini
and
Enrique Sentana

DISCUSSION PAPER 384

June 2001

FINANCIAL MARKETS GROUP
AN ESRC RESEARCH CENTRE

LONDON SCHOOL OF ECONOMICS

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

ISSN 0956-8549-384

Constrained Indirect Inference Estimation[1]

Giorgio Calzolari
(Università di Firenze)

Gabriele Fiorentini
(University of Alicante)

Enrique Sentana
(CEMFI)

March 1999

Revised: June 2001

[1]We are grateful to Filippo Altissimo, Torben Andersen, Manuel Arellano, Jon Danielsson, Ramdan Dridi, Russell Davidson, Frank Diebold, Jean-Marie Dufour, Christian Gourieroux, Lars Hansen, Jim Heckman, Joel Horowitz, Hidehiko Ichimura, Alain Monfort, Nour Meddahi, Adrian Pagan, Hashem Pesaran, Steve Satchell and Neil Shephard, as well as seminar audiences at Cambridge, CEMFI, Chicago, Ente Einaudi, Federico II (Naples), LSE, Montreal, UCL and the 1999 European Meeting of the Econometric Society for very helpful comments and suggestions. Special thanks are due to Eric Renault and George Tauchen for their 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. (email addresses: <calzolar@ds.unifi.it>, <gafi@merlin.fae.ua.es>, <sentana@cemfi.es>).

Abstract

We develop generalised indirect inference procedures that handle equality and inequality constraints on the auxiliary model parameters. We also show that the asymptotic efficiency of such estimators can never decrease by explicitly taking into account Lagrange multipliers associated with additional equality constraints, regardless of whether the restrictions are correct. Furthermore, we discuss the variety of effects on efficiency that can result from imposing some constraints on the parameters of a previously unrestricted model. As examples, we consider MA(1) estimated through AR(1), AR(1) through MA(1), and stochastic volatility through GARCH with Gaussian or t distributed errors.

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

JEL: C13, C15

1 Introduction

Consider a stochastic process, y_t, characterised by the sequence of parametric conditional densities $f(y_t|\mathbf{X}_{t-1};\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ denotes the p parameters of interest, and $\mathbf{X}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$. Consider also a possibly misspecified, auxiliary model, described by the sequence of conditional densities $\phi(y_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a q dimensional vector of parameters, with $p \leq q$. In those situations in which no closed-form expression for $f(y_t|\mathbf{X}_{t-1};\boldsymbol{\rho})$ exists, but at the same time it is easy to estimate $\boldsymbol{\theta}$, or to compute expectations of possibly nonlinear functions of y_t, either analytically, or by simulation or quadrature, the indirect inference (II) procedures of Gallant and Tauchen (1996) (GT), Gourieroux, Monfort and Renault (1993) (GMR) and Smith (1993) provide convenient estimation methods, which have made a substantial impact on the practice of econometrics over recent years. Specifically, the II procedure of GMR uses the pseudo-maximum likelihood (ML) estimators of $\boldsymbol{\theta}$ as sample statistics on which to base a classical minimum distance (CMD) estimator of $\boldsymbol{\rho}$. In contrast, the procedure proposed by GT derives a generalised method of moments (GMM) estimator of the parameters of interest on the basis of the score of the auxiliary model evaluated at the pseudo-ML estimators. Under certain conditions, both procedures lead to asymptotically normal estimators of the structural parameters $\boldsymbol{\rho}$, which, in fact, can be made equivalent by an appropriate choice of the CMD and GMM weighting matrices (see e.g. Proposition 4.3 in Gourieroux and Monfort, 1996) (GM96).

One of those conditions, though, is that the parameters of the auxiliary model are unrestricted, and consequently, that their pseudo-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). The first contribution of our paper is to show how II procedures can be generalised to handle equality and/or inequality restrictions on $\boldsymbol{\theta}$. In particular, we propose an alternative set of moment restrictions based on the first order conditions for (in)equality restricted models, which nest the ones employed by GT when there are no constraints, or when they 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 addition, we combine the "constrained" parameter estimators and Lagrange/Kuhn-Tucker multipliers to extend the original class of CMD II estimators of GMR to the possibly restricted case. We also prove that like in the unconstrained case, we can find "restricted" CMD II estimators that are asymptotically equivalent to the GMM estimators by an appropriate choice of weighting matrix. And although we concentrate for expositional purposes on pseudo-ML estimation of the auxiliary model under the assumption that the form of the density function is time-invariant, and $\square_t$ strictly stationary and ergodic, our procedures can be extended to cover any other extremum estimators of just identified auxiliary models under more general circumstances, such as M-estimators or method of moments (see section 4.1.3 of GM96). For analogous reasons, we deliberately separate the results directly related to our proposed modification of the existing II procedures from the way one would conduct numerical simulation in practice. Nevertheless, since very often one has to resort to simulation to implement II procedures, we include an appendix in which some relevant issues are discussed.

There are at least two important reasons for taking into account some inequality restrictions in the estimation of the auxiliary model in actual empirical applications. The first, and most obvious one, is that the pseudo log-likelihood function may not be well defined when certain parameter restrictions are violated, as would be the case when dealing with (transition) probabilities, (un)conditional variance/covariance structures, or some non-Gaussian distributions (see e.g. the examples in section 8.2 of GMR and section 4.1 of GT). In other cases, though, the log-likelihood function can always be computed, but some of the auxiliary parameters may become underidentified in certain regions of the auxiliary parameter space, so that we may decide to restrict it to avoid such discontinuities (see section 3.3 below, or Calzolari, Fiorentini and Sentana (2001) for examples in which both situations concur). In either case, the resulting parameter restrictions are often binding in practice.

As for the relevance of equality constraints, one just needs to realise that any parametric auxiliary model implicitly contains a vast number of maintained assumptions, which can often be written in terms of zero restrictions on some additional parameters, as the extensive literature on Lagrange multiplier-based specification testing shows. Furthermore, equality restricted procedures may be particularly useful from a computational point of view, because in many situations of significant empirical interest, it is considerably simpler to estimate a special restricted case of the auxiliary model than to maximise the unrestricted log-likelihood function.

In this context, our second contribution is an extensive discussion of the effects of the introduction of constraints on the auxiliary model parameters, and of the way we take them into account, on the efficiency of the estimators of the parameters of interest. In this respect, we show that the asymptotic efficiency of II estimators can never decrease by explicitly taking into account the Lagrange multipliers associated with the implicit zero constraints mentioned in the previous paragraph. Importantly, though, such a result in no way requires that the restrictions are correct. Thus, from a practical point of view, our result suggests a computationally very simple way to improve the efficiency of existing II estimators, which can be particularly useful when the informational content of the original auxiliary parameters about the structural parameters appears to be poor.

In addition, we illustrate the variety of effects that can be obtained when some constraints are imposed on the parameters of a previously unrestricted auxiliary model. For instance, we discuss several circumstances in which the imposition of constraints has no effect on the efficiency of the resulting II estimators, and others in which false constraints enable the restricted II estimators to achieve full efficiency. The reason for such counterintuitive results is that by adding restrictions to the auxiliary model in those circumstances in which they are not required to properly define the auxiliary objective function, we are implicitly changing the auxiliary model, and thereby, the binding functions.

For illustrative purposes, we apply our modified procedures to three time series models. The first two are (i) an MA(1) process estimated as an AR(1), possibly with an arbitrary (in)equality constraint on the autoregressive coefficient; and (ii) an AR(1) process estimated as an MA(1), possibly with a zero or non-positivity constraint on the moving average coefficient. Apart from helping guide intuition, the main role of these two examples is to illustrate the range of situations that can occur. Specifically, we show that the imposition of constraints has no effect on the efficiency of the II estimators in the case of the MA(1) process estimated as an AR(1), while it allows us to achieve full efficiency in the case of the AR(1) process estimated as an MA(1). 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 rest of the paper is organized as follows. In section 2, we include a thorough discussion of "restricted" II procedures, and of the efficiency consequences of the constraints. Detailed applications of such procedures to the three aforementioned examples can be found in section 3. Finally, our conclusions are presented in section 4. Proofs and auxiliary results are gathered in the appendix. But before deriving our main results, it may be useful to have a quick overview of II procedures and ML estimation subject to constraints.

A very simple example Let $\{y_t\}$ denote a strictly stationary and ergodic stochastic process whose data generating mechanism is $y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where $-\infty < \mu < \infty$, $\sum_{j=0}^{\infty} |a_j| < \infty$, and $\{\varepsilon_t\}$ is a sequence of $i.i.d.$ random variables with zero mean and unit variance, and suppose that we decide to use the auxiliary model $y_t \sim i.i.d.\ N(\theta, 1)$, with $-\infty < \theta < \infty$, to estimate μ. Since the (minus scaled) pseudo-log likelihood function for a sample of size T on y_t (ignoring constants) will be given by $T^{-1} \sum_{t=1}^{T} (y_t - \theta)^2$, the unconstrained pseudo-ML estimator of θ, $\tilde{\theta}_T^u$ say, will simply be the sample mean $\bar{y}_T$. For each value of μ, we can define the unrestricted binding function $\theta^u(\mu)$ as the value of θ that solves the analogue population programme $\min_\theta E\left[(y_t - \theta)^2 | \mu\right]$, where the symbol $E(\cdot|\mu)$ refers to an expected value computed with respect to the true distribution of y_t evaluated at μ. Hence, it is clear that $\theta^u(\mu) = E(y_t|\mu) = \mu$, which is the plim of $\tilde{\theta}_T^u$. In this respect, note that $\theta^u(\mu)$ satisfies the population first order conditions $E\{s; [\theta^u(\mu), 0]\} = 0$, where $E[s; (\theta, 0)] = E(y_t - \theta | \mu) = \mu - \theta$ is the expected value of the (scaled) score of the auxiliary model. In this context, an *unconstrained* CMD II estimator of μ, $\hat{\mu}_T^u$ say, will be such that it minimises the distance between the binding function $\theta^u(\mu)$ and its sample counterpart $\tilde{\theta}_T^u$ in some chosen metric. Similarly, an *unconstrained* GMM II estimator of μ, $\check{\mu}_T^u$

say, will be such that it minimises some norm of the sample moment conditions $E\left[s; (\hat{\pi}_T^u, 0)\right] = \theta - \bar{y}_T$. But since θ is an unrestricted scalar parameter, it follows that $\hat{\theta}_T^u = \tilde{\theta}_T^u = \bar{y}_T$ regardless of the metric chosen. Given our assumptions about y_t, it is then straightforward to show that $\sqrt{T}(\hat{\theta}_T^u - \theta^0) \xrightarrow{d} N\,[0, S(1)]$, where $S(L)$ is the autocovariance generating function of $\{y_t\}$, and θ^0 the true value of θ.

Suppose now that we decide to modify the auxiliary model by adding the constraint $\pi = \pi^e$. At first sight, it might seem that such an equality restricted auxiliary model no longer has any identification information about θ because the equality constrained pseudo-ML estimator of π is trivially $\hat{\pi}_T^e = \pi^e$. More formally, the equality restricted binding function would be $\pi^e(\theta) = \pi^e$, which does not depend on θ. But such a pessimistic conclusion would be ignoring the information in the Lagrange multiplier of the constraint, $\hat{\gamma}_T^e$ say, which is another sample statistic associated with the constrained model. In this respect, it is easy to see from the first order conditions of the equality restricted model that $\hat{\gamma}_T^e = \pi^e - \bar{y}_T$. By considering the analogous population programme, we can then define the binding function $\gamma^e(\theta) = \pi^e - \theta$ as the theoretical value of the Lagrange multiplier associated with the constraint $\pi = \pi^e$ for each value of θ. Not surprisingly, if $E\,[s; (\pi, \gamma)] = E\,(y_t - \pi + \gamma|\,\theta) = \theta - \pi + \gamma$ denotes the expected value of the first order derivatives of the (scaled) Lagrangian function with respect to π, it is clear that $\gamma^e(\theta)$ satisfies the population first order conditions $E\,\{s; [\pi^e(\theta), \gamma^e(\theta)]\} = 0$. On this basis, we can define an *equality constrained* CMD II estimator of θ, $\hat{\theta}_T^e$ say, as the value that minimises the distance between the binding function $\gamma^e(\theta)$ and its sample counterpart $\hat{\gamma}_T^e$ in some chosen metric. Similarly, an *equality constrained* GMM II estimator of θ, $\tilde{\theta}_T^e$ say, will be such that it minimises the norm of the sample moment conditions $E\left[s; (\hat{\pi}_T^e, \hat{\gamma}_T^e)\right] = \theta - \bar{y}_T$. Somewhat remarkably, it turns out that $\hat{\theta}_T^e = \tilde{\theta}_T^e = \bar{y}_T$ regardless of the value of π^e.

Finally, suppose that we decide to modify yet again the auxiliary model by replacing the equality constraint $\pi = \pi^e$ with the inequality constraint $\pi \geq \pi^e$. In this case, the inequality restricted pseudo-ML estimator of π, $\tilde{\pi}_T^i$ say, will coincide with the unrestricted estimator $\tilde{\pi}_T^u$ if $\tilde{\pi}_T^u \geq \pi^e$, but will be π^e otherwise. Correspondingly, the Kuhn-Tucker multiplier associated with the constraint $\pi \geq \pi^e$, $\tilde{\lambda}_T^i$ say, will be zero if $\tilde{\pi}_T^u \geq \pi^e$, but will coincide with the Lagrange multiplier $\tilde{\lambda}_T^e$ when the inequality constraint is violated by $\tilde{\pi}_T^u$. If we solve the analogous population program, we can define the inequality restricted binding functions $\pi^i(\theta) = \pi^e + (\theta - \pi^e)1(\theta \geq \pi^e)$ and $\lambda^i(\theta) = (\pi^e - \theta)\,1(\theta < \pi^e)$, where $1(\cdot)$ is the usual indicator function. As expected, these binding functions satisfy the population first-order Kuhn-Tucker conditions $E\left\{\theta; \left[\pi^i(\theta), \lambda^i(\theta)\right]\right\} = 0$, the complementary slackness restriction $\left[\pi^i(\theta) - \pi^e\right] \lambda^i(\theta) = 0$, and the sign restrictions $\pi^i(\theta) - \pi^e \geq 0$ and $\lambda^i(\theta) \geq 0$. In this context, we can define an *inequality constrained* CMD II estimator of θ, $\tilde{\theta}_T^i$ say, as the value that minimises the distance between the vector of binding functions $\left[\pi^i(\theta), \lambda^i(\theta)\right]' = \rho^i(\theta)$ and their sample counterparts $\left(\tilde{\pi}_T^i, \tilde{\lambda}_T^i\right)' = \tilde{\rho}_T^i$ for some chosen metric. It turns out that the choice of metric is once more irrelevant in this case because $\tilde{\theta}_T^i = \bar{y}_T$ will make the distance between $\rho^i(\theta)$ and $\tilde{\rho}_T^i$ equal to 0 regardless of whether $\bar{y}_T$ exceeds π^e or not. Similarly, an *inequality constrained* GMM II estimator of θ, $\check{\theta}_T^i$ say, can be defined so that it minimises the norm of the sample moment conditions $E\left[\theta; (\tilde{\pi}_T^i, \tilde{\lambda}_T^i)\right] = \theta - \bar{y}_T$. Therefore, we have once again that $\tilde{\theta}_T^i = \check{\theta}_T^i = \bar{y}_T$ irrespective of the value of π^e.

Although this example is extremely simple, it illustrates the three main points of our paper:

1. The statistical properties of II estimators are not any more difficult to obtain with constraints on the auxiliary model parameters than without them. In fact, from a formal point of view, the inequality restricted problem is not different from the equality restricted or indeed the unrestricted case, since in the latter case we can always define a multiplier $\tilde{\lambda}_T^u$ as being identically

0, so that the corresponding binding function $\square^u(\square) = 0$ is completely uninformative about the parameter of interest $\square$. Therefore, in all three cases we would be matching either the distance between the vector of binding functions $\square^r(\square)$ and their sample counterparts, $\bar{\square}^r_T$, or the norm of the sample moment conditions $\square\left[\square; (\bar{\square}^r_T \square \bar{\square}^r_T)\right]$, with $\square = \square\square\square\square$. In this respect, it is important to mention that the singularity and possible non-normality of the joint distribution of $\bar{\square}^r_T$ does not create any particular problems, because (a) the singularity is confined to the complementary slackness condition $\left(\bar{\square}^r_T - \square^e\right)\bar{\square}^r_T = 0$, which is completely uninformative about the true value of $\square$; and (b) the linear combination $\square - \bar{\square}^r_T + \bar{\square}^r_T$, which coincides with the sample moment conditions, is always asymptotically well behaved.

2. Relative to the standard practice of ignoring them, our proposed use of Lagrange and/or Kuhn-Tucker multipliers to take into account constraints on the auxiliary parameters has an unequivocal non-negative effect on the asymptotic efficiency of the resulting indirect inference estimators of the parameters of interest. In this simple example, in particular, the relative efficiency gain is infinite, since the equality restricted estimator of $\square$ does not even enable the identification of $\square$.

3. Once we take the multipliers associated with equality and/or inequality constraints into account in our proposed way, though, the efficiency implications of the precise form of the constraints is unclear. There may be situations, like the previous example, in which the imposition of constraints has no effect on the efficiency of the indirect inference estimators of $\square$, and others, in which a restricted II procedure achieves full efficiency.

2 Theoretical set up

2.1 "Restricted" II estimators

Let $l_t(\boldsymbol{\theta}) = \ln f(y_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q$, denote the density function of a possibly misspecified auxiliary model, and assume for simplicity of exposition that its functional form is time-invariant, and that y_t is strictly stationary and ergodic. The pseudo log-likelihood function for a sample of size T on y_t based on the auxiliary model (ignoring initial conditions) will therefore be given by $L_T(\boldsymbol{\theta}) = \sum_t l_t(\boldsymbol{\theta})$. Let us now define the (scaled) Lagrangian function

$$\mathcal{L}_T(\boldsymbol{\beta}) = \frac{1}{T}L_T(\boldsymbol{\theta}) + \mathbf{r}'(\boldsymbol{\theta})\boldsymbol{\mu} \tag{1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\theta}', \boldsymbol{\mu}')'$, and $\boldsymbol{\mu}$ are the r "multipliers" associated with the r constraints implicitly characterised by the vector of functions $\mathbf{r}(\boldsymbol{\theta})$, which effectively force $\boldsymbol{\theta}$ to lie in a non-empty "restricted" parameter space $\boldsymbol{\Theta}^r \subseteq \boldsymbol{\Theta}$. Such a set up is sufficiently general to cover most cases of practical interest, including a mix of equality and inequality constraints. For the sake of clarity, though, we concentrate on the three archetypal situations of (a) unconstrained estimation, (b) equality constraints, and (c) inequality constraints, which can be characterised as follows:

$$\begin{array}{lcc} \text{(a)} & \mathbf{r}(\boldsymbol{\theta}) \text{ unrestricted} & \boldsymbol{\mu} = \mathbf{0} \\ \text{(b)} & \mathbf{r}(\boldsymbol{\theta}) = \mathbf{0} & \boldsymbol{\mu} \text{ unrestricted} \\ \text{(c)} & \mathbf{r}(\boldsymbol{\theta}) \geq \mathbf{0} & \boldsymbol{\mu} \geq \mathbf{0} \end{array} \tag{2}$$

Assuming that both the pseudo-log likelihood function $L_T(\boldsymbol{\theta})$, and the vector of functions $\mathbf{r}(\boldsymbol{\theta})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$, the latter with a Jacobian matrix $\partial\mathbf{r}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ whose rank coincides with the number of effective constraints at $\boldsymbol{\theta}$, the first-order conditions that take into account the "constraints" will be given by:

$$\frac{\partial \mathcal{L}_T(\bar{\boldsymbol{\beta}}_T^r)}{\partial\boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{s}_t(\bar{\boldsymbol{\beta}}_T^r) = \mathbf{0}, \tag{3}$$

together with the complementary slackness restrictions:

$$h\left(\bar{\boldsymbol{\theta}}_T^r\right) \odot \bar{\boldsymbol{\mu}}_T^r = \mathbf{0} \tag{4}$$

plus the appropriate (in)equality restrictions on $h\left(\bar{\boldsymbol{\theta}}_T^r\right)$ and/or $\bar{\boldsymbol{\mu}}_T^r$ in (2), where

$$s_t(\boldsymbol{\beta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\boldsymbol{\mu} \tag{5}$$

is the contribution of the t^{th} observation to the modified score of the auxiliary model, $-$ indicates pseudo-ML estimators, the superscript $r = (u, e, i)$ stands for unrestricted, equality restricted and inequality restricted respectively, 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. Note that the main difference with the usual unrestricted case is that $s_t(\boldsymbol{\beta})$ not only depends on the p auxiliary model parameters $\boldsymbol{\theta}$, but also on the q multipliers $\boldsymbol{\mu}$ associated with the restrictions. In fact, given our assumptions, we can obtain from (3) the following explicit expression for $\bar{\boldsymbol{\mu}}_T^r$ in terms of $\bar{\boldsymbol{\theta}}_T^r$:

$$\bar{\boldsymbol{\mu}}_T^r = -\left[\frac{\partial h(\bar{\boldsymbol{\theta}}_T^r)}{\partial \boldsymbol{\theta}'}\boldsymbol{\Upsilon}\frac{\partial h'(\bar{\boldsymbol{\theta}}_T^r)}{\partial \boldsymbol{\theta}}\right]^{-1}\frac{\partial h(\bar{\boldsymbol{\theta}}_T^r)}{\partial \boldsymbol{\theta}'}\boldsymbol{\Upsilon}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\bar{\boldsymbol{\theta}}_T^r)}{\partial \boldsymbol{\theta}}, \tag{6}$$

where $\boldsymbol{\Upsilon}$ is any positive definite (p.d.) matrix of order p, which confirms that $\bar{\boldsymbol{\mu}}_T^r$ will be trivially zero in the unrestricted case.

Let us now define

$$\mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta}) = E\left[l_t(\boldsymbol{\theta})|\,\boldsymbol{\rho}\right] \tag{7}$$

where $E(\cdot|\boldsymbol{\rho})$ refers to an expected value computed with respect to the distribution of the model of interest evaluated at $\boldsymbol{\rho}$, and assume that $\mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta})$ is twice continuously differentiable with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$. For each value of $\boldsymbol{\rho}$, we can define the binding functions for the "constrained" auxiliary parameters $\boldsymbol{\theta}$ and the associated "multipliers" $\boldsymbol{\mu}$, $\boldsymbol{\beta}^r(\boldsymbol{\rho}) = [\boldsymbol{\theta}^{r\prime}(\boldsymbol{\rho}), \boldsymbol{\mu}^{r\prime}(\boldsymbol{\rho})]'$ say, as the values of

$\boldsymbol{\beta}$ associated with the maximum over the restricted parameter space $\boldsymbol{\Theta}^r$ of the (population) Lagrangian function

$$\mathcal{Q}(\boldsymbol{\rho};\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta}) + r'(\boldsymbol{\theta})\boldsymbol{\mu}. \tag{8}$$

As a result, if we denote by

$$\mathfrak{m}\left(\boldsymbol{\rho};\boldsymbol{\beta}\right) = \mathbb{E}\left[\, l_t(\boldsymbol{\beta})\right|\boldsymbol{\rho}\right], \tag{9}$$

the binding functions must satisfy the first-order conditions:

$$\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho})\right] = \mathbf{0}, \tag{10}$$

the exclusion restrictions

$$r\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right] \odot \boldsymbol{\mu}^r(\boldsymbol{\rho}) = \mathbf{0}, \tag{11}$$

plus the required (in)equality restrictions on $r\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right]$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho})$ in (2), as long as the differentiation and expectation operators can be interchanged, which we assume henceforth. Under those circumstances, we can obtain an explicit expression for $\boldsymbol{\mu}^r(\boldsymbol{\rho})$ in terms of $\boldsymbol{\theta}^r(\boldsymbol{\rho})$ as

$$\boldsymbol{\mu}^r(\boldsymbol{\rho}) = -\left\{\frac{\partial r\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right]}{\partial\boldsymbol{\theta}'}\boldsymbol{\Upsilon}\frac{\partial r'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right]}{\partial\boldsymbol{\theta}}\right\}^{-1}\frac{\partial r\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right]}{\partial\boldsymbol{\theta}'}\boldsymbol{\Upsilon}\mathbb{E}\left\{\left.\frac{\partial l_t\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right]}{\partial\boldsymbol{\theta}}\right|\boldsymbol{\rho}\right\} \tag{12}$$

Let $\boldsymbol{\rho}_0$ denote the true value of the parameters of interest, and $\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)$ the "constrained" pseudo-true values of $\boldsymbol{\beta}$. To guarantee the local identification of $\boldsymbol{\rho}_0$, we assume that $\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)$ is locally identified, in the sense that $\mathcal{L}\left[\boldsymbol{\rho}_0;\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)\right] > \mathcal{L}\left(\boldsymbol{\rho}_0;\boldsymbol{\theta}\right)$ for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}^r$ in a neighbourhood of $\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)$. We also assume that the systems of equations $\boldsymbol{\beta}^r(\boldsymbol{\rho}) = \boldsymbol{\beta}^r(\boldsymbol{\rho}_0)$ and $\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)\right] = \mathbf{0}$ separately admit the unique solution $\boldsymbol{\rho} = \boldsymbol{\rho}_0$, which obviously requires the order condition $\square \geq \square$ (cf. GM96). If we further assume that both functions are continuously differentiable in $\boldsymbol{\rho}$, a sufficient condition for the identification of $\boldsymbol{\rho}$ is that the Jacobian matrices $\partial\boldsymbol{\beta}^r(\boldsymbol{\rho})/\partial\boldsymbol{\rho}'$ and $\partial\mathfrak{m}(\boldsymbol{\rho};\boldsymbol{\beta})/\partial\boldsymbol{\rho}'$ have full column rank. More formally,

Assumption 1

$$\square\square\square\square \left[\frac{\square\boldsymbol{\beta}^r(\boldsymbol{\rho})}{\square\boldsymbol{\rho}'}\right] = \square$$

$$\square\square\square\square \left\{\frac{\square\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\rho}'}\right\} = \square$$

for any $\boldsymbol{\rho}$ *in a neighbourhood of* $\boldsymbol{\rho}^{\square}$*.*

As usual, such assumptions are rather difficult to check in non-linear models, but they are crucial for the consistency of the II estimators that we discuss. Intuitively, the reason is that when Assumption 1 holds, if we knew $\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})$, we could recover $\boldsymbol{\rho}^{\square}$ by either inverting the binding functions, or solving the possibly non-linear system of equations $\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right] = \mathbf{0}$ with respect to its first argument holding the second argument fixed. 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 $\boldsymbol{\rho}^{\square}$ by choosing the parameter values that minimise either some appropriately defined distance between $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ and $\bar{\boldsymbol{\beta}}^r_T$, or a given norm of the sample moments $\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^r_T)$. In particular, we can minimise with respect to $\boldsymbol{\rho}$ the following quadratic forms:

$$\square^r_T(\boldsymbol{\rho};\boldsymbol{\Omega}) = \left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \bar{\boldsymbol{\beta}}^r_T\right]' \cdot \boldsymbol{\Omega} \cdot \left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \bar{\boldsymbol{\beta}}^r_T\right]$$

or

$$\square^r_T(\boldsymbol{\rho};\boldsymbol{\Psi}) = \mathfrak{m}'(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^r_T) \cdot \boldsymbol{\Psi} \cdot \mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}_T)$$

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

$$\begin{aligned}\hat{\boldsymbol{\rho}}^r_T(\boldsymbol{\Omega}) &= \arg\min_{\boldsymbol{\rho}} \square^r_T(\boldsymbol{\rho};\boldsymbol{\Omega}) \\ \tilde{\boldsymbol{\rho}}^r_T(\boldsymbol{\Psi}) &= \arg\min_{\boldsymbol{\rho}} \square^r_T(\boldsymbol{\rho};\boldsymbol{\Psi})\end{aligned}$$

as the "restricted" CMD and GMM II estimators of $\boldsymbol{\rho}$. Obviously, without a judicious choice of metric that accounts for sample variation in the estimators of the (in)equality restricted auxiliary parameters and/or multipliers in $\bar{\boldsymbol{\beta}}_T^r$, the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}_T^r(\boldsymbol{\Omega})$ and $\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})$ is likely to be unnecessarily large.

Let us start by analysing the second criterion function. It is well known that if the sample moments $\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}_T^r)$ have a limiting normal distribution, the optimal 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 $\sqrt{T}\,\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}_T^r)$ (see e.g. Hansen, 1982). In order to derive the required asymptotic distribution, we assume the necessary conditions for $\bar{\boldsymbol{\beta}}_T^r$ to converge in probability to $\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)$ uniformly over the restricted parameter space, and for a 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 2

$$\lim_{T\to\infty}\Pr\left[\sup_{\boldsymbol{\rho}^r}\left\|\begin{array}{c}\bar{\boldsymbol{\theta}}_T^r-\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\\ \bar{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\end{array}\right\|=0\right] = 1$$

$$\lim_{T\to\infty}\Pr\left\{\left\|\frac{1}{T}\sum_t\frac{\partial^2 l_t^r(\boldsymbol{\theta}_T^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\mathcal{J}_0^r\right\|=0\right\} = 1$$

$$\sqrt{T}\frac{1}{T}\sum_t s_t\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] \to N(\mathbf{0},\mathcal{I}_0^r)$$

where $\mathcal{J}_0^r$ and $\mathcal{I}_0^r$ are non-stochastic, $\square\times\square$ matrices, with $\mathcal{I}_0^r$ p.d., and $\boldsymbol{\theta}_T^*$ is any sequence that converges in probability to $\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)$.

In this respect, it is important to note that relative to the standard unconstrained case, the main effect of adding the constant term $\{\partial \mathbf{r}'[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)]/\partial\boldsymbol{\theta}\}\,\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)$

to the original score $\partial l_t[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)]/\partial\boldsymbol{\theta}$ is to centre around zero the asymptotic distribution of $\square_t[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)]$. Therefore, if $\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)$ is in the interior of the admissible auxiliary parameter space $\boldsymbol{\Theta}$, Assumption 2 is equivalent to the high level assumptions made by GMR and GT. In addition, it should be emphasised that there are many inequality restricted situations in which the pseudo log-likelihood function is not well-defined outside the restricted parameter space, $\boldsymbol{\Theta}^r$, 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 used in section 3.3 under conditional Gaussianity, as discussed in Fiorentini, Sentana and Calzolari (2000)).

Unfortunately, we cannot directly rely on the results in GT to derive the asymptotic distribution of the sample moments $\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}_T^r)$, since the "restricted" estimator $\bar{\boldsymbol{\theta}}_T^r$ may not be asymptotically normal in large samples in the presence of inequality constraints (see Andrews (1999) and the references therein). In addition, the asymptotic distribution of $\bar{\boldsymbol{\beta}}_T^r$ is singular for $\square = (\square\square\square\square\square)$. More specifically:

Proposition 1 *Under Assumption 2,*

$$\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\odot\frac{\partial\square[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)]}{\partial\boldsymbol{\theta}'}\sqrt{T}\left[\bar{\boldsymbol{\theta}}_T^r-\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]+\square\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]\odot\sqrt{T}\left[\bar{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]=o_p(1)$$

Such a singularity is a direct consequence of the fact that the complementary slackness conditions (4) must always be satisfied by $\bar{\boldsymbol{\beta}}_T^r$. Nevertheless, it is important to mention that since their population counterparts (11) will be satisfied for any value of $\boldsymbol{\rho}$, the singular combinations of the auxiliary parameters and multipliers contain no identifying information whatsoever about the parameters of interest.

In contrast, there are $\square$ linear combinations that are asymptotically well behaved:

Proposition 2 *Under Assumption 2,*

$$\left[\mathcal{J}_0^r + \left[\boldsymbol{\mu}^r(\boldsymbol{\rho}_0) \otimes I_q\right] \frac{\partial vecd\left\{\partial \mathbf{h}'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)\right]/\partial\boldsymbol{\theta}\right\}}{\partial\boldsymbol{\theta}'}\right] \sqrt{T}\left[\bar{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}_0)\right]$$

$$+\frac{\partial \mathbf{h}'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)\right]}{\partial\boldsymbol{\theta}}\sqrt{T}\left[\bar{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}_0)\right] + \sqrt{T}\frac{1}{T}\sum_t \mathbf{s}_t\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)\right] = o_p(1)$$

Hence, even though $\bar{\boldsymbol{\theta}}_T^r$ and $\bar{\boldsymbol{\mu}}_T^r$ have a singular and possibly non-Gaussian asymptotic distribution, Proposition 2 shows that under our regularity conditions, there are always p linear combinations that are asymptotically normally distributed, irrespective of the exact nature of the restrictions, and irrespective of whether the restrictions on $\mathbf{h}\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)\right]$ and $\boldsymbol{\mu}^r(\boldsymbol{\rho}_0)$ are satisfied with equality, or strict inequality. It turns out that those p linear combinations are implicitly contained in the expected value of the modified score:

Proposition 3 *Under Assumption 2,*

$$\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}_0;\bar{\boldsymbol{\beta}}_T^r)+\sqrt{T}\frac{1}{T}\sum_t \mathbf{s}_t\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)\right] = o_p(1)$$

Therefore, $\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}_0;\bar{\boldsymbol{\beta}}_T^r)$ has indeed a limiting Gaussian distribution, and the optimal weighting matrix is precisely the inverse of $\mathcal{I}_0^r$.

The following proposition specifies the asymptotic distribution of the (infeasible) optimal GMM estimator of $\boldsymbol{\rho}$ based on the "restricted" auxiliary model:

Proposition 4 *Under Assumptions 1 and 2*

$$\sqrt{T}\left[\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right] - \boldsymbol{\rho}_0\right] \rightarrow N\left[\mathbf{0},\left\{\frac{\partial\mathfrak{m}'\left[\boldsymbol{\rho}_0;\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)\right]}{\partial\boldsymbol{\rho}}\cdot(\mathcal{I}_0^r)^{-1}\cdot\frac{\partial\mathfrak{m}\left[\boldsymbol{\rho}_0;\boldsymbol{\beta}^r(\boldsymbol{\rho}_0)\right]}{\partial\boldsymbol{\rho}'}\right\}^{-1}\right]$$

Given that this expression is completely analogous to the one derived by GT for their GMM version of the II estimator in the absence of constraints, the required matrices can also be consistently estimated using their suggested procedures. In particular, since under our assumptions

$$E\left\{ \mathbf{h}_t\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] \middle| \boldsymbol{\rho}^0 \right\} = \mathbf{0} \quad \forall t,$$

the time-invariant functional form of $\mathbf{h}_t[\boldsymbol{\beta}]$, and the strict stationarity and ergodicity of $\mathbf{y}_t$ imply that

$$\mathcal{I}_0^r = \sum_{\tau=-\infty}^{\infty} \mathbf{G}_\tau\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right],$$

where

$$\mathbf{G}_\tau(\boldsymbol{\rho};\boldsymbol{\beta}) = E\left\{ \mathbf{h}_t(\boldsymbol{\beta})\mathbf{h}'_{t-\tau}(\boldsymbol{\beta}) \middle| \boldsymbol{\rho} \right\}$$

for $\tau \geq 0$ and $\mathbf{G}_\tau(\boldsymbol{\rho};\boldsymbol{\beta}) = \mathbf{G}'_{-\tau}(\boldsymbol{\rho};\boldsymbol{\beta})$ for $\tau < 0$, provided that the autocovariance matrices are absolutely summable (see e.g. Hansen, 1982). Therefore, we could obtain a consistent estimator of the matrix $\mathcal{I}_0^r$ as

$$\overline{\mathcal{I}}_T^r = \sum_{\tau=-T^\iota}^{T^\iota} \omega(\tau)\bar{\mathbf{G}}_{\tau T}^r \tag{13}$$

with

$$\bar{\mathbf{G}}_{\tau T}^r = \frac{1}{T}\sum_{t=\tau+1}^{T} \mathbf{h}_t(\bar{\boldsymbol{\beta}}_T^r)\mathbf{h}'_{t-\tau}(\bar{\boldsymbol{\beta}}_T^r)$$

where $\omega(\tau)$ 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 $\tilde{\boldsymbol{\rho}}_T^r\left[(\overline{\mathcal{I}}_T^r)^{-1}\right]$. Alternatively, we could consider continuously updated GMM estimators à la Hansen, Heaton and Yaaron (1996), by replacing $\bar{\mathbf{G}}_{\tau T}^r$ in the above expressions with $\mathbf{G}_\tau(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}_T^r)$.

Another important implication of Proposition 4 is that the usual overidentifying restriction test

$$T\cdot J_T^r\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right];(\mathcal{I}_0^r)^{-1}\right\} = T\cdot\mathfrak{m}'\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right];\bar{\boldsymbol{\beta}}_T^r\right\}\cdot(\mathcal{I}_0^r)^{-1}\cdot\mathfrak{m}\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right];\bar{\boldsymbol{\beta}}_T\right\}$$

converges to a $\square^{\square}$ distribution with $\square - \square$ degrees of freedom as $\square \to \infty$, and hence it can be used in the standard manner to assess the adequacy of the model of interest to the data.

Let us now turn to the II estimators of $\boldsymbol{\rho}$ based on the CMD function $\square_T^r(\boldsymbol{\rho}; \boldsymbol{\Omega})$. Unfortunately, we cannot directly rely on standard CMD theory, because as we saw before, the limiting distribution of $\sqrt{\square}\left[\bar{\boldsymbol{\beta}}_T^r - \boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]$ 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 $\square + \square$:

$$
\begin{aligned}
\mathcal{K}_{\square}^r &= \left[\begin{array}{cc} \mathcal{J}_{\square}^r + [\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square}) \otimes \square_q]\,\square\square\square\square\,\{\square\square'\,[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})]\,\square\square\boldsymbol{\theta}\}\,\square\square\boldsymbol{\theta}' & \square\square'\,[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})]\,\square\square\boldsymbol{\theta} \\ \square\square\square\square\,[\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square})]\,\square\square\,[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})]\,\square\square\boldsymbol{\theta}' & \square\square\square\square\,\{\square\,[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})]\} \end{array}\right] \\
&= \left[\begin{array}{cc} \mathcal{K}_{\square\square,\square}^r & \mathcal{K}_{\square\square,\square}^r \\ \mathcal{K}_{\square\square,\square}^r & \mathcal{K}_{\square\square,\square}^r \end{array}\right]
\end{aligned}
$$

where $\square\square\square\square(\square)$ 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 CMD conditions by premultiplying them by $\mathcal{K}_{\square}^r$, we will have that the asymptotic distribution of $\sqrt{\square}\mathcal{K}_{\square}^r\left[\bar{\boldsymbol{\beta}}_T^r - \boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]$ will be normal, with the singularity confined to the last $\square$ elements. In this framework, we can prove the following result, which can be regarded as a generalisation of Proposition 4.3 in GM96, which in turn formalises earlier results in GMR:

Proposition 5 *Under Assumptions 1 and 2*

$$
\sqrt{\square}\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi}) - \hat{\boldsymbol{\rho}}_T^r(\mathcal{K}_{\square}^{r\prime}\boldsymbol{\Psi}^{\boxplus}\mathcal{K}_{\square}^r)\right] = \square_p(1)
$$

where

$$
\boldsymbol{\Psi}^{\boxplus} = \begin{pmatrix} \boldsymbol{\Psi} & 0 \\ 0 & 0 \end{pmatrix}
$$

Apart from the computational advantages highlighted by GT, which we discuss in the appendix, the GMM procedure has the additional advantage that the optimal weighting matrix can be readily computed as the variance of the limiting normal distribution of the modified score (9), irrespective of the exact nature of the inequality restrictions, and irrespective of whether the restrictions on $\square[\boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)]$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho}^\square)$ in (5) are satisfied as equalities, or strict inequalities. Nevertheless, there are some cases of practical relevance in which $\hat{\boldsymbol{\rho}}_T^r \left[\mathcal{K}_\square^{r\prime}\left(\mathcal{I}_\square^{\boxplus}\right)^\square \mathcal{K}_\square^r\right]$ is relatively easy to compute, where + denotes the Moore-Penrose generalised inverse. For instance, suppose that all □ restrictions are of the simple form $\square_j = \square_j^\dagger$ for $\square = 1\square\square\square\square\square \leq \square$. Then, it is easy to see from Proposition 2 that the $\square \times 1$ vector $(\square_\square^e \square\square\square\square\square \square_s^e \square \bar{\square}_{s\square\square}^e \square\square\square\square\square \bar{\square}_q^e)$ will have an asymptotically normal distribution, with a full rank covariance matrix, which can be used to compute the “optimal” equality constrained CMD II estimator of $\boldsymbol{\rho}$.

In addition, there is one instance in which our proposed CMD and GMM procedures yield numerically identical estimators of $\boldsymbol{\rho}$, as in Proposition 4.1 in GM96:

Proposition 6 *If $\square = \square$, so that the auxiliary model exactly identifies the parameters of interest, then $\hat{\boldsymbol{\rho}}_T^r(\boldsymbol{\Omega}) = \tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})$ for large enough $\square$.*

For instance, suppose that all the restrictions are of the simple “bounds” form, i.e. $\square_{j\,\square\square\square} \leq \square_j \leq \square_{j\,\square\square\square}$ $(\square = 1\square\square\square\square\square\square)$, with $|\square_{j\,\square\square\square}|\square|\square_{j\,\square\square\square}|$ possibly infinity, and define $\square_{j\,\square\square\square}^i \square \square_{j\,\square\square\square}^i$ as the matching pair of Kuhn-Tucker multipliers (which are set to zero by definition if the corresponding bound is $\pm\infty$, or to $\pm$ the corresponding Lagrange multiplier if $\square_{j\,\square\square\square} = \square_{j\,\square\square\square}$). Then, the value of $\boldsymbol{\rho}$ that for $\square = 1\square\square\square\square\square\square$ produced estimates of the triplets $\left[\square_j^i(\boldsymbol{\rho})\square \square_{q\,\square\square\square}(\boldsymbol{\rho})\square \square_{q\,\square\square\square}(\boldsymbol{\rho})\right]$ that are equal to (i) $(\bar{\square}_{jT}^i \square 0 \square 0)$ if $\square_{j\,\square\square\square} \square \bar{\square}_{jT}^i \square \square_{j\,\square\square\square}$, (ii) $(\square_{j\,\square\square\square} \square \bar{\square}_{q\,\square\square\square}^i, 0)$ if $\bar{\square}_{jT}^i = \square_{j\,\square\square\square}$, or (iii) $(\square_{j\,\square\square\square} \square 0 \square \bar{\square}_{q\,\square\square\square}^i)$ if $\bar{\square}_{jT}^i = \square_{j\,\square\square\square}$, would also set to zero the sample moments $\mathfrak{m}(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}_T^i)$, and therefore, would be numerically identical to $\tilde{\boldsymbol{\rho}}_T^i(\boldsymbol{\Psi})$ for all $\boldsymbol{\Psi}$.

Finally, given that both GMM and CMD can be regarded as particular cases of minimum chi-square methods (see e.g. Newey and McFadden (1994)), an attractive way of interpreting all our previous results is to think of the population moments $\mathfrak{m}(\boldsymbol{\rho};\boldsymbol{\beta})$ as a set of $\square$ new auxiliary parameters, which summarise all the information in the original parameters $\boldsymbol{\theta}$ and multipliers $\boldsymbol{\mu}$ that is useful for estimating $\boldsymbol{\rho}$. In this light, Proposition 4 simply says that the precision with which we can estimate $\boldsymbol{\rho}$ depends exclusively on (i) the precision that can be achieved in estimating those new parameters, which is given by the inverse of the covariance matrix of the modified sample score, $(\mathcal{I}_{\square}^{r})^{-\square}$, and (ii) the identification content of the same parameters, as measured by the Jacobian of the population moments with respect to its first argument, $\square\mathfrak{m}[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^{r}(\boldsymbol{\rho}^{\square})]\square\square\boldsymbol{\rho}'$.

2.2 Efficiency considerations

2.2.1 Implications of the way in which constraints are taken into account

Equality restricted 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. For instance, the estimation of a VAR(p) model is much easier than the estimation of any VARMA(p,q) model that nests it.

Nevertheless, it may seem at first sight that we could equivalently handle equality restrictions on the auxiliary parameters of the form $\square(\boldsymbol{\theta}) = \mathbf{0}$ with the existing unrestricted II procedures by re-writing the constraints in explicit form, and considering only the information in those parameters that were effectively unconstrained. More specifically, consider a homeomorphic (i.e. one-to-one and bicontinuous) transformation $\square(\square) = [\square_{\square}'(\square)\square\square_{\square}'(\square)]'$ of the auxiliary model parameters $\boldsymbol{\theta}$

into an alternative set of $(\square - \square) + \square$ parameters $\boldsymbol{\pi} = (\boldsymbol{\pi}'_{\square} \square \boldsymbol{\pi}'_{\square})'$, where $\boldsymbol{\pi}_{\square} = \square_{\square}(\boldsymbol{\theta}) = \square(\boldsymbol{\theta})$, and $\square(\boldsymbol{\theta})$ is twice continuously differentiable with rank$[\square\square'(\boldsymbol{\theta})\square\square\boldsymbol{\theta}] = \square$ in a neighbourhood of $\boldsymbol{\theta}^e(\boldsymbol{\rho}^{\square})$. For instance, a simple linear equality constraint of the form $\square_j + \square_k = 0$ can be trivially re-written in terms of $\square_j = \square_j$ and $\square_k = \square_j + \square_k$, with $\square_j$ free and $\square_k = 0$. Let $\bar{\boldsymbol{\pi}}^u_{\square T} = \square_{\square}(\bar{\boldsymbol{\theta}}^e_T)$ denote the unconstrained pseudo-ML estimator of $\boldsymbol{\pi}_{\square}$ obtained by maximising with respect to $\boldsymbol{\pi}_{\square}$ the auxiliary objective function $\square_T(\boldsymbol{\theta})$ reparametrised in terms of $\boldsymbol{\pi}$, with $\boldsymbol{\pi}_{\square}$ set to $\mathbf{0}$. Similarly, let $\boldsymbol{\pi}^u_{\square}(\boldsymbol{\rho}) = \square_{\square}[\boldsymbol{\theta}^e(\boldsymbol{\rho})]$ denote the corresponding binding function. In this context, we could define an alternative CMD unconstrained II estimator of $\boldsymbol{\rho}$, $\breve{\boldsymbol{\rho}}^u_T(\boldsymbol{\Phi})$ say, as the value of $\boldsymbol{\rho}$ that minimises the distance between $\boldsymbol{\pi}^u_{\square}(\boldsymbol{\rho})$ and $\bar{\boldsymbol{\pi}}^u_{\square T}$ in the metric of a p.s.d. matrix $\boldsymbol{\Phi}$ of order $(\square - \square)$. The rationale for such an estimator would be that since $\boldsymbol{\pi}_{\square}$ is set to zero by assumption, there is no information about the true value of $\boldsymbol{\rho}$ in those parameters that do not belong to the active set. Therefore, it is not surprising that $\breve{\boldsymbol{\rho}}^u_T(\boldsymbol{\Phi})$ is the estimator that all existing empirical implementations of II procedures have effectively used in practice.

However, it is very important to emphasise that in doing so, we would be most likely incurring in an efficiency loss relative to our proposed estimation procedure. In this respect, the following result compares the optimal equality constrained CMD II estimator of $\boldsymbol{\rho}$ described in the previous section and $\breve{\boldsymbol{\rho}}^u_T(\boldsymbol{\Phi})$.

Proposition 7 *Under Assumptions 1 and 2,* $\hat{\boldsymbol{\rho}}^e_T\left[\mathcal{K}^{r\prime}_{\square}\left(\mathcal{I}^{\boxplus}_{\square}\right)^{\square}\mathcal{K}^r_{\square}\right]$ *is asymptotically at least as efficient as* $\breve{\boldsymbol{\rho}}^u_T(\boldsymbol{\Phi})$ *for any p.s.d. matrix* $\boldsymbol{\Phi}$

Of course, there may be circumstances in which both estimators are equally efficient - for instance, when $\breve{\boldsymbol{\rho}}^u_T(\boldsymbol{\Phi})$ achieves the asymptotic Cramer-Rao bound. At the same time, there are other circumstances in which $\boldsymbol{\pi}^u_{\square}(\boldsymbol{\rho})$ would not suffice to identify $\boldsymbol{\rho}$, and hence, the relative efficiency gains from taking into account the information in $\bar{\boldsymbol{\mu}}^e_T$ would be infinite. As an extreme case, suppose that $\square = \square \geq \square$, and that $r(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\theta}^{\dagger}$, so that the only admissible value for the equality restricted estimator $\bar{\boldsymbol{\theta}}_T^e$ is precisely $\boldsymbol{\theta}^{\dagger}$, as in our introductory example. In this case, the dimension of $\boldsymbol{\pi}_u$ would be zero, and no unconstrained CMD II estimator based on those inexistent parameters could be defined. In contrast, our equality constrained CMD II procedure will work by simply matching the r equality restricted binding functions $\boldsymbol{\mu}^e(\boldsymbol{\rho})$ with the sample estimates of the r Lagrange multipliers.

Proposition 7 has important consequences for actual practice, since any auxiliary parametric model contains a potentially very large number of implicit constraints. The extensive literature on LM (or efficient score) tests provides many such examples (see e.g. Godfrey (1988) and the references therein). Therefore, given that in practice users of II procedures typically do the reduction on the auxiliary model rather than deal with the modified first order conditions, the scope for improving the efficiency of existing unconstrained II estimators by explicitly taking into account the multipliers associated with those implicit constraints could be significant. We shall investigate this issue with the example in section 3.3.

Finally, it may also seem at first sight that we could handle inequality restrictions on the parameters of the auxiliary model with the existing *unconstrained* 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 Assumptions 1 and 2 are no longer satisfied in terms of the new parameter when the inequality restricted pseudo-true value of the original parameter $\theta_j^i(\boldsymbol{\rho}^0)$ is 0, as the Jacobian of the transformation is 0 at $\theta_j^i(\boldsymbol{\rho}^0) = 0$.

2.2.2 Implications of the nature of the restrictions

If $\boldsymbol{\theta}$ were the parameters of interest, and $f(y_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ provided the correct conditional density function for y_t, the imposition of correct equality restrictions on $\boldsymbol{\theta}$ would weakly improve the efficiency of the resulting estimators (see e.g. Rothemberg (1973) for details). However, such a result is not necessarily robust to misspecification of the density function, even if both $\bar{\boldsymbol{\theta}}_T^u$ and $\bar{\boldsymbol{\theta}}_T^e$ remain consistent for the true value of $\boldsymbol{\theta}$ under misspecification of the pseudo-log likelihood function (see e.g. Arellano (1989) for a counterexample). The situation is even less clear cut in our "constrained" indirect inference set up, in which both the density function of the auxiliary model and the restrictions on $\boldsymbol{\theta}$ and/or $\boldsymbol{\mu}$ may well be incorrect. The root of the problem is that by adding restrictions to the auxiliary model in those circumstances in which they are not required to properly defined the auxiliary objective function, we are implicitly changing the auxiliary model, and thereby, the binding functions. This is generally true even if we simply shift the constraints $\square(\boldsymbol{\theta})$ by a fixed amount. Therefore, except in some special cases, a general discussion of the effects of the constraints on efficiency is just as elusive as a general discussion of the effects of replacing a parametric auxiliary model by another one, which to the best of our knowledge, remains the most important unresolved issue in the II literature.

Nevertheless, we can be more specific in certain important situations, such as when the equality restrictions are correct:

Proposition 8 *Under Assumptions 1 and 2*

$$\sqrt{\square}\left\{\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_\square^u)^{-\square}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_\square^e)^{-\square}\right]\right\}=\square_p(1)$$

if $\square\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^\square)\right]=\mathbf{0}$.

In particular, any unconstrained II estimator is asymptotically equivalent to an equality constrained II estimator that sets all the parameters of the auxiliary model to their unconstrained pseudo-true values, $\boldsymbol{\theta}^u(\boldsymbol{\rho}^\square)$.

Of course, if we knew that the equality constraints were indeed correct, we might be able to obtain more efficient estimators of the parameters of interest by using the solution proposed by Dridi (2000), who derives II estimators of $\boldsymbol{\rho}$ on the basis of a correctly overidentified auxiliary model. At the same time, the main advantage of our solution over Dridi's is that by effectively saturating his overidentifying moment conditions with Lagrange multipliers to mop up any possible bias, it produces consistent estimators of the parameters of interest even if the overidentifying restrictions are not really fulfilled by the unrestricted pseudo-true values of the auxiliary parameters.

But as we saw in the introduction, the equivalence between $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\varphi}^u)^{-1}\right]$ and $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_{\varphi}^e)^{-1}\right]$ may also hold with incorrect constraints. For instance, this is always the case when the auxiliary model is a linear autoregression, and the restrictions are linear in the autoregressive coefficients, as in section 3.1. More formally:

Proposition 9 *Under Assumptions 1 and 2*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\varphi}^u)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_{\varphi}^e)^{-1}\right]\right\}=o_p(1)$$

if

$$l_t(\boldsymbol{\theta})=-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^2-\frac{1}{2\sigma^2}(y_t-\phi_1 y_{t-1}-\cdots-\phi_k y_{t-k})^2$$

$$\mathbf{r}(\boldsymbol{\theta})=\mathbf{R}\boldsymbol{\phi}-\mathbf{q}$$

and $\mathrm{rank}(\mathbf{R}')=r$*, where* $\boldsymbol{\phi}=(\phi_1,\ldots,\phi_k)$*, and* $\boldsymbol{\theta}=(\boldsymbol{\phi}',\sigma^2)'$*.*

Note that such a result does not really depend on the nature of the true model, whose parameters only enter through the first $k+1$ theoretical "autocovariances" of y_t, $\gamma_j(\boldsymbol{\rho})=E\left(y_t y_{t-j}|\,\boldsymbol{\rho}\right)$ $(j=0,1,\ldots,k)$, but rather on the particular form of the auxiliary model used. The reason is that from the point of II estimation, the first $k+1$ sample "autocovariances" γ_{jT}^r $(j=0,1,\ldots,k)$ play the role of "sufficient statistics" of the auxiliary model from which we infer $\boldsymbol{\rho}$.

In contrast, the asymptotic relationship of the inequality restricted estimators of the parameters of interest with $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\varphi}^u)^{-1}\right]$ and $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_{\varphi}^e)^{-1}\right]$ can be derived under general circumstances. For the sake of clarity, though, we shall only present a formal result for the case a single restriction

Proposition 10 *Under Assumptions 1 and 2*

$$\sqrt{\square}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_{\square}^i)^{-\square}\right]-\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\square}^u)^{-\square}\right]\right\}=\square_p(1)$$

if $\square\,[\boldsymbol{\theta}^u(\boldsymbol{\rho}^{\square})]\ \square\ \mathbf{0}\square$

$$\sqrt{\square}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_{\square}^i)^{-\square}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_{\square}^e)^{-\square}\right]\right\}=\square_p(1)$$

if $\square\,[\boldsymbol{\theta}^u(\boldsymbol{\rho}^{\square})]\ \square\ \mathbf{0}$*, and*

$$\sqrt{\square}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_{\square}^i)^{-\square}\right]-\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\square}^u)^{-\square}\right]\right\}=\square_p(1)=\sqrt{\square}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_{\square}^i)^{-\square}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_{\square}^e)^{-\square}\right]\right\}$$

if $\square\,[\boldsymbol{\theta}^u(\boldsymbol{\rho}^{\square})]=\mathbf{0}$*.*

In fact, the *inequality constrained* and *unconstrained* II procedures will yield numerically identical results if the inequality restriction is not binding in a given sample, since in that case $\bar{\boldsymbol{\theta}}_T^i$ coincides with the unconstrained pseudo-ML estimator, $\bar{\boldsymbol{\theta}}_T^u$ (and $\bar{\boldsymbol{\mu}}_T^i$ with $\bar{\boldsymbol{\mu}}_T^u=\mathbf{0}$). Similarly, the *inequality* and *equality constrained* procedures will yield numerically identical results if the inequality restriction is binding in a given sample, since in that case $\bar{\boldsymbol{\theta}}_T^i$ coincides with the *equality constrained* pseudo-ML estimator, $\bar{\boldsymbol{\theta}}_T^e$, and consequently $\bar{\boldsymbol{\mu}}_T^i$ with $\bar{\boldsymbol{\mu}}_T^e$.

In the case of multiple inequality constraints, $\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_{\square}^i)^{-\square}\right]$ will numerically coincide with either the unrestricted estimator, or an equality restricted estimator that imposes the subset of the $\square$ constraints that happen to be satisfied with equality by $\bar{\boldsymbol{\theta}}_T^i$. Therefore, it is not surprising that the *inequality constrained* optimal II estimator will be asymptotically equivalent to $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_{\square}^u)^{-\square}\right]$ if $\square\,[\boldsymbol{\theta}^u(\boldsymbol{\rho}^{\square})]\ \square\ \mathbf{0}$, or to some equality restricted estimator otherwise. We shall provide a detailed illustration of Propositions 8, 9 and 10 in section 3.1 below.

Finally, it would be interesting to characterise under which circumstances "restricted" II estimators of $\boldsymbol{\rho}$ achieve the usual asymptotic Cramer-Rao efficiency bound. Proposition 2 in GT provides a leading example in the context of unrestricted II estimation, namely, when the auxiliary model "smoothly embeds" the true model (see Definition 1 in GT). Unfortunately, it is often the case that the auxiliary model does not nest the true model, as the examples in section 3 illustrate. However, this is not by any means the only such situation. As the example in section 3.2 shows, there are other cases in which we can achieve full efficiency by adding completely false constraints to a badly misspecified auxiliary model. More generally, we can state the following result:

Proposition 11 *Under Assumptions 1 and 2, the optimal CMD II estimator* $\hat{\boldsymbol{\rho}}_T^r\left[\mathcal{K}_{\square}^{r\prime}\left(\mathcal{I}_{\square}^{\boxplus}\right)^{\square}\mathcal{K}_{\square}^r\right]$ *is asymptotically as efficient as the ML estimator of* $\boldsymbol{\rho}$ *based on the true model if the latter only depends on the data through a continuously differentiable function of the first* $\square$ *elements of* $\mathcal{K}_{\square}^r\bar{\boldsymbol{\beta}}_T^r$*, and* $\mathcal{K}_{\square}^r\boldsymbol{\beta}^r(\boldsymbol{\rho})$ *is twice continuously differentiable in* $\boldsymbol{\rho}$ *on a neighbourhood of* $\boldsymbol{\rho}^{\square}$.

2.3 Extensions

2.3.1 Partially optimised unconstrained procedures

One approach commonly followed by users of II estimation methods is to select a simple auxiliary model that closely resembles the model of interest, but whose pseudo-log likelihood function is easy to evaluate, so that they can fully optimise it very rapidly. Many other empirical researchers, though, prefer 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 (see GM96). 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 unconstrained GMM-based II procedures, and assume that the numerical procedure used to maximise the pseudo log-likelihood function $L_T(\boldsymbol{\theta})$ is the Newton-Raphson method without line searches, 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 $\bar{\boldsymbol{\theta}}_T^{(k)}$ denotes the value of the parameters after iteration k $(1 \leq k \leq k_{max})$, we assume that the recursive formula employed is

$$\bar{\boldsymbol{\theta}}_T^{(k)} = \bar{\boldsymbol{\theta}}_T^{(k-1)} + \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t\left(\bar{\boldsymbol{\theta}}_T^{(k-1)}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial l_t\left(\bar{\boldsymbol{\theta}}_T^{(k-1)}\right)}{\partial \boldsymbol{\theta}}$$

Let us initially consider the case of $k_{max} = 0$, so that no optimisation whatsoever takes place. If the initial value $\bar{\boldsymbol{\theta}}_T^{(0)}$ is non-stochastic, we simply have a special case of the equality constrained GMM-based II estimator, with the restrictions $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$. In effect, this transforms the GMM II procedure in a CMD II procedure in which we match the values of the multipliers $\bar{\boldsymbol{\mu}}_T^{(0)}$ in the actual sample and the population. Nevertheless, note that if the value of $\boldsymbol{\theta}^{(0)}$ is not sensibly chosen by the practitioner, it may well fail to satisfy the required conditions in Assumptions 1 and 2. Typically, however, $\bar{\boldsymbol{\theta}}_T^{(0)}$ 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 (see e.g. Calzolari, Fiorentini and Sentana, 2001). If that is the case, the results in sections 2.1 imply that the *fully non-optimised* GMM II estimator of $\boldsymbol{\rho}$ based on $\bar{\boldsymbol{\theta}}_T^{(0)}$ and $\bar{\boldsymbol{\mu}}_T^{(0)}$, $\tilde{\boldsymbol{\rho}}_T^{(0)}$ say, will be consistent and asymptotically normal, as long as the regularity conditions in Assumptions 1 and 2 (with $\partial r'(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \mathbf{I}_q$) remain valid when (i) $\bar{\boldsymbol{\theta}}_T^r$ is replaced by $\bar{\boldsymbol{\theta}}_T^{(0)}$, (ii) $\boldsymbol{\theta}^r(\boldsymbol{\rho}_0)$ by the pseudo-true value of $\bar{\boldsymbol{\theta}}_T^{(0)}$, $\boldsymbol{\theta}^{(0)}(\boldsymbol{\rho}_0)$ say, (iii) $\bar{\boldsymbol{\mu}}_T^r$ by $\bar{\boldsymbol{\mu}}_T^{(0)}$, which are the Lagrange multipliers required to satisfy the sample first-order conditions (3) at $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_T^{(0)}$, and (iv) $\boldsymbol{\mu}^r(\boldsymbol{\rho}_0)$ by the corresponding pseudo-true value,

$\boldsymbol{\mu}^{\square\square\square}(\boldsymbol{\rho}^{\square})$.

Let us now consider the more interesting case of $\square_{\square\square\square} = 1$. It is then clear that $\bar{\boldsymbol{\theta}}_T^{\square\square\square}$ and $\bar{\boldsymbol{\mu}}_T^{\square\square\square}$ will also be stochastic, with pseudo-true values given by

$$\begin{aligned} \boldsymbol{\theta}^{\square\square\square}(\boldsymbol{\rho}^{\square}) &= \boldsymbol{\theta}^{\square\square\square}(\boldsymbol{\rho}^{\square}) - \boldsymbol{\mu}^{\square\square\square}(\boldsymbol{\rho}^{\square})\mathcal{J}_{\square}^{\square\square\square} \\ \boldsymbol{\mu}^{\square\square\square}(\boldsymbol{\rho}^{\square}) &= -\square \left\{ \left. \frac{\square\square_t \left[\boldsymbol{\theta}^{\square\square\square}(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\theta}} \right| \boldsymbol{\rho}^{\square} \right\} \end{aligned}$$

If, *mutatis mutandi*, the regularity conditions in Assumptions 1 and 2 remain valid, then the *one-step optimised* GMM estimator of $\boldsymbol{\rho}$ based on $\bar{\boldsymbol{\theta}}_T^{\square\square\square}$ and $\bar{\boldsymbol{\mu}}_T^{\square\square\square}$, $\tilde{\boldsymbol{\rho}}_T^{\square\square\square}$ say, will also be consistent and asymptotically normal. But since the above argument does not really depend on $\square_{\square\square\square}$ being 1, or the way in which $\bar{\boldsymbol{\theta}}_T^{\square\square\square}$ was obtained, it remains valid for any $\square_{\square\square\square}$.

Although situations in which an applied researcher knowingly decides to proceed with a partially optimised auxiliary model may seem hard to envisage, there are at least two practical cases in which the results of this subsection may be of some use: (i) to allow for the fact that the numerical algorithm used to optimise the auxiliary objective function may have converged very close to, but not exactly at the optimum, as we do in section 3.3, and (ii) to cater for an increasing number of practitioners who use the SNP auxiliary model suggested by GT with a ever growing number of terms in the Hermite expansions to obtain what has become commonly known as EMM estimators of $\boldsymbol{\rho}$. In both cases, 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 II estimator of the parameters of interest $\boldsymbol{\rho}$, as long as the moment conditions used include Lagrange multipliers to reflect the lack of convergence of the algorithm.

2.3.2 Pre-test estimators

For reasons analogous to the ones discussed at the beginning of the previous subsection, 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 II 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, provided that in the latter case the information in the corresponding Lagrange multiplier is taken into account. If the specification test is consistent (in the sense that it rejects the null hypothesis with probability approaching one as the sample size increases when the unrestricted pseudo-true value of the relevant parameter is different from zero), then the limiting distribution of the pre-test II estimator of $\boldsymbol{\rho}$ is the same as the limiting distribution of the fully optimised unconstrained II estimator. In contrast, if the limiting unrestricted pseudo-true value is exactly zero, then the limiting distribution of the pre-test estimator of $\boldsymbol{\rho}$ will be a mixture of the equality restricted estimator, and the unconstrained estimator. But since equality restricted and un-

3 Examples

3.1 MA(1) estimated as AR(1)

3.1.1 True and auxiliary models

Consider the following Gaussian MA(1) process:

$$y_t = \varepsilon_t - \psi\varepsilon_{t-1}, \quad \varepsilon_t|I_{t-1};\boldsymbol{\rho} \sim N(0,\sigma^2), \quad |\psi| \leq 1, \quad 0 < \sigma^2 < \infty \tag{14}$$

where the parameters of interest are $\boldsymbol{\rho} = (\psi, \sigma^2)'$. It is well known that $E(y_t|\boldsymbol{\rho}) = 0$, and that its autocovariance structure is given by

$$\gamma_0(\boldsymbol{\rho}) = (1+\psi^2)\sigma^2, \qquad \gamma_1(\boldsymbol{\rho}) = -\psi\sigma^2, \qquad \gamma_j(\boldsymbol{\rho}) = 0, \; j > 1. \tag{15}$$

In order to estimate $\boldsymbol{\rho}$ by II, we are going to consider the following first-order autoregression:

$$y_t = \phi y_{t-1} + u_t, \quad u_t|I_{t-1};\boldsymbol{\theta} \sim N(0,\omega), \quad \omega \geq 0$$

possibly subject to the restrictions $\phi = \phi_{max}$ or $\phi \geq \phi_{min}$, with $|\phi_{min}| < \infty$ but otherwise arbitrary, so that $\boldsymbol{\theta} = (\phi, \omega)'$. In this respect, note that the unrestricted auxiliary model only nests the true model if $\psi^0 = 0$.

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:

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

and the (scaled) Lagrangian function by

$$\Lambda_T(\boldsymbol{\beta}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\omega - \frac{1}{2\omega}\frac{1}{T}\sum_t (y_t - \phi y_{t-1})^2 + (\phi - \phi_{max})\lambda_\phi + \omega\lambda_\omega$$

where $\boldsymbol{\mu} = (\square_\phi \square \square_\omega)'$ are the multipliers associated with the (in)equality restrictions $\square \geq \square_{\square\square\square}$ and $\square \geq 0$ respectively. Therefore, the sample first-order conditions that take into account those constraints will be given by:

$$
\begin{aligned}
\frac{1}{\bar{\square}_T^r} \frac{1}{\square} \sum_t (\square_t - \bar{\square}_T^r \square_{t-\square}) \square_{t-\square} + \bar{\square}_{\phi T}^r &= 0 \\
\frac{1}{2\bar{\square}_T^r} \frac{1}{\square} \sum_t \left[\frac{(\square_t - \bar{\square}_T^r \square_{t-\square})^\square}{\bar{\square}_T^r} - 1 \right] + \bar{\square}_{\omega T}^r &= 0
\end{aligned}
\tag{16}
$$

together with the complementary slackness conditions:

$$
\begin{gathered}
(\bar{\square}_T^r - \square_{\square\square\square}) \cdot \bar{\square}_{\phi T}^r = 0 \\
\bar{\square}_T^r \cdot \bar{\square}_{\omega T}^r = 0
\end{gathered}
$$

plus the appropriate (in)equality restrictions on parameters and/or multipliers. But since in all three cases

$$
\bar{\square}_T^r = \frac{1}{\square} \sum_t (\square_t - \bar{\square}_T^r \square_{t-\square})^\square \geq 0\square
$$

we can safely take $\bar{\square}_{\omega T}^r$ as 0 in what follows. Also note that since

$$
\bar{\square}_{\phi T}^r = -\frac{1}{\bar{\square}_T^r} \frac{1}{\square} \sum_t (\square_t - \bar{\square}_T^r \square_{t-\square}) \square_{t-\square}
$$

we can interpret the other multiplier as (minus) the coefficient in the OLS regression of $\square_{t-\square}$ on the "restricted" residuals $(\square_t - \bar{\square}_T^r \square_{t-\square})$ (see Gourieroux, Holly and Monfort, 1982). Therefore, we will have that $\bar{\square}_T^i$, $\bar{\square}_T^i$ and $\bar{\square}_{\phi T}^i$ will be equal to the unrestricted estimators

$$
\bar{\square}_T^u = \frac{\bar{\square}_{\square\square T}}{\bar{\square}_{\square\square T}}\square \qquad \bar{\square}_T^u = \bar{\square}_{\square\square T} - \frac{\bar{\square}_{\square\square T}^\square}{\bar{\square}_{\square\square T}}\square \qquad \bar{\square}_{\phi T}^u = 0
$$

if $\bar{\square}_T^u \geq \square_{\square\square\square}$, or to the equality restricted estimators

$$
\bar{\square}_T^e = \square_{\square\square\square}\square \;\; \bar{\square}_T^e = \bar{\square}_{\square\square T} + \square_{\square\square\square}^\square \bar{\square}_{\square\square T} - 2\square_{\square\square\square} \bar{\square}_{\square\square T}\square \;\; \bar{\square}_{\phi T}^e = \frac{-\left(\bar{\square}_{\square\square T} - \square_{\square\square\square} \bar{\square}_{\square\square T}\right)}{\bar{\square}_{\square\square T} + \square_{\square\square\square}^\square \bar{\square}_{\square\square T} - 2\square_{\square\square\square} \bar{\square}_{\square\square T}}
$$

otherwise, where:

$$\bar{\mathbf{\Sigma}}_T = \begin{pmatrix} \bar{\square}_{\square\square T} & \bar{\square}_{\square\square T} \\ \bar{\square}_{\square\square T} & \bar{\square}_{\square\square T} \end{pmatrix} = \frac{1}{\square}\sum_t \begin{pmatrix} \square_t \\ \square_{t-\square} \end{pmatrix} \begin{pmatrix} \square_t & \square_{t-\square} \end{pmatrix}$$

is the sample second moment matrix.

3.1.3 Population moments and binding functions

If we now define

$$\mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta}) = \square\left[\square_t(\boldsymbol{\theta})\right|\boldsymbol{\rho}] = -\frac{1}{2}\ln 2\square - \frac{1}{2}\ln\square - \frac{(1+\square^{\square})\square_{\square}(\boldsymbol{\rho}) - 2\square\square_{\square}(\boldsymbol{\rho})}{2\square}\square$$

it is clear that the binding functions $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ satisfy the moment conditions

$$\begin{aligned} \mathfrak{m}_\phi\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho})\right] &= \mathbf{0} \\ \mathfrak{m}_\omega\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho})\right] &= \mathbf{0} \end{aligned}$$

together with the exclusion restrictions

$$\begin{aligned} [\square^r(\boldsymbol{\rho}) - \square_{\square\square\square}]\cdot\square^r_\phi(\boldsymbol{\rho}) &= 0 \\ \square^r(\boldsymbol{\rho})\cdot\square^r_\omega(\boldsymbol{\rho}) &= 0 \end{aligned}$$

plus the appropriate (in)equality restrictions on $\boldsymbol{\theta}^r(\boldsymbol{\rho})$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho})$, where

$$\begin{aligned} \mathfrak{m}_\phi(\boldsymbol{\rho};\boldsymbol{\beta}) &= \frac{1}{\square}\left[\square_{\square}(\boldsymbol{\rho}) - \square\square_{\square}(\boldsymbol{\rho})\right] + \square_\phi \\ \mathfrak{m}_\omega(\boldsymbol{\rho};\boldsymbol{\beta}) &= \frac{1}{2\square}\left[\frac{-2\square\square_{\square}(\boldsymbol{\rho}) + (1+\square^{\square})\square_{\square}(\boldsymbol{\rho})}{\square} - 1\right] + \square_\omega \end{aligned} \quad (17)$$

and the dependence of $\square_j$ on $\boldsymbol{\rho}$ comes from (15).

From here, it is easy to see that

$$\square^r(\boldsymbol{\rho}) = \square\left\{\left[\square_t - \square^r(\boldsymbol{\rho})\square_{t-\square}\right]^{\square}\right|\boldsymbol{\rho}\} = \left\{1 + [\square^r(\boldsymbol{\rho})]^{\square}\right\}\square_{\square}(\boldsymbol{\rho}) - 2\square^r(\boldsymbol{\rho})\square_{\square}(\boldsymbol{\rho}) \geq 0$$

so that $\square^r_\omega(\boldsymbol{\rho}) = 0$. As for the other elements, in principle there may be two different situations. Specifically, $\square^i(\boldsymbol{\rho})$, $\square^i(\boldsymbol{\rho})$ and $\square^i_\phi(\boldsymbol{\rho})$ will be equal to

$$\square^u(\boldsymbol{\rho}) = \frac{\square_\square(\boldsymbol{\rho})}{\square_\square(\boldsymbol{\rho})}\square \qquad \square^u(\boldsymbol{\rho}) = \square_\square(\boldsymbol{\rho}) - \frac{\square^\square_\square(\boldsymbol{\rho})}{\square_\square(\boldsymbol{\rho})}\square \qquad \square^u_\phi(\boldsymbol{\rho}) = 0$$

respectively when $-\square\square(1+\square^\square) \geq \square_{\square\square\square}$. Otherwise, they will be equal to

$$\square^e(\boldsymbol{\rho}) = \square_{\square\square\square}\square \;\; \square^e(\boldsymbol{\rho}) = (1+\square^\square_{\square\square\square})\square_\square(\boldsymbol{\rho}) - 2\square_\square(\boldsymbol{\rho})\square \;\; \square^e_\phi(\boldsymbol{\rho}) = \frac{-\left[\square_\square(\boldsymbol{\rho}) - \square_{\square\square\square}\square_\square(\boldsymbol{\rho})\right]}{(1+\square^\square_{\square\square\square})\square_\square(\boldsymbol{\rho}) - 2\square_\square(\boldsymbol{\rho})}$$

Figure 1 plots the binding functions $\square^u(\boldsymbol{\rho})$ and $\square^e_\phi(\boldsymbol{\rho})$ for $-1 \leq \square \leq 1$ and $\square_{\square\square\square} = 0$. Note that in this case, $\square^i(\boldsymbol{\rho}) = \max\left[\square^u(\boldsymbol{\rho})\square 0\right]$ and $\square^i_\phi(\boldsymbol{\rho}) = \max\left[\square^e_\phi(\boldsymbol{\rho})\square 0\right]$.

3.1.4 Asymptotic distributions of pseudo-ML estimators and sample moments

Given the different expressions for the inequality restricted pseudo-ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ discussed previously, the sample counterparts to (17) will be given by either:

$$\mathfrak{m}_\phi(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^i_T) = \frac{\left[\square_\square(\boldsymbol{\rho}) - (\bar{\square}_{\square\square T}\square\bar{\square}_{\square\square T})\square_\square(\boldsymbol{\rho})\right]}{\bar{\square}_{\square\square T} - \bar{\square}^\square_{\square\square T}\square\bar{\square}_{\square\square T}}$$

$$\mathfrak{m}_\omega(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^i_T) = \frac{\left[-2(\bar{\square}_{\square\square T}\square\bar{\square}_{\square\square T})\square_\square(\boldsymbol{\rho}) + (1+\bar{\square}^\square_{\square\square T}\square\bar{\square}^\square_{\square\square T})\square_\square(\boldsymbol{\rho}) - (\bar{\square}_{\square\square T} - \bar{\square}^\square_{\square\square T}\square\bar{\square}_{\square\square T})\right]}{2\left(\bar{\square}_{\square\square T} - \bar{\square}^\square_{\square\square T}\square\bar{\square}_{\square\square T}\right)^\square}$$

when $\bar{\square}_{\square\square T}\square\bar{\square}_{\square\square T} \geq \square_{\square\square\square}$, which are precisely the sample moments that we would use in an *unrestricted* GMM-based II procedure, or

$$\mathfrak{m}_\phi(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^i_T) = \frac{\left[\square_\square(\boldsymbol{\rho}) - \bar{\square}_{\square\square T}\right] - \square_{\square\square\square}\left[\square_\square(\boldsymbol{\rho}) - \bar{\square}_{\square\square T}\right]}{\bar{\square}_{\square\square T} + \square^\square_{\square\square\square}\bar{\square}_{\square\square T} - 2\square_{\square\square\square}\bar{\square}_{\square\square T}}$$

$$\mathfrak{m}_\omega(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^i_T) = \frac{\left[-2\square_{\square\square\square}\square_\square(\boldsymbol{\rho}) + (1+\square^\square_{\square\square\square})\square_\square(\boldsymbol{\rho})\right] - \left(\bar{\square}_{\square\square T} + \square^\square_{\square\square\square}\bar{\square}_{\square\square T} - 2\square_{\square\square\square}\bar{\square}_{\square\square T}\right)}{2\left(\bar{\square}_{\square\square T} + \square^\square_{\square\square\square}\bar{\square}_{\square\square T} - 2\square_{\square\square\square}\bar{\square}_{\square\square T}\right)^\square}$$

when $\bar{\square}_{\square\square T}\square\bar{\square}_{\square\square T} \leq \square_{\square\square\square}$, which are the sample moments that correspond to the *equality constrained* GMM-based procedure that imposes the restriction $\square = \square_{\square\square\square}$.

Let us 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) $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] \,\square\, \square_{\square\square\square}$, (ii) $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] \,\square\, \square_{\square\square\square}$, and (iii) $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] = \square_{\square\square\square}$. 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* $\square_t$ *is given by the Gaussian* MA*(1) model (14), the first sample autocorrelation* $\bar{\square}^u_T$ *is* $\square^{\square/\square}$*-consistent for the first population autocorrelation* $\square^u(\boldsymbol{\rho}^{\square})$*, with the following limiting distribution*

$$\sqrt{\square}\left[\bar{\square}^u_T - \square^u(\boldsymbol{\rho}^{\square})\right] \xrightarrow{d} \square\left\{0\square\,\frac{1+(\square^{\square})^{\square}+4(\square^{\square})^{\square}+(\square^{\square})^{\square}+(\square^{\square})^{\square}}{\left[1+(\square^{\square})^{\square}\right]^{\square}}\right\}$$

Note that the asymptotic variance of $\bar{\square}^u_T$, which not surprisingly is the same for a non-invertible MA(1) process with parameter $1\square\square$, achieves its maximum (=1) for $\square^{\square}=0$ and its minimum (=1/2) for $\square^{\square}=\pm1$. As a result, we will have that

$$\lim_{T\to\infty}\square\left[\sqrt{\square}(\bar{\square}^u_T-\square_{\square\square\square})\,\square\,0\right]=\lim_{T\to\infty}\square(\sqrt{\square}\,\square^e_T\,\square\,0)=\begin{cases}1 & \text{if } \square^u(\boldsymbol{\rho}^{\square})\,\square\,\square_{\square\square\square}\\ 1\square2 & \text{if } \square^u(\boldsymbol{\rho}^{\square})=\square_{\square\square\square}\\ 0 & \text{if } \square^u(\boldsymbol{\rho}^{\square})\,\square\,\square_{\square\square\square}\end{cases}$$

Hence, when $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] \,\square\, \square_{\square\square\square}$, $\sqrt{\square}\left(\bar{\square}^i_T-\bar{\square}^u_T\right)$ and $\sqrt{\square}\,\bar{\square}^i_{\square T}$ are both $\square_p(1)$, and the *inequality restricted* II estimators of $\boldsymbol{\rho}$ are asymptotically equivalent to the usual *unrestricted* II estimators. In contrast, when $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] \,\square\, \square_{\square\square\square}$, $\sqrt{\square}(\bar{\square}^i_T-\square_{\square\square\square})$ and $\sqrt{\square}\,(\bar{\square}^i_{\square T}-\bar{\square}^e_{\square T})$ are $\square_p(1)$, and the *inequality restricted* II estimators of $\boldsymbol{\rho}$ will then coincide in large samples with the *equality restricted* ones. The most interesting situation arises when $-\square^{\square}\square\left[1+(\square^{\square})^{\square}\right] = \square_{\square\square\square}$. In this case, $\bar{\boldsymbol{\beta}}^i_T$ has a non-normal asymptotic distribution, as it will be equal to either $(\bar{\square}^u_T\square\,\bar{\square}^u_T\square\,0)'$ or $(\square_{\square\square\square}\square\,\bar{\square}^e_T\square\,\bar{\square}^e_{\square T})'$ with probability approximately one half each. As a consequence, the sample moment conditions will also be $\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^u_T)$ fifty per cent of the time, and $\mathfrak{m}(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^e_T)$ the other fifty. Nevertheless, given that we can easily prove that $\sqrt{\square}\left[\mathfrak{m}_\phi(\boldsymbol{\rho}^\square;\bar{\boldsymbol{\beta}}^u_T)-\mathfrak{m}_\phi(\boldsymbol{\rho}^\square;\bar{\boldsymbol{\beta}}^e_T)\right]=\square_p(1)$ when $-\square^\square\square\left[1+\left(\square^\square\right)^\square\right]=\square_{\square\square\square}$, the limiting distribution of $\sqrt{\square}\,\mathfrak{m}_\phi(\boldsymbol{\rho}^\square;\bar{\boldsymbol{\beta}}^i_T)$ will also be normal, with an analogous result for the other moment. In this respect, the above results can be regarded as an illustration of Propositions 8 and 10.

3.1.5 Indirect inference estimators

If the parameters of interest of the true model were $\boldsymbol{\gamma}=(\square_\square\square\square_\square)'$ rather than $\boldsymbol{\rho}$, the solution of the linear system of equations $\mathfrak{m}[\boldsymbol{\gamma}\square\bar{\boldsymbol{\beta}}^r_T]=0$ with respect to its first argument would give us the "*restricted*" GMM-based estimator of these autocovariances, $\tilde{\boldsymbol{\gamma}}^r_T$. In this context, it is easy to see that if $\square_{\square\square T}=\square_{\square\square T}$, we would have that all three GMM-based estimators of $\boldsymbol{\gamma}$ would be numerically identical to

$$\begin{pmatrix}\square^r_{\square T}\\ \square^r_{\square T}\end{pmatrix}=\begin{pmatrix}\square_{\square\square T}\\ \square_{\square\square T}\end{pmatrix} \tag{18}$$

regardless of the sign of $\square^u(\boldsymbol{\rho}^\square)-\square_{\square\square\square}$, and indeed, regardless of the value of $\square_{\square\square\square}$. In finite samples, of course, $\square_{\square\square T}\neq\square_{\square\square T}$, but given that $\square_{\square\square T}-\square_{\square\square T}=\square_p(\square^{-\square})$ for any value of $\square^\square$, it is straightforward to show that $\tilde{\boldsymbol{\gamma}}^u_T$, $\tilde{\boldsymbol{\gamma}}^e_T$ and $\tilde{\boldsymbol{\gamma}}^i_T$ are always asymptotically equivalent.

In addition, since $\square_\square$ and $\square_\square$ are free parameters, and the auxiliary model exactly identifies them, their CMD-based II estimators will be numerically identical to the GMM-based II estimators in large samples, as indicated by Proposition 6.

The common asymptotic distribution of $\tilde{\boldsymbol{\gamma}}^u_T\square\tilde{\boldsymbol{\gamma}}^e_T$ and $\tilde{\boldsymbol{\gamma}}^i_T$ can be directly obtained as a special case of Theorem 8.4.2 in Anderson (1971):

Lemma 2 *When $\square_t$ is given by the Gaussian* MA*(1) model (14), $\square^r_{\square T}$ and $\square^r_{\square T}$ are $\square^{\square/\square}$-consistent for $\square_\square(\boldsymbol{\rho}^\square)$ and $\square_\square(\boldsymbol{\rho}^\square)$ respectively, with the following limiting distribution*

$$\sqrt{\square}\left[\tilde{\boldsymbol{\gamma}}^r_T-\square(\boldsymbol{\rho}^\square)\right]\xrightarrow{d}\square\left[0\square\mathbf{V}(\boldsymbol{\rho}^\square)\right]$$

where

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

But even though $\boldsymbol{\gamma}$ are not really the parameters of interest, we can regard their II estimators as "sufficient statistics" of the auxiliary model from which we can estimate $\boldsymbol{\rho}$. In particular, if $\left|\hat{\square}_T\right| \leq \square 5$, we can obtain all the II estimators of $\boldsymbol{\rho}$ (irrespective of the weighting matrix) by solving numerically the nonlinear system of equations (15).

The asymptotic equivalence of the different constrained II estimators is somewhat surprising, for in principle, it may seem that by choosing $\square_{\square\square\square} = -\square 5$ one should obtain more efficient estimators of the parameters of interest. The intuition would be that since the inequality constraint $\square \geq -\square 5$ is trivially satisfied by the unrestricted binding function $\square^u(\boldsymbol{\rho})$, and the inequality restricted pseudo-ML estimator $\bar{\square}_T^i$ must necessarily have less sampling variance than the unrestricted estimator $\bar{\square}_T^u$ around the pseudo-true value $\square^i(\boldsymbol{\rho}^{\square}) = \square^u(\boldsymbol{\rho}^{\square})$, then the inequality restricted II estimator of $\boldsymbol{\rho}$ should be more efficient than the unrestricted one. However, our previous results, which can be regarded as an illustration of Proposition 9, imply that such an intuition is not correct, as both II estimators are asymptotically equivalent regardless of the value of $\square_{\square\square\square}$.

An analogous line of reasoning applies to *pretest* 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 $\square = \square_{\square\square\square}$, or the *unrestricted* estimators when it does. Since as we have just seen, $\tilde{\boldsymbol{\rho}}_T^u$ and $\tilde{\boldsymbol{\rho}}_T^e$ have the same asymptotic distribution regardless of the value of $\square^{\square}$, such a common distribution will be inherited by the *pretest* estimators.

Finally, note that since the unrestricted auxiliary model "smoothly embeds" the true model when $\square^{\square} = 0$, Theorem 2 in GT implies that in this particular case, the unrestricted estimator $\tilde{\boldsymbol{\rho}}_T^u$ is asymptotically equivalent to maximum likelihood, and the same obviously applies to all the other estimators. However, the asymptotic efficiency of $\tilde{\boldsymbol{\rho}}_T^r$ relative to the ML estimator decreases as $\left|\square^{\square}\right|$ increases. In particular, if $\square^{\square} = 1$, the ML estimator of $\square$ is superconsistent (i.e. consistent at the rate $\square$; see Sargan and Bhargava, 1983), while the asymptotic distribution of $\sqrt{\square}(\tilde{\square}_T^r - \square^{\square})$ is half normal. The reason is that $\lim_{T\to\infty} \square \left[\sqrt{\square}(\bar{\square}_T^u - \square 5) \square 0\right] = 1\square 2$ by virtue of Lemma 1, which in turn means that the system of equations (15) evaluated at $\boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}}_T^r$ does not have a real solution in $\boldsymbol{\rho}$ half the time. In those cases, one attractive possibility involves the minimisation of the optimal (continuously updated) CMD criterion:

$$\begin{pmatrix} \bar{\square}_{\square T}^r - \square_{\square}(\boldsymbol{\rho}) & \bar{\square}_{\square T}^r - \square_{\square}(\boldsymbol{\rho}) \end{pmatrix} \mathbf{V}^{-\square}(\boldsymbol{\rho}) \begin{pmatrix} \bar{\square}_{\square T}^r - \square_{\square}(\boldsymbol{\rho}) \\ \bar{\square}_{\square T}^r - \square_{\square}(\boldsymbol{\rho}) \end{pmatrix}$$

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

$$\left.\begin{array}{c} \tilde{\square}_T^r = 0 \\ \tilde{\square}_T^r = \bar{\square}_{\square\square T} \end{array}\right\} \qquad \text{if } \bar{\square}_T^u = 0$$

$$\left.\begin{array}{c} \tilde{\square}_T^r = \left[-1 + \sqrt{1 - 4\left(\bar{\square}_T^u\right)^{\square}}\right] \square (2\bar{\square}_T^u) \\ \tilde{\square}_T^r = \bar{\square}_{\square\square T} \square \left[1 + (\tilde{\square}_T^r)^{\square}\right] \end{array}\right\} \qquad \text{if } 0 \square \left(\bar{\square}_T^u\right)^{\square} \leq \square 25$$

$$\left.\begin{array}{c} \tilde{\square}_T^r = -\square\square\square\square(\bar{\square}_T^u) \\ \tilde{\square}_T^r = \bar{\square}_{\square\square T}\left[7 + 12\left(\bar{\square}_T^u\right)^{\square} - 16\left|\bar{\square}_T^u\right|\right] \square \left(6 - 4\left|\bar{\square}_T^u\right|\right) \end{array}\right\} \qquad \text{if } \left(\bar{\square}_T^u\right)^{\square} \square \square 25$$

Asymptotically equivalent estimators will be obtained if we replace the above CMD criterion with the optimal (continuously updated) GMM criterion based on the moment conditions

$$\square\left[\square_t^{\square} - \square_{\square}(\boldsymbol{\rho})\right] = 0$$

$$\square\left[\square_t \square_{t-\square} - \square_{\square}(\boldsymbol{\rho})\right] = 0$$

3.2 AR(1) estimated as MA(1)

3.2.1 True and auxiliary models

Consider now the following stationary AR(1) process:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \varepsilon_t | y_{t-1}, \ldots \sim N(0, \omega), \quad |\alpha| < 1, 0 < \omega < \infty \tag{19}$$

where the parameters of interest this time are $\boldsymbol{\rho} = (\alpha, \omega)'$. It is well known that $E(y_t | \boldsymbol{\rho}) = 0$, and that its autocovariance structure is given by

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

In order to estimate $\boldsymbol{\rho}$ by II, we are going to use the following MA(1) model:

$$y_t = u_t - \psi u_{t-1}, \quad u_t | y_{t-1}, \ldots \sim N(0, \sigma), \quad \sigma \geq 0$$

possibly subject to the restrictions $\psi = 0$ or $\psi \leq 0$, so that $\boldsymbol{\theta} = (\psi, \sigma)'$. In this respect, note that the unrestricted auxiliary model only nests the true model if $\alpha^2 = 0$.

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:

$$L_T(\boldsymbol{\theta}) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln \sigma - \frac{1}{2\sigma}\sum_t [y_t - \mu_t(\psi)]^2$$

with

$$\mu_t(\psi) = -\sum_{j=1}^{\infty} \psi^j y_{t-j},$$

and the (scaled) Lagrangian function by

$$\Lambda_T(\boldsymbol{\beta}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma - \frac{1}{2\sigma}\frac{1}{T}\sum_t [y_t - \mu_t(\psi)]^2 + \delta\lambda_\delta + \psi\lambda_\psi$$

where $\boldsymbol{\mu} = (□_\delta □ □_\psi)'$ are the multipliers associated with the (in)equality restrictions □ ≤ 0 and □ ≥ 0 respectively. Therefore, the sample first-order conditions will be:

$$\begin{aligned} \frac{1}{\bar{□}_T^r} \frac{1}{□} \sum_t □_t(\bar{□}_T^r) \frac{□□_t(\bar{□}_T^r)}{□□} + \bar{□}_{\delta T}^r &= 0 \\ \frac{1}{2\bar{□}_T^r} \frac{1}{□} \sum_t \left[\frac{□_t^□(\bar{□}_T)}{\bar{□}_T^r} - 1 \right] + \bar{□}_{\psi T}^r &= 0 \end{aligned}$$

together with the exclusion constraints

$$\begin{aligned} \bar{□}_T^r \cdot \bar{□}_{\delta T}^r &= 0 \\ \bar{□}_T^r \cdot \bar{□}_{\psi T}^r &= 0 \end{aligned}$$

plus the appropriate (in)equality restrictions, where

$$□_t(□) = \sum_{j□□}^{\infty} □^j □_{t-j} \tag{21}$$

$$\frac{□□_t(□)}{□□} = -\sum_{j□□}^{\infty} □□^{j-□} □_{t-j} \tag{22}$$

But as

$$\bar{□}_T^r = \frac{1}{□} \sum_t □_t^□(\bar{□}_T^r) \geq 0$$

we can safely take $\bar{□}_{\psi T}^r = 0$ in what follows. Also since

$$\bar{□}_{\delta T}^r = -\frac{1}{\bar{□}_T^r} \frac{1}{□} \sum_t □_t(\bar{□}_T^r) \frac{□□_t(\bar{□}_T^r)}{□□}$$

we can interpret this multiplier as (minus) the coefficient in the OLS regression of $□□_t(\bar{□}_T^r)□□□$ on the "restricted" residuals $□_t(\bar{□}_T^r)$ (see Gourieroux, Holly and Monfort, 1980). Therefore, $\bar{□}_{\delta T}^i$ will be 0 if the inequality restriction is satisfied, or the usual Lagrange multiplier associated with the *equality* constraint □ $= 0$ otherwise. Not surprisingly, the Lagrange multiplier is simply

$$\bar{□}_{\delta T}^e = -\frac{□^{-□} \sum_t □_t □_{t-□}}{□^{-□} \sum_t □_t^□} = -\frac{\bar{□}_{□□T}}{\bar{□}_{□□T}}$$

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

$$\bar{□}_T^e = \frac{1}{□}\sum_t □_t^{□} = \bar{□}^e_{□□T}$$

i.e. the sample variance with denominator □.

3.2.3 Population moments and binding functions

If we now define

$$\mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta}) = □\left[□_t(\boldsymbol{\theta})\right|\boldsymbol{\rho}] = -\frac{1}{2}\ln 2□ - \frac{1}{2}\ln □ - \frac{1}{2□}□\left\{□_t^{□}(□)\right|\boldsymbol{\rho}\}$$

it is clear that the binding functions $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ satisfy the moment conditions

$$\begin{aligned}\mathfrak{m}_\delta\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho})\right] &= \mathbf{0}\\ \mathfrak{m}_\psi\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho})\right] &= \mathbf{0}\end{aligned}$$

together with the exclusion restrictions

$$\begin{aligned}□^r(\boldsymbol{\rho})\cdot □_\delta^r(\boldsymbol{\rho}) &= 0\\ □^r(\boldsymbol{\rho})\cdot □_\psi^r(\boldsymbol{\rho}) &= 0\end{aligned}$$

plus the appropriate (in)equality restrictions on $\boldsymbol{\theta}^r(\boldsymbol{\rho})$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho})$, where

$$\begin{aligned}\mathfrak{m}_\delta(\boldsymbol{\rho};\boldsymbol{\beta}) &= □\left[\frac{1}{□}□_t(□)\frac{□□_t(□)}{□□} + □_\delta\right|\boldsymbol{\rho}\Bigg]\\ \mathfrak{m}_\psi(\boldsymbol{\rho};\boldsymbol{\beta}) &= □\left\{\frac{1}{2□}\left[\frac{□_t^{□}(□)}{□} - 1\right] + □_\psi\right|\boldsymbol{\rho}\Bigg\}\end{aligned} \tag{23}$$

Using the results in the appendix, we can write (23) as

$$\mathfrak{m}_\delta(\boldsymbol{\rho};\boldsymbol{\beta}) = -\frac{□_□(\boldsymbol{\rho})}{□(1-□^{□})^{□}}\left\{□ + \sum_{l□□}^{\infty}\left[2□^{□} + \left(1-□^{□}\right)□\right]□^{l-□}\frac{□_l(\boldsymbol{\rho})}{□_□(\boldsymbol{\rho})}\right\} + □_\delta$$

$$= \frac{\square}{\square(1-\square^{\square})\left(1-\square^{\square}\right)^{\square}(1-\square\square)^{\square}}\left(\square^{\square}\square^{\square}+\square\square^{\square}-\square-\square\right)+\square_{\delta}$$

$$\mathfrak{m}_{\psi}(\boldsymbol{\rho};\boldsymbol{\beta}) = \frac{1}{2\square^{\square}}\left[\frac{\square_{\square}(\boldsymbol{\rho})}{1-\square^{\square}}\left(1+2\sum_{l\square\square}^{\infty}\square^{l}\frac{\square_{l}(\boldsymbol{\rho})}{\square_{\square}(\boldsymbol{\rho})}\right)-\square\right]+\square_{\psi}$$

$$= \frac{1}{2\square^{\square}}\left[\frac{\square}{\left(1-\square^{\square}\right)\left(1-\square^{\square}\right)}\cdot\frac{1-\square\square}{1+\square\square}-\square\right]+\square_{\psi}$$

where the intermediate expressions are valid whatever the true model, while the final expressions are obtained by replacing (20) in the intermediate ones.

From here, it is easy to see that

$$\square^{r}(\boldsymbol{\rho}) = \square\left\{\square_{t}^{\square}\left[\square^{r}(\boldsymbol{\rho})\right]\middle|\,\boldsymbol{\rho}\right\} = \frac{\square_{\square}(\boldsymbol{\rho})}{1-\left[\square^{i}(\boldsymbol{\rho})\right]^{\square}}\left(1+2\sum_{l\square\square}^{\infty}\left[\square^{r}(\boldsymbol{\rho})\right]^{l}\frac{\square_{l}(\boldsymbol{\rho})}{\square_{\square}(\boldsymbol{\rho})}\right)\geq 0$$

and consequently, that $\square_{\psi}^{r}(\boldsymbol{\rho}) = 0$.

From the above moment expressions, we also have that the unconstrained binding function for $\square$, $\square^{u}(\boldsymbol{\rho})$, will be the real root of the following third order equation

$$\square^{\square}\left[\square^{u}(\boldsymbol{\rho})\right]^{\square}+\square\left[\square^{u}(\boldsymbol{\rho})\right]^{\square}-\square^{u}(\boldsymbol{\rho})-\square = 0$$

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

As a result, if $\square^{u}(\boldsymbol{\rho})\leq 0$, then $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}) = \boldsymbol{\beta}^{u}(\boldsymbol{\rho})$, where

$$\square^{u}(\boldsymbol{\rho}) = \frac{\square}{\left\{1-\left[\square^{u}(\boldsymbol{\rho})\right]^{\square}\right\}\left(1-\square^{\square}\right)}\cdot\frac{1-\square^{u}(\boldsymbol{\rho})\square}{1+\square^{u}(\boldsymbol{\rho})\square}$$

$$\square_{\delta}^{u}(\boldsymbol{\rho}) = 0$$

are the remaining unconstrained binding functions, while if $\square^{u}(\boldsymbol{\rho})\ \square\ 0$, then $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}) = \boldsymbol{\beta}^{e}(\boldsymbol{\rho})$, where

$$\square^{e}(\boldsymbol{\rho}) = 0$$

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

$$\square^e(\boldsymbol{\rho}) = \square_\square(\boldsymbol{\rho}) = \frac{\square}{1-\square^\square} \tag{24}$$
$$\square^e_\delta(\boldsymbol{\rho}) = -\frac{\square_\square(\boldsymbol{\rho})}{\square_\square(\boldsymbol{\rho})} = -\square \geq 0$$

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

Figure 2 plots the binding functions $\square^u(\boldsymbol{\rho})$ and $\square^e_\delta(\boldsymbol{\rho})$ for $-1 \square \square \square 1$. Note that in this framework, $\square^i(\boldsymbol{\rho}) = \min[\square^u(\boldsymbol{\rho})\square 0]$ while $\square^i_\delta(\boldsymbol{\rho}) = \max[\square^e_\delta(\boldsymbol{\rho})\square 0]$.

3.2.4 Asymptotic distributions of pseudo-ML estimators and sample moments

First of all, let us 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 $\square_t$ is given by the Gaussian* AR*(1) model (19), $\bar{\square}_{\square\square T}$ and $\bar{\square}_{\square\square T}$ are $\square^{\square/\square}$-consistent for $\square_\square(\boldsymbol{\rho}^\square)$ and $\square_\square(\boldsymbol{\rho}^\square)$ in (20) respectively, with the following limiting distribution*

$$\sqrt{\square}\left[\bar{\boldsymbol{\gamma}}_T - \square(\boldsymbol{\rho}^\square)\right] \xrightarrow{d} \square\left[0\square \mathbf{V}(\boldsymbol{\rho}^\square)\right]$$

where

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

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

$$\mathfrak{m}_\delta(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^e) = -\frac{\square}{\bar{\square}_{\square\square T}(1-\square^\square)}\square - \frac{\bar{\square}_{\square\square T}}{\bar{\square}_{\square\square T}}$$

$$\mathfrak{m}_\psi(\boldsymbol{\rho};\bar{\boldsymbol{\beta}}^e) = \frac{1}{2\bar{\square}_{\square\square T}}\left[\frac{\square}{(1-\square^\square)} - \bar{\square}_{\square\square T}\right]$$

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 $\bar{\Box}^e_{\delta T} = -\bar{\Box}_{\Box\Box T}\Box\bar{\Box}_{\Box\Box T}$ and $\bar{\Box}^e_T = \bar{\Box}_{\Box\Box T}$. 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 $\bar{\Box}^u_T$ and $\bar{\Box}^u_T$ is rather more laborious to obtain, as we need to derive closed form expressions for the matrices $\mathcal{I}^u_{\Box}$ and $\mathcal{J}^u_{\Box}$. For simplicity, we shall only do it for the case of $\Box^{\Box} = 0$, which as we saw before, corresponds to $\Box^u(\boldsymbol{\rho}^{\Box}) = 0$ and $\Box^u(\boldsymbol{\rho}^{\Box}) = \Box^{\Box}$. 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:

$$\frac{1}{\Box^{\Box}}\frac{1}{\Box}\sum_t \Box_t\Box_{t-\Box} = \frac{1}{\Box^{\Box}}\bar{\Box}_{\Box\Box T}$$

$$\frac{1}{2\Box^{\Box}}\frac{1}{\Box}\sum_t\left(\frac{\Box^{\Box}_t}{\Box^{\Box}} - 1\right) = \frac{1}{2(\Box^{\Box})^{\Box}}\left[\bar{\Box}_{\Box\Box T} - \Box^{\Box}\right]$$

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

$$\mathcal{I}^u_{\Box} = \begin{bmatrix} 1 & 0 \\ 0 & 1\Box 2(\Box^{\Box})^{\Box} \end{bmatrix}$$

Similarly, it is also easy to prove that for $\boldsymbol{\rho}^{\Box} = (0\Box\,\Box^{\Box})'$

$$\mathcal{J}^u_{\Box} = \begin{bmatrix} 1 & 0 \\ 0 & 1\Box 2(\Box^{\Box})^{\Box} \end{bmatrix}$$

so that

$$\sqrt{\Box}\begin{pmatrix} \bar{\Box}^u_T \\ \bar{\Box}^u_T - \Box^{\Box} \end{pmatrix} \xrightarrow{d} \Box\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\Box\begin{bmatrix} 1 & 0 \\ 0 & 2(\Box^{\Box})^{\Box} \end{bmatrix}\right\}$$

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 $\Box$, $\Box$, and $\Box_\delta$, there may be three different situations, according to whether $\Box^{\Box} \Box 0$, $\Box^{\Box} \Box 0$ or $\Box^{\Box} = 0$. In the first case, it is easy to see from Propositions 1 and 2 that $\sqrt{\Box}(\bar{\Box}^i_T - \bar{\Box}^u_T)$, $\sqrt{\Box}(\bar{\Box}^i_T - \bar{\Box}^u_T)$ and $\sqrt{\Box}\,\Box_{\delta T}$ are all $\Box_p(1)$, while in the second case the same applies to $\sqrt{\Box}\,\bar{\Box}^i_T$, $\sqrt{\Box}(\bar{\Box}^i_T - \bar{\Box}^e_T)$ and $\sqrt{\Box}(\Box^i_T - \Box^e_T)$. Once more, the most interesting case arises when $\Box^{\Box} = 0$, because $\sqrt{\Box}\,\bar{\Box}^i_T$ and $\sqrt{\Box}\,\Box_{\delta T}$ have half normal asymptotic distributions. Nevertheless, from Proposition 2 we will again have that $\sqrt{\Box}(\bar{\Box}^i_T - \Box^i_{\delta T})$ will share an asymptotic $\Box(0\Box 1)$ distribution with $\sqrt{\Box}(\bar{\Box}^u_T - \Box^u_{\delta T}) = \sqrt{\Box}\,\bar{\Box}^u_T$ and $\sqrt{\Box}(\bar{\Box}^e_T - \Box^e_{\delta T}) = -\sqrt{\Box}\,\Box^e_{\delta T}$.

3.2.5 Indirect inference estimators

Given the two different expressions for the *inequality restricted* pseudo-ML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ discussed previously, the sample counterparts to the population moments (23) will be given by either $\mathfrak{m}(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^u_T)$, which correspond to the sample moments used by an *unrestricted* GMM-based II procedure, or $\mathfrak{m}(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^e_T)$, which will be the moments used by the *equality constrained* one. But since when we solve for $\boldsymbol{\rho}$ from the system of equations $\mathfrak{m}(\boldsymbol{\rho}; \bar{\boldsymbol{\beta}}^e_T) = 0$ we get

$$\begin{aligned} \tilde{\Box}^e_T &= -\bar{\Box}^e_{\delta T} = \frac{\bar{\Box}_{\Box\Box T}}{\bar{\Box}_{\Box\Box T}} \\ \Box^e_T &= \bar{\Box}^e_T \left[1 - \left(\tilde{\Box}^e_T\right)^{\Box}\right] = \bar{\Box}_{\Box\Box T} - \frac{\bar{\Box}^{\Box}_{\Box\Box T}}{\bar{\Box}_{\Box\Box T}} \end{aligned}$$

it is clear that the *equality constrained* II estimator converges in probability to the maximum likelihood estimator of the parameters of interest. Note that this is true regardless of the sign of $\Box^u(\boldsymbol{\rho}^{\Box})$, and therefore independently of whether or not the restriction $\Box = 0$ is correct. Of course, if we knew that $\Box^u(\boldsymbol{\rho}^{\Box}) = 0$, or any other value for that matter, we could recover $\Box^{\Box}$ from the binding function directly without estimation error (cf. Dridi, 2000). Given that the auxiliary model exactly identifies the parameters of interest, the same result applies to the corresponding *equality constrained* CMD II estimators, which minimise the objective function

$$\square^e_T(\boldsymbol{\rho};\mathbf{I}_\square) = \left[\square^e_{\delta T} + \frac{\square_\square(\boldsymbol{\rho})}{\square_\square(\boldsymbol{\rho})}\right]^\square + \left[\bar{\square}^e_T - \square_\square(\boldsymbol{\rho})\right]^\square$$

The reason for this seemingly counterintuitive result is that $\square^e_{\delta T}$ and $\bar{\square}^e_T$ are sufficient statistics for the true AR(1) model, so that Proposition 11 applies.

As for the *inequality restricted* estimators, it depends on whether or not the pseudo-true value $\square^i(\boldsymbol{\rho}^\square)$ is 0 or strictly negative (or the associated Kuhn-Tucker multiplier $\square^i_\delta(\boldsymbol{\rho}^\square)$ is 0 or strictly positive). If $\square^\square \; \square \; 0$, then $\tilde{\boldsymbol{\rho}}^i_T$ will be asymptotically equivalent to the unrestricted estimator $\tilde{\boldsymbol{\rho}}^u_T$ because the sign restriction on $\tilde{\square}_T$ is not binding in large samples, as predicted by the first part of Proposition 10. As a result, the *inequality restricted* estimators will be less efficient than the *equality constrained* ones. If on the other hand, $\square^\square \; \square \; 0$, the restriction is almost certainly binding in the limit, and therefore $\tilde{\boldsymbol{\rho}}^i_T$ will be asymptotically equivalent to the equality restricted estimator $\tilde{\boldsymbol{\rho}}^e_T$, as predicted by the second part of Proposition 10. Finally, since the unrestricted pseudo log-likelihood nests the true log-likelihood when $\square^\square = 0$, the *unrestricted* estimators will also be as efficient as ML by virtue of Theorem 2 in GT. But since the *inequality restricted estimators* will be a 50:50 mixture of $\tilde{\boldsymbol{\rho}}^u_T$ and $\tilde{\boldsymbol{\rho}}^e_T$ in large samples, it will share their common asymptotic distribution, as indicated by the last statement in Proposition 10.

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

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

$$
\begin{aligned}
y_t &= \sqrt{h_t}\varepsilon_t \\
\ln h_t &= \omega + \alpha \ln h_{t-1} + \sigma_v \xi_t
\end{aligned}
\qquad (25)
$$

where $|\alpha| < 1, 0 < \sigma_v < \infty$, and $(\varepsilon_t, \xi_t)' | I_{t-1}; \rho \sim N(0, \mathbf{I}_2)$. 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 y_t based on its own past values alone, despite the fact that its distribution conditional on $h_t, I_{t-1}; \rho$ 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 $\boldsymbol{\rho} = (\omega, \alpha, \sigma_v)'$ (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 most general model that we will estimate is given by

$$
\begin{aligned}
y_t &= \sqrt{h_t}\varepsilon_t \\
h_t &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1}
\end{aligned}
$$

where $\varepsilon_t|I_{t-1};\boldsymbol{\theta}$ follows a standardised Student's t distribution with η^{-1} degrees of freedom,[2] so that $\boldsymbol{\theta} = (\omega, \alpha, \beta, \eta)'$. As is well known, the standardised t distribution nests the standard normal for $\eta = 0$, but has otherwise 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 r_t is Gaussian white noise.

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 σ_t^2 will be nonnegative with probability one if $\omega \geq 0$, $\alpha \geq 0$ and $\beta \geq 0$.

2. The pseudo-ML estimators of $\boldsymbol{\theta}$ may not be well behaved when $\alpha + \beta \geq 1$ (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 $\alpha = 0$, β 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:

$$\omega \geq 0, \quad \alpha \geq \varphi_{\min}, \quad \beta \geq 0, \quad \alpha + \beta \leq 1, \quad 0 \leq \eta \leq \eta_{\max} \tag{26}$$

where $\varphi_{\min}$, and $1/2 - \eta_{\max}$ are arbitrarily chosen small values.[3]

[2] 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 mean $1/\eta$ and variance $2/\eta$.

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

Unfortunately, the tail-thickness parameter η is often very imprecisely estimated even if the sample size is reasonably large. This is due to the fact that the log-likelihood function 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. 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 takes into account the value of the corresponding multiplier from the relevant first order condition. Finally, we consider a third estimator that is also based on the Gaussian pseudo log-likelihood function, but which discards the information in the multiplier, as discussed in section 2.2.1. For the sake of brevity, we shall refer to the estimator that allows η to vary freely within its bounds as the "inequality restricted" II estimator, to the second one as the "equality restricted" II estimator, and to the third one as the "unrestricted" II estimator. In all three cases, though, 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 σ_t, κ say, where

$$\kappa^2 = \frac{V(\sigma_t)}{E^2(\sigma_t)} = \exp(\frac{\sigma_v^2}{1-\rho^2}) - 1$$

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 □$^{□}$'□ (around □3) 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 □$^{□}$, and the columns by the autocorrelation coefficient for log volatility, □. Finally, the remaining parameter □ 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 □ = 500 observations, we have also considered □ = 1□000 and 2□000.

For convenience, we first optimise the pseudo log-likelihood function in terms of some unrestricted parameters $\boldsymbol{\theta}^*$, where □ = □$^{*□}_{□}$, □ = □$_{□□□}$ + (1 − □$_{□□□}$) sin$^{□}$(□$^{*}_{□}$), □ = (1 − □) sin$^{□}$(□$^{*}_{□}$) and □ = sin$^{□}$(□$^{*}_{□}$)□$_{□□□}$. Then, we compute the score in terms of the original parameters $\boldsymbol{\theta}$ = (□□ □□ □□ □)$'$ using the analytical expressions derived by Fiorentini, Sentana and Calzolari (2000) to avoid large numerical errors, 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 □□, with □ = 10, as explained in the appendix. A larger value of □ should in theory reduce the Monte Carlo variability of the II estimators according to the relation (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 $\boldsymbol{\rho}^*$, with □ = □$^{*}_{□}$, □ = □$_{□□□}$ sin(□$^{*}_{□}$) and □$_v$ = □$^{*□}_{□}$, where □$_{□□□}$ = □9999, so as to ensure that |□| □ 1 and □$_v \geq 0$.

Tables 1, 2 and 3 contain the proportion of inequality and equality restricted pseudo-ML estimators of $\boldsymbol{\theta}$ that satisfy with equality the different restrictions in (26). In this respect, note that the auxiliary model estimated by the unrestricted procedure coincides with the model estimated by the equality restricted one.

When $□^{□}$ is 1, such restrictions are hardly ever binding, especially for $□ = 2□000$. However, when $□^{□}$ is rather large (=10), most of the estimated GARCH models are of the IGARCH variety. This is particularly true when □ 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 situations in which $□^{□}$ is very 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 $□ = 500$ and $□ = □98$, 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 II procedures are not only of theoretical interest, but also highly relevant in practice.

Tables 4 to 12 present the means, root mean square errors, mean biases and standard deviations of the "unrestricted", "equality restricted", and "inequality restricted" GMM-based II estimators of the parameters of interest $\boldsymbol{\rho}$ for the case in which the optimal 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.[4] In this respect, note that by including a multiplier in each

[4] Note that since the "unrestricted" II estimator is effectively using a just-identified auxiliary model, it is invariant to the weighting matrix. Nevertheless, by using the optimal weighting

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, in the sense that the score of the auxiliary model is close to being a vector martingale difference sequence, we have not included any correction for serial correlation (cf. GT).

In line with the existing literature, we find that the different estimators of the autoregressive parameter □ are systematically downward biased. This is particularly so when $□^{□}$ is high, and/or $□_v^{□}$ 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 □ is transmitted into an upward bias in the absolute value of the estimates of the mean constant, □, and the standard deviation of the log-volatility innovations $□_v$. Therefore, it is not surprising that the most important determinant of the performance of the estimators is precisely $□^{□}$, which effectively plays the role of a signal to noise ratio.

As for the comparison between the "unrestricted" and "equality restricted" II estimators, the most noticeable effect of taking into account the information in the multiplier associated with the zero constraint on □ is that the precision with which we estimate the volatility of volatility parameter, $□_v$, increases substantially, the more so the smaller the signal to noise ratio. This is due to the fact that $□_v$ is the parameter that most directly determines the degree of leptokurtosis of the conditional distribution of $□_t$, which is mainly captured in the GARCH model through the value of □, or its associated multiplier. With respect to the other structural parameters, the reported simulation evidence also confirms the result stated in Proposition 7, with the exception of □ = □1, where large sample sizes are required for the asymptotics to apply.

matrix, we ensure that the objective function is evenly scaled across parameters, which improves the numerical properties of the optimisation algorithm.

In contrast, neither of the two restricted versions of the II estimator seems to dominate the other across all Monte Carlo designs. When κ° is 10, the inequality restricted II 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° is .1, the equality restricted II 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, the auxiliary parameter η is poorly determined. As a result, our attempts to estimate simultaneously the reciprocal of the degrees of freedom result in a deterioration of the estimators of the GARCH parameters relative to the Gaussian case. At the same time, since as we explained above the first order condition for η is the most directly related to the degree of leptokurtosis of the observed data, the equality restricted II 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 II estimator tends to have a smaller root mean square error than the equality restricted one, at the cost of a slightly higher mean bias. In this respect, please note that the mixed conclusions that we obtain from these simulations are partly the result of the pseudo-true values of η (computed on the basis of 500,000 observations) being systematically different from zero (cf. Propositions 8 and 10). In addition, they also reflect the fact that the accuracy with which one can estimate the equality and inequality restricted auxiliary parameters crucially depends on the values of the true parameters. In practice, of course, an applied econometrician will often know the range of values of $\boldsymbol{\rho}$ that she is likely to obtain with her data, in which case the ambiguity disappears.

Finally, a comparison of our results with the ones reported by JPR suggests that our II 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, our II 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 II estimators is partly due to those simulations in which [illegible] is estimated as being negative. For instance, the root mean square error of the equality restricted estimator of [illegible] in row 2, column 3 of Table 5 decreases from .0765 to .0524 if we exclude the only two negative estimates of [illegible] found in 1,000 replications.

4 Conclusions

In this paper, we generalise the II approaches of GT and GMR to those situations in which there are equality and/or inequality constraints on the parameters of the auxiliary model. Specifically, we propose an alternative set of moment restrictions based on the first order conditions for (in)equality restricted models, which nest the ones employed by GT when there are no constraints, or when they well as those that impose the constraints depending on the significance of some preliminary specification test.

Inequality restrictions must often be considered in practice because either the pseudo log-likelihood function may not be well defined when certain parameter restrictions are violated, or some of the auxiliary parameters may become underidentified in certain regions of the auxiliary parameter space. In addition, 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. In this respect, we show that the asymptotic efficiency of II estimators can never decrease by explicitly taking into account the Lagrange multipliers associated with additional equality constraints, regardless of whether such restrictions are correct. This result is particularly important in practice, as any parametric auxiliary model implicitly contains a vast number of maintained assumptions, which can often be written in terms of zero restrictions on some additional parameters.

We also illustrate the variety of effects that can be obtained when some constraints are imposed on the parameters of a previously unrestricted auxiliary model. For instance, we discuss several circumstances in which the imposition of constraints has no effect on the efficiency of the resulting II estimators, and others in which false constraints enable the restricted II estimators to achieve full efficiency. The reason for these seemingly counterintuitive results is that by adding restrictions to the auxiliary model in those circumstances in which they are not required to properly define the auxiliary objective function, we are implicitly changing the auxiliary model, and thereby, the binding functions.

For illustrative purposes, we discuss the usual example of MA(1) estimated as AR(1), and show that inequality restricted II estimators are asymptotically equivalent to the unrestricted estimators, and indeed, to equality restricted II estimators that set the autoregressive parameter in the auxiliary model to any arbitrary bounded value, but include either the corresponding first order condition in the set of moments, or the Lagrange multiplier in the distance function. Importantly, the equivalence of the different 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 II estimators that impose a white noise restriction not only dominate the unrestricted estimators, but also become as efficient as ML, even though the auxiliary model does not nest the true one, and the restriction is false. 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 document that when the auxiliary model is estimated under Gaussianity, we can increase the efficiency of the usual II estimators by including the information in the multiplier corresponding to the reciprocal of the degrees of freedom. Finally, we find that although neither version of the restricted II estimator systematically outperforms the other, replacing the pseudo-ML estimator of the tail-thickness parameter by its multiplier in those situations in which there is little information in the data about this auxiliary parameter results in more efficient estimators of the parameters of interest.

Appendix

Proofs of results

Proposition 1

If we linearise the complementary slackness conditions

$$\square(\bar{\boldsymbol{\theta}}_T^r) \odot \bar{\boldsymbol{\mu}}_T^r = 0$$

around $\boldsymbol{\beta}^r(\boldsymbol{\rho}^\square)$, taking into account that $\square\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right] \odot \boldsymbol{\mu}^r(\boldsymbol{\rho}^\square) = 0$, and that Hadamard products are commutative, we obtain:

$$\boldsymbol{\mu}_T^* \odot \frac{\square\square'(\boldsymbol{\theta}_T^*)}{\square\boldsymbol{\theta}} \sqrt{\square}\left[\bar{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right] + \square(\boldsymbol{\theta}_T^*) \odot \sqrt{\square}\left[\bar{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^\square)\right] = 0$$

where $\boldsymbol{\beta}_T^* = (\boldsymbol{\theta}_T^{*\prime}\square\,\boldsymbol{\mu}_T^{*\prime})'$ is an "intermediate" value (in fact, a different one for each row). Then, given that in view of our high level assumptions, $\boldsymbol{\mu}_T^* - \boldsymbol{\mu}^r(\boldsymbol{\rho}^\square) = \square_p(1)$, $\square(\boldsymbol{\theta}_T^*) - \square\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right] = \square_p(1)$, and $\square\square(\boldsymbol{\theta}_T^*)\square\square\boldsymbol{\theta} - \square\square\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\theta} = \square_p(1)$, the result follows. $\square$

Proposition 2

If we linearise the first-order conditions

$$\frac{\sqrt{\square}}{\square}\sum_t \square_t(\bar{\boldsymbol{\theta}}_T^r) = \mathbf{0}$$

around $\boldsymbol{\beta}_T^r(\boldsymbol{\rho}^\square)$, we obtain:

$$\frac{\sqrt{\square}}{\square}\sum_t \square_t\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right]$$

$$+\frac{1}{\square}\sum_t\left\{\frac{\square\square_t^\square(\boldsymbol{\theta}_T^\star)}{\square\boldsymbol{\theta}\square\boldsymbol{\theta}'} + (\boldsymbol{\mu}_T^\star \otimes \square_q)\frac{\square\square\square\square\left[\square\square'(\boldsymbol{\theta}_T^\star)\square\square\boldsymbol{\theta}\right]}{\square\boldsymbol{\theta}'}\right\}\sqrt{\square}\left[\bar{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^\square)\right]$$

$$+\frac{\square\square'(\boldsymbol{\theta}_T^\star)}{\square\boldsymbol{\theta}}\sqrt{\square}\left[\bar{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^\square)\right]$$

where $\boldsymbol{\beta}_T^{\star} = (\boldsymbol{\theta}_T^{\star\prime}\square\,\boldsymbol{\mu}_T^{\star\prime})'$ is another "intermediate" value. Then, since in view of Assumption 2

$$\frac{1}{\square}\sum_t \left\{\frac{\square\square_t^{\square}(\boldsymbol{\theta}_T^{\star})}{\square\boldsymbol{\theta}\square\boldsymbol{\theta}'}\right\} = \mathcal{J}_{\square}^{i} + \square_p(1)$$

$$(\boldsymbol{\mu}_T^{\star}\otimes\square_q)\frac{\square\square\square\square\left[\square\square'(\boldsymbol{\theta}_T^{\star})\square\square\boldsymbol{\theta}\right]}{\square\boldsymbol{\theta}'} = \left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square})\otimes\square_q\right]\frac{\square\square\square\square\left\{\square\square'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\theta}\right\}}{\square\boldsymbol{\theta}'} + \square_p(1)$$

$$\frac{\square\square'(\boldsymbol{\theta}_T^{\star})}{\square\boldsymbol{\theta}} = \frac{\square\square'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\theta}} + \square_p(1)$$

a straightforward application of Crámer's theorem completes the proof. $\square$

Proposition 3

Let us now linearise the sample moments $\mathfrak{m}(\boldsymbol{\rho}^{\square};\bar{\boldsymbol{\beta}}_T^r)$ around $\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})$ to obtain

$$\sqrt{\square}\,\mathfrak{m}(\boldsymbol{\rho}^{\square};\bar{\boldsymbol{\beta}}_T^r) = \sqrt{\square}\,\mathfrak{m}\left[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]$$

$$+\frac{\square\mathfrak{m}(\boldsymbol{\rho}^{\square};\boldsymbol{\beta}_T^{\diamond})}{\square\boldsymbol{\theta}'}\sqrt{\square}\left[\bar{\boldsymbol{\theta}}_T-\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]+\frac{\square\mathfrak{m}(\boldsymbol{\rho}^{\square};\boldsymbol{\beta}_T^{\diamond})}{\square\boldsymbol{\mu}'}\sqrt{\square}\left[\bar{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square})\right]$$

where $\boldsymbol{\beta}_T^{\diamond}$ is yet another "intermediate" value. This implies that under Assumption 2, $\sqrt{\square}\,\mathfrak{m}(\boldsymbol{\rho}^{\square};\bar{\boldsymbol{\beta}}_T^r)$ has the same asymptotic distribution as

$$\frac{\square\mathfrak{m}'\left[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\theta}}\sqrt{\square}\left[\bar{\boldsymbol{\theta}}_T^r-\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]+\frac{\square\mathfrak{m}'\left[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\mu}}\sqrt{\square}\left[\bar{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square})\right]$$

where

$$\frac{\square\mathfrak{m}\left[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\theta}'} = \mathcal{J}_{\square}^{r}+\left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^{\square})\otimes\square_q\right]\frac{\square\square\square\square\left\{\square\square'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\theta}\right\}}{\square\boldsymbol{\theta}'}=\mathcal{K}_{\square\square,\square}^{r}$$

$$\frac{\square\mathfrak{m}\left[\boldsymbol{\rho}^{\square};\boldsymbol{\beta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\mu}'} = \frac{\square\square'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^{\square})\right]}{\square\boldsymbol{\theta}}=\mathcal{K}_{\square\square,\square}^{r}$$

But then, Proposition 2 directly yields the required result $\square$

Proposition 4

The first order conditions associated with $\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]$ can be written as

$$\frac{\partial \mathfrak{m}'\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right];\bar{\boldsymbol{\beta}}_T^r\right\}}{\partial \boldsymbol{\rho}}\cdot(\mathcal{I}_0^r)^{-1}\cdot\sqrt{T}\mathfrak{m}\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right];\bar{\boldsymbol{\beta}}_T^r\right\}=\mathbf{0}$$

Expanding around $\boldsymbol{\rho}^0$ yields

$$\begin{aligned}&\frac{\partial \mathfrak{m}'\left(\boldsymbol{\rho}^0;\bar{\boldsymbol{\beta}}_T^r\right)}{\partial \boldsymbol{\rho}}\cdot(\mathcal{I}_0^r)^{-1}\cdot\sqrt{T}\mathfrak{m}\left(\boldsymbol{\rho}^0;\bar{\boldsymbol{\beta}}_T^r\right)\\&+\frac{\partial \mathfrak{m}'\left(\boldsymbol{\rho}_T^*;\bar{\boldsymbol{\beta}}_T^r\right)}{\partial \boldsymbol{\rho}}\cdot(\mathcal{I}_0^r)^{-1}\cdot\frac{\partial \mathfrak{m}\left(\boldsymbol{\rho}^*;\bar{\boldsymbol{\beta}}_T^r\right)}{\partial \boldsymbol{\rho}}\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]-\boldsymbol{\rho}^0\right\}\\&+\left[(\mathcal{I}_0^r)^{-1}\cdot\mathfrak{m}\left(\boldsymbol{\rho}_T^*;\bar{\boldsymbol{\beta}}_T^r\right)\otimes I_d\right]\frac{\partial\,\mathrm{vec}\left[\partial\mathfrak{m}'(\boldsymbol{\rho}_T^\star;\bar{\boldsymbol{\beta}}_T^r)/\partial\boldsymbol{\rho}\right]}{\partial\boldsymbol{\rho}'}\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]-\boldsymbol{\rho}^0\right\}\end{aligned}$$

where $\boldsymbol{\rho}_T^*$ is some "intermediate" value. But since $\mathfrak{m}\left(\boldsymbol{\rho}_T^*;\bar{\boldsymbol{\beta}}_T^r\right)$ is $o_p(1)$, and $\partial\mathfrak{m}\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\rho}'$ has full column rank, we finally have that

$$\begin{aligned}\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]-\boldsymbol{\rho}^0\right\}&=\left\{\frac{\partial \mathfrak{m}'\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\rho}}\cdot(\mathcal{I}_0^r)^{-1}\cdot\frac{\partial \mathfrak{m}\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\rho}'}\right\}^{-1}\\&\quad\times\frac{\partial \mathfrak{m}'\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\rho}}(\mathcal{I}_0^r)^{-1}\sqrt{T}\mathfrak{m}\left(\boldsymbol{\rho}^0;\bar{\boldsymbol{\beta}}_T^r\right)+o_p(1)\end{aligned}$$

as required. □

Proposition 5

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

$$\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0;\bar{\boldsymbol{\beta}}_T^r)-\left\{\mathcal{K}_{00,0}^r\sqrt{T}\left[\bar{\boldsymbol{\theta}}_T^r-\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]+\mathcal{K}_{00,0}^i\sqrt{T}\left[\bar{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]\right\}=o_p(1)$$

Proposition 6

By definition, $\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})$ must always satisfy the first-order equations:

$$\frac{\partial \mathfrak{m}'\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi});\bar{\boldsymbol{\beta}}_T^r\right]}{\partial\boldsymbol{\rho}}\cdot\boldsymbol{\Psi}\cdot\mathfrak{m}\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi});\bar{\boldsymbol{\beta}}_T^r\right]=\mathbf{0}.$$

If $\square = \square$ and $\square$ is large enough, though, our assumptions imply that $\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})$ will in fact be the solution to the system of equations

$$\mathfrak{m}\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi});\bar{\boldsymbol{\beta}}_T^r\right] = \mathbf{0}$$

independently of $\boldsymbol{\Psi}$. But since

$$\mathfrak{m}\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi});\bar{\boldsymbol{\beta}}_T^r\right] = \square\left[\square_t\left(\bar{\boldsymbol{\beta}}_T^r\right)\middle|\,\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})\right]\square$$

the first order conditions that characterise the binding functions imply that

$$\boldsymbol{\beta}^r\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})\right] - \bar{\boldsymbol{\beta}}_T^r = 0,$$

which means that $\boldsymbol{\beta}^r\left[\tilde{\boldsymbol{\rho}}_T^r(\boldsymbol{\Psi})\right]$ trivially minimises $\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \bar{\boldsymbol{\beta}}_T^r\right]' \cdot \boldsymbol{\Omega} \cdot \left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \bar{\boldsymbol{\beta}}_T^r\right]$ for any $\boldsymbol{\Omega}$.

Proposition 7

Given our assumptions about the relationship between $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$, we can use the results of Ferguson (1958) to show that $\hat{\boldsymbol{\rho}}_T^e\left[\mathcal{K}_\square^{r\prime}\left(\mathcal{I}_\square^{\boxplus}\right)^{\square}\mathcal{K}_\square^r\right]$ is asymptotically equivalent to the optimal CMD II estimator of $\boldsymbol{\rho}$ that minimises the distance between the "sample statistics" $(\bar{\boldsymbol{\pi}}_{\square T}^{u\prime}\square\mathbf{0}\square\bar{\boldsymbol{\mu}}_{\square T}^{e\prime})'$ and their population counterparts $\left[\boldsymbol{\pi}_\square^{u\prime}(\boldsymbol{\rho})\square\mathbf{0}\square\boldsymbol{\mu}_\square^{e\prime}(\boldsymbol{\rho})\right]'$, where $\bar{\boldsymbol{\mu}}_{\square T}^e$ are the Lagrange multipliers associated with the equality constraints $\square(\boldsymbol{\theta}) = \boldsymbol{\pi}_\square = \mathbf{0}$, and $\boldsymbol{\mu}_\square^e(\boldsymbol{\rho})$ the corresponding binding functions. But such an estimator is at least as efficient as the optimal CMD II estimator of $\boldsymbol{\rho}$ that simply minimises the distance between $\bar{\boldsymbol{\pi}}_{\square T}^u$ and $\boldsymbol{\pi}_\square^u(\boldsymbol{\rho})$.

Proposition 8

If the equality constraints are satisfied by the unrestricted pseudo-true values of $\boldsymbol{\theta}$, in the sense that $\square\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^\square)\right] = 0$, then $\boldsymbol{\theta}^u(\boldsymbol{\rho}^\square) = \boldsymbol{\theta}^e(\boldsymbol{\rho}^\square)$, $\boldsymbol{\mu}^u(\boldsymbol{\rho}^\square) = \boldsymbol{\mu}^e(\boldsymbol{\rho}^\square) = \mathbf{0}$, and $\square_t\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right] = \square_t\left[\boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)\right]\ \forall\square$. As a result, $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right] = \mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)\right]$ for all $\boldsymbol{\rho}$ in a neighbourhood of $\boldsymbol{\rho}^\square$, so that $\square\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\rho} = \square\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\rho}$. For analogous reasons, $\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}_0, \bar{\boldsymbol{\beta}}_T^e\right] - \sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}_0, \bar{\boldsymbol{\beta}}_T^u\right] = O_p(1)$ in view of Proposition 3, so that $\mathcal{I}_0^u = \mathcal{I}_0^e$. The required result then follows from Proposition 4.

Proposition 9

For simplicity of notation, let us define $\mathbf{z}_t = (x_{t-1}, \ldots, x_{t-k})'$, $\gamma_{xx}(\boldsymbol{\rho}) = E\left(x_t^2 \middle| \boldsymbol{\rho}\right)$, $\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}) = E\left(\mathbf{z}_t x_t \middle| \boldsymbol{\rho}\right)$ and $\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho}) = E\left(\mathbf{z}_t \mathbf{z}_t' \middle| \boldsymbol{\rho}\right)$. It is then straightforward to see that

$$
\begin{aligned}
\mathfrak{m}_\phi(\boldsymbol{\rho}, \boldsymbol{\beta}) &= \frac{1}{\omega}\left[\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}) - \boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}\right] + \mathbf{R}'\boldsymbol{\mu} \\
\mathfrak{m}_\omega(\boldsymbol{\rho}, \boldsymbol{\beta}) &= \frac{1}{2\omega^2}\left[\gamma_{xx}(\boldsymbol{\rho}) + \boldsymbol{\phi}'_{zz}\boldsymbol{\Sigma}(\boldsymbol{\rho})\boldsymbol{\phi} - 2\boldsymbol{\sigma}'_{zx}(\boldsymbol{\rho})\boldsymbol{\phi} - \omega\right]
\end{aligned}
$$

from where we can obtain the following binding functions

$$
\begin{aligned}
\boldsymbol{\phi}^u(\boldsymbol{\rho}) &= \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}) \\
\boldsymbol{\mu}^u(\boldsymbol{\rho}) &= \mathbf{0} \\
\omega^u(\boldsymbol{\rho}) &= \gamma_{xx}(\boldsymbol{\rho}) - \boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})'\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\phi}^e(\boldsymbol{\rho}) &= \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}) + \boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\mathbf{R}'\left[\mathbf{R}\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\mathbf{R}'\right]^{-1}\left[\mathbf{r} - \mathbf{R}\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})\right] \\
\boldsymbol{\mu}^e(\boldsymbol{\rho}) &= \frac{1}{\omega^e(\boldsymbol{\rho})}\left[\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\mathbf{R}'\right]^{-1}\left[\mathbf{r} - \mathbf{R}\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho})\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})\right] \\
\omega^e(\boldsymbol{\rho}) &= \gamma_{xx}(\boldsymbol{\rho}) + \boldsymbol{\phi}^e(\boldsymbol{\rho})'\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^e(\boldsymbol{\rho}) - 2\boldsymbol{\sigma}'_{zx}(\boldsymbol{\rho})\boldsymbol{\phi}^e(\boldsymbol{\rho})
\end{aligned}
$$

Therefore, we will have that

$$
\begin{aligned}
\mathfrak{m}_\phi\left[\boldsymbol{\rho}, \boldsymbol{\beta}^u(\boldsymbol{\rho}_0)\right] &= \frac{\gamma_{zx}(\boldsymbol{\rho}) - \boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho}_0)\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}_0)}{\omega^u(\boldsymbol{\rho}_0)} \\
\mathfrak{m}_\omega\left[\boldsymbol{\rho}, \boldsymbol{\beta}^u(\boldsymbol{\rho}_0)\right] &= \frac{\gamma_{xx}(\boldsymbol{\rho}) + \boldsymbol{\phi}^u(\boldsymbol{\rho}_0)'\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^u(\boldsymbol{\rho}_0) - 2\boldsymbol{\sigma}'_{zx}(\boldsymbol{\rho})\boldsymbol{\phi}^u(\boldsymbol{\rho}_0) - \omega^u(\boldsymbol{\rho}_0)}{2\left[\omega^u(\boldsymbol{\rho}_0)\right]^2}
\end{aligned}
$$

and

$$
\mathfrak{m}_\phi\left[\boldsymbol{\rho}, \boldsymbol{\beta}^e(\boldsymbol{\rho}_0)\right] = \frac{\left[\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}) - \boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^e(\boldsymbol{\rho}_0)\right]}{\omega^e(\boldsymbol{\rho}_0)} + \omega^e(\boldsymbol{\rho}_0)\mathbf{R}'\boldsymbol{\mu}^e(\boldsymbol{\rho}_0)
$$

$$\mathfrak{m}_{\omega}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{\square})\right] = \frac{\square_{xx}(\boldsymbol{\rho}) + \boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{\square})'\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{\square}) - 2\boldsymbol{\sigma}'_{zx}(\boldsymbol{\rho})\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{\square}) - \square^{e}(\boldsymbol{\rho}^{\square})}{2\left[\square^{e}(\boldsymbol{\rho}^{\square})\right]^{\square}}$$

Let us now define the $\square+1$ vector of functions $\boldsymbol{\gamma}(\boldsymbol{\rho}) = [\square_{\square}(\boldsymbol{\rho})\square\square_{\square}(\boldsymbol{\rho})\square\square\square\square\square\square_{k}(\boldsymbol{\rho})]'$, so that all the elements of $\square_{xx}(\boldsymbol{\rho})$, $\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})$ and $\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})$ can be trivially written as functions of $\boldsymbol{\gamma}(\boldsymbol{\rho})$. Then, tedious but otherwise straightforward algebra shows that both $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})\right]$ and $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{\square})\right]$ can be written as homeomorphic functions of $\boldsymbol{\gamma}(\boldsymbol{\rho})$. As a result, the estimators of $\boldsymbol{\rho}$ based on minimising the optimal norm of $\mathfrak{m}\left[\boldsymbol{\rho}\square\bar{\boldsymbol{\beta}}^{u}_{T}\right]$ or $\mathfrak{m}\left[\boldsymbol{\rho}\square\bar{\boldsymbol{\beta}}^{u}_{T}\right]$, will be asymptotically equivalent to the CMD estimators based on minimising the optimal norm of $\boldsymbol{\gamma}(\boldsymbol{\rho}) - \bar{\boldsymbol{\gamma}}_{T}$, where $\bar{\boldsymbol{\gamma}}_{T}$ contains the first $\square+1$ sample autocovariances of $\square_{t}$.

Proposition 10

The proof of these three cases, which correspond to an asymptotically *strictly unconstrained* auxiliary model, an asymptotically *strictly constrained* auxiliary model, and an asymptotically *correctly equality constrained* auxiliary model follows the lines of the proof of Proposition 8.

In the first case, we have that $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square}) = \boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})$, so that $\square_{t}\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right] = \square_{t}\left[\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})\right]$ $\forall\square$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right] = \mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})\right]$ for all $\boldsymbol{\rho}$ in a neighbourhood of $\boldsymbol{\rho}^{\square}$, which implies that $\square\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\rho} = \square\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\rho}$. In addition, $\sqrt{\square}\bar{\boldsymbol{\mu}}^{i}_{T} = \square_{p}(1)$ and $\sqrt{\square}(\bar{\boldsymbol{\theta}}^{i}_{T}-\bar{\boldsymbol{\theta}}^{u}_{T}) = \square_{p}(1)$ from Propositions 1 and 2 respectively, and $\sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\bar{\boldsymbol{\beta}}^{i}_{T}\right] - \sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\bar{\boldsymbol{\beta}}^{u}_{T}\right] = \square_{p}(1)$ in view of Proposition 3, so that $\mathcal{I}^{i}_{\square} = \mathcal{I}^{u}_{\square}$.

In the second case, $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square}) = \boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{\square})$, so that $\square_{t}\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right] = \square_{t}\left[\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{\square})\right]$ $\forall\square$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right] = \mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{\square})\right]$ for all $\boldsymbol{\rho}$ in a neighbourhood of $\boldsymbol{\rho}^{\square}$, which implies that $\square\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\rho} = \square\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{\square})\right]\square\square\boldsymbol{\rho}$. Similarly, Propositions 1 to 3 also imply that $\sqrt{\square}(\bar{\boldsymbol{\mu}}^{i}_{T}-\bar{\boldsymbol{\mu}}^{e}_{T}) = \square_{p}(1)$, $\sqrt{\square}(\bar{\boldsymbol{\theta}}^{i}_{T}-\bar{\boldsymbol{\theta}}^{e}_{T}) = \square_{p}(1)$, and $\sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\bar{\boldsymbol{\beta}}^{i}_{T}\right] - \sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^{\square}\square\bar{\boldsymbol{\beta}}^{e}_{T}\right] = \square_{p}(1)$, so that $\mathcal{I}^{i}_{\square} = \mathcal{I}^{e}_{\square}$.

In the last case, of course, $\boldsymbol{\beta}^i(\boldsymbol{\rho}^\square) = \boldsymbol{\beta}^u(\boldsymbol{\rho}^\square) = \boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)$, so that $\square_t\left[\boldsymbol{\beta}^i(\boldsymbol{\rho}^\square)\right] = \square_t\left[\boldsymbol{\beta}^i(\boldsymbol{\rho}^\square)\right] = \square_t\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right] \forall\square$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^i(\boldsymbol{\rho}^\square)\right] = \mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)\right] = \mathfrak{m}\left[\boldsymbol{\rho}\square\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right]$ for all $\boldsymbol{\rho}$ in a neighbourhood of $\boldsymbol{\rho}^\square$, which implies that $\square\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\boldsymbol{\beta}^i(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\rho} = \square\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\boldsymbol{\beta}^u(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\rho} = \square\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\boldsymbol{\beta}^e(\boldsymbol{\rho}^\square)\right]\square\square\boldsymbol{\rho}$. But in contrast, even if $\square$ is large, $\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\bar{\boldsymbol{\beta}}^i_T\right]$ will only coincide with $\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\bar{\boldsymbol{\beta}}^u_T\right]$ approximately half the time, while it will coincide with $\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\bar{\boldsymbol{\beta}}^e_T\right]$ the other half. Nevertheless, since in this case $\sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\bar{\boldsymbol{\beta}}^e_T\right] - \sqrt{\square}\mathfrak{m}\left[\boldsymbol{\rho}^\square\square\bar{\boldsymbol{\beta}}^u_T\right] = \square_p(1)$ from Proposition 8, all three estimators are asymptotically equivalent.

Proposition 11

The proof of this result follows directly from the proof of Theorem 6 in Chiang (1956), with the first $\square$ elements of $\mathcal{K}^r_\square\bar{\boldsymbol{\beta}}^r_T$ and $\mathcal{K}^r_\square\boldsymbol{\beta}^r(\boldsymbol{\rho})$ playing the roles of the consistent and asymptotically normal "sufficient statistics", and their plims respectively. In this respect, note that the only role of $\mathcal{K}^r_\square$ is to relegate the $\square$ singular combinations of Proposition 1 to the last positions.

The expected value of the score of an MA(1) model

In order to find

$$\begin{aligned}
\mathfrak{m}_\delta(\boldsymbol{\rho};\boldsymbol{\beta}) &= \square\left[\frac{1}{\square}\square_t(\square)\frac{\square\square_t(\square)}{\square\square} + \square_\square\middle|\boldsymbol{\rho}\right] \\
\mathfrak{m}_\psi(\boldsymbol{\rho};\boldsymbol{\beta}) &= \square\left[\frac{1}{\square}\left[\frac{\square^\square_t(\square)}{\square} - 1\right] + \square_\square\middle|\boldsymbol{\rho}\right]
\end{aligned}$$

it is convenient to write

$$\square_t(\square) = \sum_{j\square\square}^{\infty}\square^j\square_{t-j} = \frac{1}{1-\square\square}\square_t$$

and

$$\frac{\square\square_t(\square)}{\square\square} = -\sum_{j\square\square}^{\infty} \square\square^{j-\square}\square_{t-j} = \frac{-\square}{(1-\square\square)^{\square}}\square_t$$

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

$$\begin{aligned}
&\frac{1}{1-\square\square} \cdot \Gamma_{x_t}(\square) \cdot \frac{1}{1-\square\square^{-\square}} \\
= \; &\frac{1}{1-\square^{\square}} \left(\begin{array}{c} \square_{\square}(\boldsymbol{\rho}) \left[1 + \sum_{j\square\square}^{\infty} \square^j (\square^j + \square^{-j})\right] \\ + \sum_{l\square\square}^{\infty} \square_l(\boldsymbol{\rho})\square^l \left[1 + \sum_{j\square\square}^{\infty} \square^j (\square^j + \square^{-j})\right] \\ + \sum_{l\square\square}^{\infty} \square_l(\boldsymbol{\rho})\square^{-l} \left[1 + \sum_{j\square\square}^{\infty} \square^j (\square^j + \square^{-j})\right] \end{array} \right)
\end{aligned}$$

Hence,

$$\square \left[\square_t^{\square}(\square) \middle| \boldsymbol{\rho} \right] = \frac{\square_{\square}(\boldsymbol{\rho})}{1-\square^{\square}} \left(1 + 2\sum_{l\square\square}^{\infty} \square^l \frac{\square_l(\boldsymbol{\rho})}{\square_{\square}(\boldsymbol{\rho})} \right)$$

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

$$\square \left[\square_t^{\square}(\square) \middle| \boldsymbol{\rho} \right] = \frac{\square}{\left(1-\square^{\square}\right)\left(1-\square^{\square}\right)} \cdot \frac{1-\square\square}{1+\square\square}$$

In fact, given that we can write

$$\square_t(\square) = \frac{1}{1-\square\square}\square_t = \frac{1}{(1-\square\square)(1-\square\square)}\square_t\square$$

it is not surprising that $\square \left[\square_t^{\square}(\square) \middle| \boldsymbol{\rho} \right]$ coincides with the unconditional variance of an AR(2) process with autoregressive roots $\square$ and $\square$, and innovation variance $\square$.

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

$$\begin{aligned}
&\frac{\square}{(1-\square\square)^{\square}} \cdot \Gamma_{x_t}(\square) \cdot \frac{1}{1-\square\square^{-\square}} \\
= \; &\sum_{j\square\square}^{\infty} \square\square^{j-\square}\square^j \times \left(\square_{\square}(\boldsymbol{\rho}) + \sum_{l\square\square}^{\infty} \square_l(\boldsymbol{\rho})\square^l + \sum_{l\square\square}^{\infty} \square_l(\boldsymbol{\rho})\square^{-l} \right) \times \left(1 + \sum_{k\square\square}^{\infty} \square^k \square^{-k} \right)
\end{aligned}$$

$$= \square_{\square}(\boldsymbol{\rho}) \sum_{j\square\square}^{\infty} \square\square^{j-\square}\square^{j} + \sum_{j\square\square}^{\infty}\sum_{l\square\square}^{\infty} \square\square^{j-\square}\square_{l}(\boldsymbol{\rho})(\square^{l}+\square^{-l})\square^{j}$$

$$+\square_{\square}(\boldsymbol{\rho}) \sum_{j\square\square}^{\infty}\sum_{k\square\square}^{\infty} \square\square^{j-\square}\square^{k}\square^{j}\square^{-k} + \sum_{j\square\square}^{\infty}\sum_{k\square\square}^{\infty}\sum_{l\square\square}^{\infty} \square\square^{j-\square}\square^{k}\square_{l}(\boldsymbol{\rho})(\square^{l}+\square^{-l})\square^{-k}\square^{j}$$

Therefore, the coefficient associated with the constant term will be

$$\square_{\square}(\boldsymbol{\rho}) \sum_{j\square\square}^{\infty} \square\square^{\square j-\square} + \sum_{l\square\square}^{\infty} \square\square^{l-\square}\square_{l}(\boldsymbol{\rho}) + 2\sum_{l\square\square}^{\infty} \square^{l}\square_{l}(\boldsymbol{\rho}) \sum_{j\square\square}^{\infty} \square\square^{\square j-\square} + \sum_{l\square\square}^{\infty} \square\square^{l}\square_{l}(\boldsymbol{\rho}) \sum_{j\square\square}^{\infty} \square^{\square j-\square}$$

But since for $|\square| \square 1$

$$\sum_{j\square\square}^{\infty} \square^{\square j-\square} = \square \sum_{j\square\square}^{\infty} \square^{\square j} = \frac{\square}{1-\square^{\square}}$$

$$\sum_{j\square\square}^{\infty} \square\square^{\square j-\square} = \square \sum_{j\square\square}^{\infty} (\square+1)\square^{\square j} = \frac{\square}{\left(1-\square^{\square}\right)^{\square}}$$

we will have that

$$\square\left[\square_{t}(\square)\frac{\square\square_{t}(\square)}{\square\square}\bigg|\,\boldsymbol{\rho}\right] = -\frac{\square_{\square}(\boldsymbol{\rho})}{(1-\square^{\square})^{\square}}\left\{\square + \sum_{l\square\square}^{\infty}\left[2\square^{\square}+\left(1-\square^{\square}\right)\square\right]\square^{l-\square}\frac{\square_{l}(\boldsymbol{\rho})}{\square_{\square}(\boldsymbol{\rho})}\right\}$$

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

$$\square\left[\square_{t}(\square)\frac{\square\square_{t}(\square)}{\square\square}\bigg|\,\boldsymbol{\rho}\right] = \frac{\square}{(1-\square^{\square})\left(1-\square^{\square}\right)^{\square}(1-\square\square)^{\square}}\left(\square^{\square}\square^{\square}+\square^{\square}\square-\square-\square\right)$$

Simulation-based estimators

For the sake of clarity, we have assumed throughout that analytical expressions for (7) and (9) can be readily obtained, as in sections 3.1 and 3.2. 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 $\square$

realizations of size □ of the true process simulated with parameter values equal to $\boldsymbol{\rho}$. Specifically,

$$\begin{aligned}\mathcal{L}_T(\boldsymbol{\rho};\boldsymbol{\theta}) &\simeq \mathcal{L}_{HT}(\boldsymbol{\rho};\boldsymbol{\theta}) = \frac{1}{\square}\sum_h \frac{1}{\square}\sum_t \frac{\square_t(\boldsymbol{\theta})}{\square\boldsymbol{\theta}} \\ \mathfrak{m}_T(\boldsymbol{\rho};\boldsymbol{\beta}) &\simeq \square_{HT}(\boldsymbol{\rho};\boldsymbol{\beta}) = \frac{1}{\square}\sum_h \frac{1}{\square}\sum_t \frac{\square\square_t(\boldsymbol{\theta})}{\square\boldsymbol{\theta}} + \frac{\square\square'(\boldsymbol{\theta})}{\square\boldsymbol{\theta}}\boldsymbol{\mu}\end{aligned}$$

where we can make the last terms arbitrarily close in a numerical sense to the first ones as □ $\to \infty$. Nevertheless, it is important to bear in mind that these simulated functions will seldom be differentiable with respect to $\boldsymbol{\rho}$ unless the underlying uniform variates are kept fixed across simulations, there are no discrete variables in $\square_t$, and smooth transformations of the underlying uniforms are used to obtain the desired distributions. In this respect, we would like to stress that in the only example in which we relied on simulations to compute the required moments (see section 3.3), all three conditions were fulfilled.

Alternatively, since we are assuming that $\square_t$ is strictly stationary and ergodic, there is, mact,45URmC-4AalAIA;rternatPAIA;rve sPAIA;rmulAIA;ratPAIA;ro45URmC-4
expectats by their sample □ $\to\infty$ a45P5-RaCTRloguCP5-ReCD54RsC36I65DRi45UROT5ORaC39O4Rs

On ths basis, we canaproximate the different bnding functions $\boldsymbol{\beta}$

($\boldsymbol{\rho}$ by

mea45U5DRsC3TOD5-RoC-56R $\boldsymbol{\rho}$ C39I656RaCagoac5-56Rpps5U5DRsimuli-ons5RpfsizC--56RtelyC3TOI54AcC

□ of estimators obtained from each simulated sample. The main attraction of the second procedure is that it may often improve the small sample properties of the estimators of $\boldsymbol{\rho}$ (see e.g. Gourieroux, Renault and Touzi, 2000).

From a computational point of view, though, the crucial advantage of GMM-based estimators over CMD-ones is that they avoid the calculation of the possibly constrained estimators for each simulation of the process. However, given that $\bar{\boldsymbol{\mu}}_T^u = 0$, we can always regard the GMM-based II procedure as a CMD procedure that matches the value in the observed sample of a vector that contains one multiplier per auxiliary parameter with the (average) value of the same vector in the simulated sample(s). At the same time, since the term $\left[\square\square'(\bar{\boldsymbol{\theta}}_T^r)\square\square\boldsymbol{\theta}\right] \cdot \bar{\boldsymbol{\mu}}_T^r$ is fixed across simulations, what we effectively do in practice is to minimise the distance between the score in the actual sample and the (average) score in the simulated samples.

Finally, note that the autocovariance matrices $\square_\tau(\boldsymbol{\rho}; \boldsymbol{\beta}_T)$ used in the computation of the optimal weighting matrix for the continuously updated GMM-based II estimators can also be arbitrarily approximated by replacing the required expected values by their sample counterparts in a long simulation of length $\square \cdot \square$. Nevertheless, it is important to bear in mind that since $\square$ is finite in practice, the asymptotic covariance matrix of the GMM and CMD II estimators in Proposition 4 must be multiplied by the scalar quantity $(1 + \square^{-\square})$ (see GMR).

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.

Arellano, M. (1989): "On the efficient estimation of simultaneous equations with covariance restrictions", *Journal of Econometrics* 42, 247-265.

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

Calzolari, G., Fiorentini, G. and Sentana, E. (2001): "Indirect inference estimation of conditionally heteroskedastic factor models", mimeo, CEMFI.

Chiang, C.L. (1956): "On regular best asymptotically normal estimates", *Annals of Mathematical Statistics*, 27, 336-351.

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

Ferguson, T.S. (1958): "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities", *Annals of Mathematical Statistics* 29, 1046-1062.

Fiorentini, G., Sentana, E. and Calzolari, G. (2000): "The score of conditionally heteroskedastic dynamic regression models with student t innovations, and an LM test for multivariate normality", CEMFI Working Paper 0007.

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

Godfrey, L.G. (1988): *Misspecification 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.

Gourieroux, C., Renault, E. and Touzi, N. (2000): "Calibration by simulation for small sample bias correction", in R. Mariano, T. Shuermann and M. Weeks, ed. *Simulation-based inference in econometrics. Methods and applications*, Cambridge University Press.

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.

Newey, W.K. and McFadden, D. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, pp. 2111-2245.

Rothemberg, T.J. (1973): *Efficient estimation with a priori information*, Cowles Foundation Monograph 23, Yale University Press.

Sargan, J.D. and Bhargava, A. (1983): "Maximum likelihood estimation of regression models with first 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. Barndorff-Nielsen (eds.), *Time series models with econometric, finance 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 misspecified models", *Econometrica* 50, 1-28.

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
$\eta = 0$	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
$\eta = 0$	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
$\eta = 0$	.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
$\eta = 0$	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
$\eta = 0$	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
$\eta = 0$	.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	
$\eta = 0$		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	
$\eta = 0$		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	
$\eta = 0$		.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 unrestricted II 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	-1.0831	0.8723	.7395	-.5901	.9299	.5302	-.3329	.9600	.3498
rmse	.7874	.0794	.2711	.4514	.0499	.1666	.3122	.0363	.1074
mean bias	-.2621	-.0277	.0645	-.1795	-.0201	.0467	-.1687	-.0200	.0418
std. dev.	.7425	.0744	.2633	.4142	.0456	.1600	.2627	.0303	.0990
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-1.0574	.8596	.4315	-.6082	.9183	.3130	-.4144	.9445	.2269
rmse	.8319	.1025	.2212	.5205	.0677	.1422	.5799	.0722	.1354
mean bias	-.3214	-.0404	.0685	-.2402	-.0317	.0530	-.2672	-.0355	.0609
std. dev.	.7673	.0942	.2104	.4617	.0598	.1320	.5146	.0629	.1210
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.7255	.7648	.2757	-1.1924	.8360	.1999	-.7804	.8916	.1376
rmse	2.2915	.3035	.3788	1.8727	.2499	.2793	1.5052	.2039	.1949
mean bias	-1.0195	-.1352	.1407	-.8394	-.1140	.1035	-.6394	-.0884	.0762
std. dev.	2.0522	.2717	.3517	1.6741	.2223	.2594	1.3626	.1838	.1794

Table 5

Mean, root mean square error, mean bias and standard deviation of the equality restricted II 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 II 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 7

Mean, root mean square error, mean bias and standard deviation of the unrestricted II 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	-.9546	.8861	.7069	-.5013	.9398	.5066	-.2486	.9702	.3310
rmse	.4518	.0493	.1660	.2655	.0302	.1034	.1805	.0209	.0700
mean bias	-.1336	-.0139	.0319	-.0907	-.0102	.0231	-.0844	-.0098	.0230
std. dev.	.4315	.0473	.1629	.2496	.0284	.1008	.1596	.0184	.0661
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.8919	.8800	.3958	-.4906	.9338	.2894	-.2529	.9658	.1943
rmse	.4694	.0607	.1241	.3164	.0413	.0868	.2208	.0294	.0662
mean bias	-.1559	-.0200	.0328	-.1226	-.0162	.0294	-.1057	-.0142	.0283
std. dev.	.4427	.0573	.1197	.2917	.0381	.0817	.1938	.0257	.0598
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.6044	.7787	.2506	-1.0513	.8543	.1907	-.6128	.9141	.1248
rmse	1.9730	.2643	.2946	1.4970	.2012	.2350	1.0655	.1470	.1517
mean bias	-.8984	-.1213	.1156	-.6983	-.0957	.0943	-.4718	-.0659	.0634
std. dev.	1.7565	.2349	.2709	1.3241	.1770	.2152	.9553	.1315	.1378

Table 8

Mean, root mean square error, mean bias and standard deviation of the equality restricted II 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 9

Mean, root mean square error, mean bias and standard deviation of the inequality restricted II 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 10

Mean, root mean square error, mean bias and standard deviation of the urestricted II 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	-.9079	0.8908	.7013	-.4660	.9438	.5002	-.2099	.9747	.3213
rmse	.2977	.0328	.1180	.1693	.0192	.0726	.1063	.0123	.0472
mean bias	-.0869	-.0092	.0263	-.0554	-.0062	.0167	-.0457	-.0053	.0133
std. dev.	.2847	.0315	.1150	.1600	.0182	.0706	.0959	.0111	.0453
1	*-.736*	*.9*	*.363*	*-.368*	*.95*	*.26*	*-.1472*	*.98*	*.166*
mean	-.8231	.8887	.3840	-.4336	.9413	.2774	-.2027	.9725	.1821
rmse	.2690	.0353	.0806	.1620	.0215	.0535	.1122	.0150	.0400
mean bias	-.0871	-.0113	.0210	-.0656	-.0087	.0174	-.0555	-.0075	.0161
std. dev.	.2545	.0334	.0778	.1481	.0197	.0506	.0975	.0130	.0366
.1	*-.706*	*.9*	*.135*	*-.353*	*.95*	*.0964*	*-.141*	*.98*	*.0614*
mean	-1.3798	.8078	.2229	-.8413	.8822	.1619	-.4565	.9358	.1094
rmse	1.4439	.1946	.2187	1.0238	.1397	.1477	.6963	.0956	.1080
mean bias	-.6738	-.0922	.0879	-.4883	-.0678	.0655	-.3155	-.0442	.0480
std. dev.	1.2770	.1714	.2003	.8999	.1221	.1324	.6208	.0848	.0967

Table 11

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

Table 12

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

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

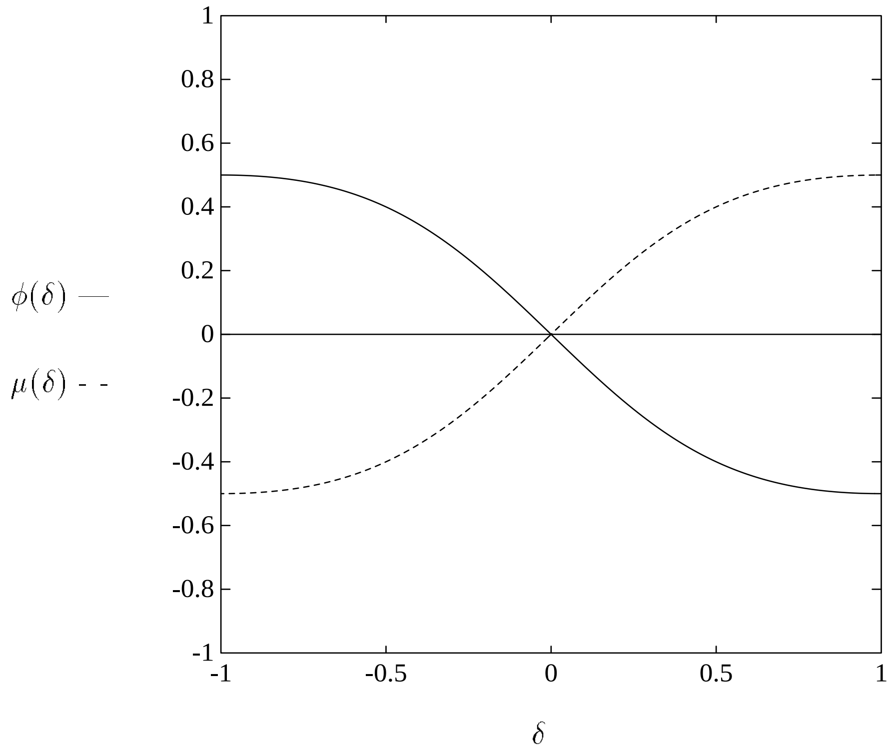

Figure 2: Binding Functions for AR(1) estimated as MA(1)