Constrained Indirect Estimation*

Giorgio Calzolari
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<calzolar@ds.unifi.it>

Gabriele Fiorentini
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<fiorentini@ds.unifi.it>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

March 1999
Revised: January 2004

Abstract

We develop generalised indirect estimation procedures that handle equality and inequality constraints on the auxiliary model parameters by extracting information from the relevant multipliers, and compare their asymptotic efficiency to maximum likelihood. We also show that regardless of the validity of the restrictions, the asymptotic efficiency of such estimators can never decrease by explicitly considering the multipliers associated with additional equality constraints. Furthermore, we discuss the variety of effects on efficiency that can result from imposing constraints on a previously unrestricted model. As an example, we consider stochastic volatility process estimated through a GARCH model with Gaussian or t distributed errors.

Keywords: EMM, Indirect Inference, GMM, Minimum Distance, Stochastic Volatility

JEL: C13, C15

*This is a thoroughly revised version of Calzolari, Fiorentini and Sentana (2000). We are grateful to Filippo Altissimo, Torben Andersen, Manuel Arellano, Jon Danielsson, Antonis Demos, Ramdan Dridi, Russell Davidson, Frank Diebold, Jean-Marie Dufour, Ron Gallant, Christian Gouriéroux, Lars Hansen, Jim Heckman, Joel Horowitz, Hide 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 (Rome), Federico II (Naples), LSE, Montreal, North Carolina Research Triangle, 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. The suggestions of three anonymous referees have also substantially improved the paper. Of course, the usual caveat applies. Financial support from CICYT, IVIE and MIUR through the project “Specification, estimation and testing of latent variable models. Applications to the analysis and forecasting of economic and financial time series” is gratefully acknowledged.

1 Introduction

Consider a stochastic process, x_t, characterised by the sequence of parametric conditional densities $p(x_t|\mathbf{X}_{t-1};\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ denotes the d parameters of interest, and $\mathbf{X}_{t-1} = \{x_{t-1}, x_{t-2}, \ldots\}$. Consider also a possibly misspecified auxiliary model, described by the sequence of conditional densities $f(x_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a c dimensional vector of parameters, with $d \leq c$. In those situations in which no closed-form expression for $p(x_t|\mathbf{X}_{t-1};\boldsymbol{\rho})$ exists, but at the same time it is easy to estimate $\boldsymbol{\theta}$, or to compute expectations of functions of x_t, either analytically, or by simulation or quadrature, the indirect estimation procedures of Gallant and Tauchen (1996) (GT96), Gouriéroux, 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 indirect estimation procedure of GMR uses the pseudo-maximum likelihood (PML) 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 GT96 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 PML 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 Gouriéroux and Monfort, 1996) (GM96).

One of those conditions, though, is that the parameters of the auxiliary model are unrestricted, and consequently, that their PML estimators have an asymptotically normal distribution with a full rank covariance matrix under standard regularity conditions (see e.g. Gouriéroux, Monfort and Trognon (1984) or White (1982) for a discussion of unconstrained PML estimation, and its relationship to the Kullback discrepancy between $f(x_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ and $p(x_t|\mathbf{X}_{t-1};\boldsymbol{\rho})$). The first contribution of our paper is to show how indirect estimation 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 GT96 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 indirect estimators of GMR to the possibly restricted case. We also prove that we can find "restricted" CMD indirect estimators that are asymptotically equivalent to the GMM estimators by an appropriate choice of weighting matrix. And although we concentrate for expositional purposes on PML estimation of the auxiliary model under the assumption that the form of the density function is time-invariant, and x_t strictly stationary and ergodic, our procedures can be extended to cover other extremum estimators of just identified auxiliary models with strictly exogenous regressors in more general contexts (see section 4.1.3 of GM96). For analogous reasons, we deliberately separate the results directly related to our proposed modification of the existing indirect estimation procedures from the way one would conduct numerical simulation in practice. However, since very often one has to resort to simulation to implement indirect estimation procedures, we include an appendix in which some relevant issues are discussed.

There are at least three 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 GT96). In other cases, though, the log-likelihood function can always be computed, but some of the auxiliary parameters may be poorly identified, if at all, in certain regions of the auxiliary parameter space, so that we may decide to restrict it to avoid such discontinuities (see section 3 below, or Calzolari, Fiorentini and Sentana (2004) for examples in which both situations concur). Finally, there may be also non-statistical reasons for imposing inequality constraints; for instance, to guarantee that an auxiliary model always generates a positive nominal short interest rate. In all cases, 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 shown by the extensive literature on Lagrange multiplier specification tests. Furthermore, equality restricted procedures may be particularly useful from a computational point of view, because in many situations of 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 resulting indirect estimators. To do so, we first explicitly relate the asymptotic efficiency of our indirect estimators to the usually infeasible maximum likelihood (ML) estimator. Then, we show that the asymptotic efficiency of indirect estimators can never decrease by considering the Lagrange multipliers associated with the implicit zero constraints mentioned above. 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 indirect estimators, which can be particularly useful when the informational content of the original auxiliary parameters about the structural parameters appears to be poor. Finally, 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 indirect estimators, and others in which false constraints enable the restricted indirect estimators to achieve full efficiency.

For illustrative purposes, we apply our modified procedures to 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 organised as follows. In section 2, we include a thorough discussion of "restricted" indirect estimation procedures, and of the efficiency consequences of the constraints. Detailed applications of such procedures to the aforementioned example can be found in section 3. Finally, our conclusions are presented in section 4. Proofs and auxiliary results are gathered in the appendix.

2 Theoretical set up

2.1 "Restricted" indirect estimators

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

$$Q_T(\boldsymbol{\beta}) = \bar{l}_T(\boldsymbol{\theta}) + h'(\boldsymbol{\theta})\boldsymbol{\mu} \tag{1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\theta}', \boldsymbol{\mu}')'$, and $\boldsymbol{\mu}$ are the s "multipliers" associated with the s constraints implicitly characterised by the vector of functions $h(\boldsymbol{\theta})$, which effectively force $\boldsymbol{\theta}$ to lie in a compact and 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:

$$\begin{array}{llll} \text{(a)} & h(\boldsymbol{\theta}) \text{ unrestricted} & \boldsymbol{\mu} = \mathbf{0} & \boldsymbol{\Theta}^r \equiv \boldsymbol{\Theta} \\ \text{(b)} & h(\boldsymbol{\theta}) = \mathbf{0} & \boldsymbol{\mu} \text{ unrestricted} & \boldsymbol{\Theta}^r \equiv \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : h(\boldsymbol{\theta}) = \mathbf{0}\} \\ \text{(c)} & h(\boldsymbol{\theta}) \geq \mathbf{0} & \boldsymbol{\mu} \geq \mathbf{0} & \boldsymbol{\Theta}^r \equiv \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : h(\boldsymbol{\theta}) \geq \mathbf{0}\} \end{array} \tag{2}$$

Assuming that both the average pseudo-log likelihood function $\bar{l}_T(\boldsymbol{\theta})$, and the vector of functions $h(\boldsymbol{\theta})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$, the latter with a Jacobian matrix $\partial h'(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$

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:

$$\frac{\partial Q_T(\hat{\boldsymbol{\beta}}_T^r)}{\partial \boldsymbol{\theta}} = \bar{m}_T(\hat{\boldsymbol{\beta}}_T^r) = \mathbf{0}, \tag{3}$$

where $\bar{m}_T(\boldsymbol{\beta})$ is the sample mean of

$$m_t(\boldsymbol{\beta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\mu},$$

which is the contribution of the t^{th} observation to the modified score of the auxiliary model, $\widehat{\ }$ indicates (P)ML estimators, and the superscript $r = (u, e, i)$ stands for unrestricted, equality restricted and inequality restricted respectively. In addition, $\hat{\boldsymbol{\beta}}_T^r$ must satisfy the complementary slackness restrictions

$$h(\hat{\boldsymbol{\theta}}_T^r) \odot \hat{\boldsymbol{\mu}}_T^r = \mathbf{0}, \tag{4}$$

plus the appropriate (in)equality restrictions on $h(\hat{\boldsymbol{\theta}}_T^r)$ and/or $\hat{\boldsymbol{\mu}}_T^r$ in (2), where 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 $m_t(\boldsymbol{\beta})$ not only depends on the c auxiliary model parameters $\boldsymbol{\theta}$, but also on the s multipliers $\boldsymbol{\mu}$ associated with the restrictions.

Let us now define

$$\mathcal{L}(\boldsymbol{\rho}; \boldsymbol{\theta}) = E\left[\bar{l}_T(\boldsymbol{\theta}) \middle| \boldsymbol{\rho}\right], \tag{5}$$

where $E(.|\boldsymbol{\rho})$ refers to an expected value computed with respect to the distribution of the model of interest evaluated at $\boldsymbol{\rho}$. In what follows, we assume that

Assumption 1 $\bar{l}_T(\boldsymbol{\theta})$ *converges almost surely to* $\mathcal{L}(\boldsymbol{\rho}; \boldsymbol{\theta})$ *uniformly in* $(\boldsymbol{\theta}, \boldsymbol{\rho})$ *as* T *goes to infinity, where* $\mathcal{L}(\boldsymbol{\rho}; \boldsymbol{\theta})$ *is twice continuously differentiable with respect to both its arguments.*

For each value of $\boldsymbol{\rho}$, we can define the binding functions for the "constrained" auxiliary parameters $\boldsymbol{\theta}$ and the associated "multipliers" $\boldsymbol{\mu}$, $\boldsymbol{\beta}^r(\boldsymbol{\rho}) = \left[\boldsymbol{\theta}^{r\prime}(\boldsymbol{\rho}), \boldsymbol{\mu}^{r\prime}(\boldsymbol{\rho})\right]'$ say, as the values of $\boldsymbol{\beta}$ associated with the maximum over the restricted parameter space $\boldsymbol{\Theta}^r$ of the (population) Lagrangian function

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

As a result, if we denote by

$$\mathfrak{m}\left(\boldsymbol{\rho}; \boldsymbol{\beta}\right) = E\left[\bar{m}_T(\boldsymbol{\beta}) | \boldsymbol{\rho}\right], \tag{6}$$

the binding functions must satisfy the first-order conditions:

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

the exclusion restrictions

$$h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho})\right] \odot \boldsymbol{\mu}^r(\boldsymbol{\rho}) = \mathbf{0}, \tag{7}$$

plus the required (in)equality restrictions on $h[\boldsymbol{\theta}^r(\boldsymbol{\rho})]$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho})$ in (2), as long as the differentiation and expectation operators can be interchanged, which we assume henceforth. In addition, we assume that $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ is unique, in the sense that $\mathcal{L}[\boldsymbol{\rho};\boldsymbol{\theta}^r(\boldsymbol{\rho})] > \mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}^r$ in an open neighbourhood of $\boldsymbol{\theta}^r(\boldsymbol{\rho})$. As a consequence, we can use standard PML results to prove the strong consistency of $\hat{\boldsymbol{\beta}}_T^r$ for $\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)$, where $\boldsymbol{\rho}^0$ denotes the true value of the parameters of interest, and $\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)$ the "constrained" pseudo-true values of $\boldsymbol{\beta}$.

To ensure the local identification of $\boldsymbol{\rho}^0$, we 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 $c \geq d$ (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 2

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

and

$$rank\left\{\frac{\partial\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\rho}'}\right\} = d$$

for any $\boldsymbol{\rho}$ in an open neighbourhood of $\boldsymbol{\rho}^0$.

As usual, such assumptions are rather difficult to check in non-linear models, but they are crucial for the consistency of the indirect estimators that we discuss. Intuitively, the reason is that when Assumption 2 holds, if we knew $\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)$, we could recover $\boldsymbol{\rho}^0$ by either inverting the binding functions, or solving the possibly non-linear system of equations $\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] = \mathbf{0}$ with respect to its first argument holding the second argument fixed. In practice, though, we do not know the pseudo true values, but since they are consistently estimated by the auxiliary model, we can obtain consistent estimators of $\boldsymbol{\rho}^0$ by choosing the parameter values that minimise either some appropriately defined distance between $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ and $\hat{\boldsymbol{\beta}}_T^r$, or a given norm of the sample moments $\mathfrak{m}(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}_T^r)$. In particular, we can minimise with respect to $\boldsymbol{\rho}$ the following quadratic forms:

$$D^r(\boldsymbol{\rho};\boldsymbol{\Omega},\hat{\boldsymbol{\beta}}_T^r) = \left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \hat{\boldsymbol{\beta}}_T^r\right]' \cdot \boldsymbol{\Omega} \cdot \left[\boldsymbol{\beta}^r(\boldsymbol{\rho}) - \hat{\boldsymbol{\beta}}_T^r\right]$$

or

$$G(\boldsymbol{\rho};\boldsymbol{\Psi},\hat{\boldsymbol{\beta}}_T^r) = \mathfrak{m}'(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}_T^r) \cdot \boldsymbol{\Psi} \cdot \mathfrak{m}(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}_T^r)$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{\Psi}$ are positive semidefinite weighting matrices of orders $c + s$ and c respectively, and the letters D and G 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

$$\begin{aligned}\tilde{\boldsymbol{\rho}}^r_{DT}(\boldsymbol{\Omega}) &= \arg\min_{\boldsymbol{\rho}} D^r(\boldsymbol{\rho};\boldsymbol{\Omega},\hat{\boldsymbol{\beta}}^r_T)\\ \tilde{\boldsymbol{\rho}}^r_{GT}(\boldsymbol{\Psi}) &= \arg\min_{\boldsymbol{\rho}} G(\boldsymbol{\rho};\boldsymbol{\Psi},\hat{\boldsymbol{\beta}}^r_T)\end{aligned}$$

as the "restricted" CMD and GMM indirect 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 $\hat{\boldsymbol{\beta}}^r_T$, the asymptotic covariance matrix of $\tilde{\boldsymbol{\rho}}^r_{DT}(\boldsymbol{\Omega})$ and $\tilde{\boldsymbol{\rho}}^r_{GT}(\boldsymbol{\Psi})$ is likely to be unnecessarily large in those overidentified situations in which $c > d$.

Let us start by analysing the second criterion function. It is well known that if the sample moments $\mathfrak{m}(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}^r_T)$ 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 positive semidefinite) is given by the inverse of the asymptotic variance of $\sqrt{T}\mathfrak{m}(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}^r_T)$ (see e.g. Hansen, 1982). In order to derive the required asymptotic distribution, we assume the necessary conditions for a law of large numbers and a central limit theorem to apply to the average Hessian and modified score of the log-likelihood of the auxiliary model respectively. More formally,

Assumption 3

$$\frac{\partial \bar{l}^2_T(\boldsymbol{\theta}^*_T)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \mathcal{J}^r_0 = o_p(1),$$

and

$$\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] \to N(\mathbf{0},\mathcal{I}^r_0),$$

where $\mathcal{J}^r_0$ and $\mathcal{I}^r_0$ are non-stochastic $c\times c$ matrices, with $\mathcal{I}^r_0$ positive definite, and $\boldsymbol{\theta}^*_T$ is any sequence such that $\boldsymbol{\theta}^*_T - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0) = o_p(1)$.

In this respect, it is important to note that relative to the standard unconstrained case, the main effect of adding the constant term $\left\{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\right\}\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)$ to the original score $\partial l_t\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}$ is to centre around zero the asymptotic distribution of $m_t\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$. Therefore, if $\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)$ is in the interior of the admissible auxiliary parameter space $\boldsymbol{\Theta}^r$, Assumption 3 is equivalent to the high level assumptions made by GMR and GT96. 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 under conditional Gaussianity, as discussed in Fiorentini, Sentana and Calzolari (2003)).

Unfortunately, we cannot directly rely on the results in GT96 to derive the asymptotic distribution of the sample moments $\mathfrak{m}(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}^r_T)$, since the "restricted" estimator $\hat{\boldsymbol{\theta}}^r_T$ may not be asymptotically normal in large samples in the presence of inequality constraints (see Andrews (1999) and the references therein).[1] In addition, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_T^r$ is singular for $r = (u, e, i)$. More specifically:

Proposition 1 *Under Assumptions 1,2, and 3*

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

Such a singularity is a direct consequence of the fact that the complementary slackness conditions (4) must always be satisfied by $\hat{\boldsymbol{\beta}}_T^r$. Nevertheless, it is important to mention that since their population counterparts (7) 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 c linear combinations that are asymptotically well behaved:

Proposition 2 *Under Assumptions 1,2, and 3*

$$\left[\mathcal{J}_0^r + \left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^0) \otimes I_c\right] \frac{\partial vec\left\{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial \boldsymbol{\theta}\right\}}{\partial \boldsymbol{\theta}'}\right] \sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]$$

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

Hence, even though $\hat{\boldsymbol{\theta}}_T^r$ and $\hat{\boldsymbol{\mu}}_T^r$ have a singular and possibly non-Gaussian asymptotic distribution, Proposition 2 shows that under our regularity conditions, there are always c linear combinations that are asymptotically normally distributed, irrespective of the exact nature of the restrictions, and irrespective of whether the restrictions on $h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]$ and $\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)$ are satisfied with equality, or strict inequality. It turns out that those c linear combinations are implicitly contained in the expected value of the modified score:

Proposition 3 *Under Assumptions 1, 2, and 3*

$$\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0; \hat{\boldsymbol{\beta}}_T^r) + \sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] = o_p(1).$$

Therefore, $\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0; \hat{\boldsymbol{\beta}}_T^r)$ 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, 2 and 3*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_{GT}^r\left[(\mathcal{I}_0^r)^{-1}\right] - \boldsymbol{\rho}^0\right\} \to N\left[0, (\mathcal{C}_0^r)^{-1}\right],$$

where

$$\mathcal{C}_0^r = \frac{\partial \mathfrak{m}'\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial \boldsymbol{\rho}} \cdot (\mathcal{I}_0^r)^{-1} \cdot \frac{\partial \mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial \boldsymbol{\rho}'}. \tag{8}$$

[1] It may seem at first sight that we could handle inequality restrictions on the parameters of the auxiliary model with the existing *unconstrained* indirect estimation 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 $(\theta_j^*)^2$, where $-\infty < \theta_j^* < \infty$. Unfortunately, the regularity conditions in Assumptions 2 and 3 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 $\theta_j^i(\boldsymbol{\rho}^0) = 0$.

Given that this expression is completely analogous to the one derived by GT96 for their GMM version of the indirect estimator in the absence of constraints, the required matrices can also be consistently estimated using their suggested procedures. In particular, since under our assumptions

$$E\left\{\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho})\right]\middle|\,\boldsymbol{\rho}\right\}=\mathbf{0}\quad\forall T,$$

the time-invariant functional form of $m_t(\boldsymbol{\beta})$, and the strict stationarity and ergodicity of x_t imply that

$$\mathcal{I}_0^r=\lim_{T\to\infty}V\left\{\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]\middle|\,\boldsymbol{\rho}^0\right\}=\sum_{\tau=-\infty}^{\infty}S_\tau\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right],\tag{9}$$

where

$$S_\tau(\boldsymbol{\rho};\boldsymbol{\beta})=E\left\{m_t(\boldsymbol{\beta})m'_{t-\tau}(\boldsymbol{\beta})\middle|\,\boldsymbol{\rho}\right\}$$

for $\tau\geq 0$, and $S_\tau(\boldsymbol{\rho};\boldsymbol{\beta})=S'_{-\tau}(\boldsymbol{\rho};\boldsymbol{\beta})$ 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

$$\overline{\mathcal{I}}_T^r=\sum_{\tau=-T^\iota}^{T^\iota}w(\tau)\bar{S}_{\tau T}^r\tag{10}$$

with

$$\bar{S}_{\tau T}^r=\frac{1}{T}\sum_{t=\tau+1}^{T}m_t(\hat{\boldsymbol{\beta}}_T^r)m'_{t-\tau}(\hat{\boldsymbol{\beta}}_T^r)$$

where $w(\tau)$ are weights suggested by a standard heteroskedasticity and autocorrelation consistent (HAC) covariance estimation procedure, and ι the corresponding rate (see e.g. de Jong and Davidson (2000) and the references therein). Then, a feasible two-step optimal GMM estimator will be given by $\tilde{\boldsymbol{\rho}}_{GT}^r\left[(\overline{\mathcal{I}}_T^r)^{-1}\right]$. Alternatively, we could consider continuously updated GMM estimators à la Hansen, Heaton and Yaron (1996), by replacing $\bar{S}_{\tau T}^r$ in the above expressions with $S_\tau(\boldsymbol{\rho};\hat{\boldsymbol{\beta}}_T^r)$.

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

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

converges to a χ^2 distribution with $c-d$ degrees of freedom as $T\to\infty$, 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 indirect estimators of $\boldsymbol{\rho}$ based on the CMD function $D^r(\boldsymbol{\rho};\boldsymbol{\Omega};\hat{\boldsymbol{\beta}}_T^r)$. Unfortunately, we cannot directly rely on standard CMD theory, because as we saw before, the limiting distribution of $\sqrt{T}\left[\hat{\boldsymbol{\beta}}_T^r-\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ is singular and possibly non-normal. To overcome this difficulty, it is convenient to write down the linear transformations in Propositions 1 and 2 together in terms of the following square matrix of order $c+s$:

$$\begin{aligned}\mathcal{K}_0^r &= \begin{bmatrix}\mathcal{J}_0^r+\left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\otimes I_c\right]\partial vec\left\{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}' & \partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\\ diag\left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]\partial h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}' & diag\left\{h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]\right\}\end{bmatrix}\\ &= \begin{bmatrix}\mathcal{K}_{11,0}^r & \mathcal{K}_{12,0}^r\\ \mathcal{K}_{21,0}^r & \mathcal{K}_{22,0}^r\end{bmatrix},\end{aligned}$$

where $diag\,(.)$ is the operator that transforms a vector into a diagonal matrix of the same order by placing its elements along the main diagonal. Then, if we transform the CMD conditions by premultiplying them by $\mathcal{K}_0^r$, we will have that the asymptotic distribution of $\sqrt{T}\mathcal{K}_0^r\left[\hat{\boldsymbol{\beta}}_T^r-\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ will be normal, with the singularity confined to the last s elements. In this framework, we can prove the following generalisation of Proposition 4.3 in GM96, which in turn formalises earlier results in GMR:

Proposition 5 *Under Assumptions 1, 2 and 3*

$$\sqrt{T}\left[\tilde{\boldsymbol{\rho}}_{GT}^r(\boldsymbol{\Psi})-\tilde{\boldsymbol{\rho}}_{DT}^r(\mathcal{K}_0^{r\prime}\boldsymbol{\Psi}^{\boxplus}\mathcal{K}_0^r)\right]=o_p(1),$$

where

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

Given the equivalence between both estimators, in what follows we shall drop the D and G subscripts when no confusion arises. Apart from the computational advantages highlighted by GT96, 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 (6), irrespective of the exact nature of the restrictions, and irrespective of whether the restrictions on $h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]$ and/or $\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)$ are satisfied as equalities, or strict inequalities. However, 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 $c=d$, so that the auxiliary model exactly identifies the parameters of interest, then $\tilde{\boldsymbol{\rho}}_{DT}^r(\boldsymbol{\Omega})=\tilde{\boldsymbol{\rho}}_{GT}^r(\boldsymbol{\Psi})$ for large enough T irrespective of $\boldsymbol{\Omega}$ and $\boldsymbol{\Psi}$.*

2.2 Efficiency considerations

Given that both GMM and CMD can be regarded as particular cases of minimum chi-square methods (see e.g. Newey and McFadden (1994) (NM)), an attractive way of interpreting our previous results is to think of the population moments $\mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ as a set of c 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, $(\mathcal{I}_0^r)^{-1}$, and (ii) the identification content of the same parameters, as measured by the Jacobian of the population moments with respect to its first argument, $\partial\mathfrak{m}\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\rho}'$. This Jacobian matrix can be given a rather intuitive interpretation. Let $\bar{q}_T(\boldsymbol{\rho})$ denote the sample average of the log-likelihood score of the structural model, so that

$$q_t(\boldsymbol{\rho})=\frac{\partial\ln p(x_t|\mathbf{X}_{t-1};\boldsymbol{\rho})}{\partial\boldsymbol{\rho}}.$$

Then, a variation of the generalised information matrix equality implies that

$$\begin{aligned}\frac{\partial \mathfrak{m}\left[\boldsymbol{\rho};\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial \boldsymbol{\rho}'} &= \lim_{T\to\infty}\frac{\partial}{\partial \boldsymbol{\rho}'}E\left\{\left(\frac{1}{T}\sum_{t=1}^{T} m_t\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]\right)\middle|\,\boldsymbol{\rho}\right\}\\ &= \lim_{T\to\infty} cov\left\{\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right],\sqrt{T}\bar{q}_T(\boldsymbol{\rho})\middle|\,\boldsymbol{\rho}\right\} \qquad (11)\end{aligned}$$

(see GM96, Tauchen (1996), and NM for precise regularity conditions). Therefore, the second part of Assumption 2 guarantees that this covariance matrix has full column rank in an open neighbourhood of $\boldsymbol{\rho}^0$.

Expression (11) also allows us to formally characterise the asymptotic efficiency of our proposed estimator $\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]$ relative to the possibly infeasible ML estimator of $\boldsymbol{\rho}$, $\hat{\boldsymbol{\rho}}_T$.

Proposition 7 *Define*

$$\mathcal{D}_0^r = \lim_{T\to\infty} V\left\{\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0) - \frac{\partial \mathfrak{m}'\left[\boldsymbol{\rho}^0;\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial \boldsymbol{\rho}}\left[\mathcal{I}_0^r\right]^{-1}\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]\middle|\,\boldsymbol{\rho}^0\right\}$$

as the asymptotic residual variance of the limiting least squares projection of $\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)$ on $\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$. Then, under Assumptions 1, 2 and 3, the inverse of the asymptotic covariance matrix of $\tilde{\boldsymbol{\rho}}_T^r\left[(\mathcal{I}_0^r)^{-1}\right]$ given in (8), which can be interpreted in this context as the indirect asymptotic information matrix, will be given by

$$\mathcal{C}_0^r = \mathcal{B}_0 - \mathcal{D}_0^r,$$

where

$$\mathcal{B}_0 = \lim_{T\to\infty} V\left\{\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)\middle|\,\boldsymbol{\rho}^0\right\}$$

is the usual asymptotic information matrix.

This result, which is related to Theorem 5.1 in NM, generalises to the constrained case Proposition 4.7 in GM96, as well as the analogous result in Tauchen (1996).[2] Translated into words, Proposition (7) essentially says that the higher the (multivariate) correlation between the (average) modified score of the auxiliary model and the (average) score of the true model, the higher the efficiency of the constrained indirect estimator relative to the asymptotically efficient but potentially infeasible ML estimator. In particular, it immediately follows from it that the optimal "restricted" indirect estimators of $\boldsymbol{\rho}$ will achieve the usual asymptotic Cramer-Rao efficiency bound if and only if the residual covariance matrix $\mathcal{D}_0^r$ is zero.[3]

Proposition 2 in GT96 provides a leading example that guarantees this condition in the context of unrestricted indirect estimation. Specifically, GT96 show that full efficiency will be achieved if the auxiliary model "smoothly embeds" the true model, in the sense that there is an open neighbourhood of $\boldsymbol{\rho}^0$ in which the unrestricted binding function $\boldsymbol{\theta}^u(\boldsymbol{\rho})$ is twice continuously differentiable and $p(x_t|\mathbf{X}_{t-1};\boldsymbol{\rho}) = f\left[x_t|\mathbf{X}_{t-1};\boldsymbol{\theta}^u(\boldsymbol{\rho})\right]$.

[2] In Tauchen's case, $q_t(\boldsymbol{\rho}^0)$ and $l_t\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]$ are effectively strictly stationary and ergodic martingale difference sequences. As a consequence, $\mathcal{D}_0^u$ is simply the residual variance of the linear projection of $q_t(\boldsymbol{\rho}^0)$ on $l_t\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]$.

[3] In that case, we can show that the ML estimator of $\boldsymbol{\rho}$ will effectively depend on the data only through a continuously differentiable function of the first q elements of $\mathcal{K}_0^r\hat{\boldsymbol{\beta}}_T^r$ (cf. Chiang, 1959).

Regrettably, it is often the case that the auxiliary model does not nest the true model. However, as the following corollary illustrates, there are other cases in which we can achieve full efficiency by adding completely false constraints to a badly misspecified auxiliary model:

Corollary 1 *Consider the following stationary* AR*(1) process:*

$$x_t = \phi x_{t-1} + v_t, \quad v_t|\mathbf{X}_{t-1} \sim N(0, \omega), \quad |\phi| < 1, \quad 0 < \omega < \infty,$$

where the parameters of interest are $\boldsymbol{\rho} = (\phi, \omega)'$. *Then,*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right] - \hat{\boldsymbol{\rho}}_T\right\} = o_p(1)$$

regardless of the value of $\boldsymbol{\rho}^0$ *if the estimated auxiliary model is the following* MA*(1) model*

$$x_t = u_t - \delta u_{t-1}, \quad u_t|\mathbf{X}_{t-1} \sim N(0, \psi), \quad \psi \geq 0,$$

subject to the constraint $\delta = 0$.

Intuitively, the reason is that the auxiliary "parameter" estimators $\hat{\mu}_{\delta T}^e = \sum_t x_t x_{t-1} / \sum_t x_t^2$ and $\hat{\psi}_T^e = \sum_t x_t^2 / T$ are sufficient statistics for the true AR(1) log-likelihood. In contrast, the unrestricted indirect estimator of $\boldsymbol{\rho}$ is efficient only if $\phi^0 = 0$, which is precisely the only instance in which the unrestricted pseudo log-likelihood will smoothly embed the true log-likelihood.

Other more subtle examples of full asymptotic efficiency arise when the number of parameters of the auxiliary model is allowed to go to infinity. For instance, in the context of density estimation, Gallant and Tauchen (1999) (GT99) use earlier results by Gallant and Long (1997) and Tauchen (1996) to show that the score associated with the semi-non-parametric (SNP) density proposed by Gallant and Nychka (1987) - which multiplies a standard Gaussian density by a squared Hermite polynomial expansion - spans the true score in the limiting case in which the degree of the expansion goes to infinity. Similarly, GT99 also indicate that a GMM estimator based on an increasing sequence of integer moments of x_t can achieve full efficiency in the limit for those distributions that have a well-defined moment generating function. In this respect, we can show that such a GMM estimator is asymptotically equivalent for any finite integer power to an equality restricted indirect estimator based on the score of a SNP model of the same degree in which all the coefficients of the Hermite polynomial expansion are restricted to 0. Nevertheless, it is important to note that the appropriate rate at which extra terms can be added while preserving standard root-T asymptotics is unknown in both cases.

Proposition 7 may also seem to suggest that if we consider a somewhat more complicated auxiliary model, which implies that the number of components in $\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ will increase, then the new indirect estimators will be at least as efficient as those based on the original model because the limiting residual covariance matrix of the regression of $\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)$ on $\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ cannot increase by adding new "regressors". However, as GT99 point out, such a monotonicity property does not necessarily apply to unrestricted indirect estimators (see panel (a) in Figure 3 of GT99 for a counterexample). The reason is that when we unrestrictedly estimate an augmented auxiliary model, we are not simply adding new elements to its score, but also changing the parameter values at which we evaluate the original components.

In contrast, by considering the Lagrange multipliers associated with the implicit constraints that allow the nesting of the original model into the augmented one, we will always achieve asymptotic efficiency gains, in the sense that the equality constrained indirect estimators of $\boldsymbol{\rho}$ that take into account the information contained in those multipliers will be at least as efficient as the unconstrained indirect estimators based on the original auxiliary model.

More formally, consider a homeomorphic (i.e. one-to-one and bicontinuous) transformation $g(.) = [g_1'(.), g_2'(.)]'$ of the auxiliary model parameters $\boldsymbol{\theta}$ into an alternative set of $(c-s)+s$ parameters $\boldsymbol{\pi}=(\boldsymbol{\pi}_1', \boldsymbol{\pi}_2')'$, where $\boldsymbol{\pi}_2=g_2(\boldsymbol{\theta})=h(\boldsymbol{\theta})$, and $g(\boldsymbol{\theta})$ is twice continuously differentiable with rank$[\partial g'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]$ $= c$ in a neighbourhood of $\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)$. The purpose of this reparametrisation is to write the original implicit constraints $h(\boldsymbol{\theta}) = \mathbf{0}$ in explicit form as $\boldsymbol{\pi}_2 = \mathbf{0}$. Let $\hat{\boldsymbol{\pi}}_{1T}^u = g_1(\hat{\boldsymbol{\theta}}_T^e)$ denote the unconstrained PML estimator of $\boldsymbol{\pi}_1$ obtained by maximising with respect to $\boldsymbol{\pi}_1$ the auxiliary objective function $\bar{l}_T(\boldsymbol{\theta})$ reparametrised in terms of $\boldsymbol{\pi}$, with $\boldsymbol{\pi}_2$ set to $\mathbf{0}$. Similarly, let $\boldsymbol{\pi}_1^u(\boldsymbol{\rho}) = g_1[\boldsymbol{\theta}^e(\boldsymbol{\rho})]$ and $\mathfrak{m}_{\boldsymbol{\pi}_1}(\boldsymbol{\rho};\boldsymbol{\pi}_1) = E\left[\partial\bar{l}_T(\boldsymbol{\pi}_1;\mathbf{0})/\partial\boldsymbol{\pi}_1\middle|\,\boldsymbol{\rho}\right]$, denote the corresponding binding function and population moment condition, respectively, so that $\mathfrak{m}_{\boldsymbol{\pi}_1}[\boldsymbol{\rho};\boldsymbol{\pi}_1^u(\boldsymbol{\rho})] = 0$. In this context, we can define the unconstrained indirect estimators of $\boldsymbol{\rho}$ based on the original model, $\breve{\boldsymbol{\rho}}_T^u(\boldsymbol{\Phi})$ say, as the values of $\boldsymbol{\rho}$ that minimise the norm of $\mathfrak{m}_{\boldsymbol{\pi}_1}(\boldsymbol{\rho};\hat{\boldsymbol{\pi}}_{1T}^u)$ in the metric of a positive semidefinite matrix $\boldsymbol{\Phi}$ of order $(c-s)$, or some chosen distance between $\boldsymbol{\pi}_1^u(\boldsymbol{\rho})$ and $\hat{\boldsymbol{\pi}}_{1T}^u$. The rationale for such estimators would be that since $\boldsymbol{\pi}_2$ 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 $\breve{\boldsymbol{\rho}}_T^u(\boldsymbol{\Phi})$ is the estimator that all existing empirical implementations of indirect estimation procedures have effectively used in practice. As the following proposition shows, though, ignoring the Lagrange multipliers associated with the constraints $\boldsymbol{\pi}_2 = \mathbf{0}$ usually leads to asymptotic efficiency losses:

Proposition 8 *Under Assumptions 1, 2 and 3:*

1. $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^r)^{-1}\right]$ *is asymptotically at least as efficient as* $\breve{\boldsymbol{\rho}}_T^u(\boldsymbol{\Phi})$ *for any positive semidefinite matrix* $\boldsymbol{\Phi}$*, and*

2. *the optimal* $\breve{\boldsymbol{\rho}}_T^u(\boldsymbol{\Phi})$ *is asymptotically just as efficient as* $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^r)^{-1}\right]$ *if and only if the limiting covariance matrix between* $\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)$ *and* $\sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^e - \boldsymbol{\mu}^e(\boldsymbol{\rho}^0)\right]$ *is* $\mathbf{0}$ *after partialling out the effect of* $\sqrt{T}\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\boldsymbol{\pi}_1$.

Not surprisingly, we can show using (11) that the condition in part 2 of Proposition 8 is analogous in our context to condition (B) in Theorem 1 of Breusch et al. (1999). Therefore, there will be no efficiency gains in using $\hat{\boldsymbol{\mu}}_T^e$ if and only if the additional moment conditions associated with $\partial l_t\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\boldsymbol{\pi}_2$ have no incremental identification information about $\boldsymbol{\rho}$. At the same time, there are other circumstances in which $\boldsymbol{\pi}_1^u(\boldsymbol{\rho})$ would not suffice to identify $\boldsymbol{\rho}$, and hence, the relative efficiency gains from taking into account the information in $\hat{\boldsymbol{\mu}}_T^e$ would be infinite.[4]

Proposition 8 has important consequences for actual practice because any auxiliary parametric model contains a potentially very large number of implicit constraints, as the extensive literature on LM (or efficient score) tests illustrates (see e.g. Godfrey (1988) and the references therein). Moreover, in many situations of interest, it is considerably simpler to estimate a special restricted case of the auxiliary model than the unrestricted model itself. Therefore, given that in practice users of indirect estimation 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 indirect 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.

Finally, if $\boldsymbol{\theta}$ were the parameters of interest, and $f(x_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ provided the correct conditional density function for x_t, the imposition of correct equality restrictions on $\boldsymbol{\theta}$ would weakly improve the efficiency of the resulting estimators (see e.g. Rothenberg (1973) for details). However, such a result is not necessarily robust to misspecification of the density function, even if both $\hat{\boldsymbol{\theta}}_T^u$ and $\hat{\boldsymbol{\theta}}_T^e$ 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 estimation set up, in which both the density function of the auxiliary model and the restrictions on $\boldsymbol{\theta}$ 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 define the auxiliary objective function, we are implicitly changing the auxiliary model, and thereby, the binding functions. Therefore, a discussion of the efficiency consequences of imposing equality constraints on a previously unrestricted auxiliary model will typically require us to compare the residual covariance matrices $\mathcal{D}_0^u$ and $\mathcal{D}_0^e$ defined in Proposition 7. Nevertheless, we can state the following sufficient condition for asymptotic equivalence:

Proposition 9 *Under Assumptions 1, 2 and 3*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]\right\}=o_p(1)$$

if

$$\lim_{T\to\infty} V\left\{\sqrt{T}\bar{m}_T\left[\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)\right]-\mathcal{H}_0\sqrt{T}\bar{m}_T\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]\right\}=0,$$

where $\mathcal{H}_0$ is the matrix of limiting projection coefficients, and rank$(\mathcal{H}_0)=s$.

Intuitively, this condition says that the two estimators are asymptotically equivalent if their modified scores generate the same linear span. A particularly important example is given by the following result:

Corollary 2 *Under Assumptions 1, 2 and 3*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]\right\}=o_p(1)\quad \text{if } h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]=\mathbf{0}.$$

[4] As an extreme case, suppose that $s=c\geq d$, and that $h(\boldsymbol{\theta})=\boldsymbol{\theta}-\boldsymbol{\theta}^\dagger$, so that the only admissible value for the equality restricted estimator $\hat{\boldsymbol{\theta}}_T^e$ is precisely $\boldsymbol{\theta}^\dagger$. In this case, the dimension of $\boldsymbol{\pi}_1$ would be zero, and no unconstrained indirect estimator based on those inexistent parameters could be defined. In contrast, our equality constrained indirect estimation procedures will work by simply matching the c equality restricted binding functions $\boldsymbol{\mu}^e(\boldsymbol{\rho})$ with the sample estimates of the c Lagrange multipliers.

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 indirect 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 the equivalence between $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]$ and $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]$ may also hold with incorrect constraints. For instance, this is always the case when the auxiliary model is a linear autoregression with drift, and the restrictions are linear in the autoregression coefficients. More formally:

Corollary 3 *Under Assumptions 1, 2 and 3*

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]\right\}=o_p(1)$$

if

$$l_t(\boldsymbol{\theta})=-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\omega-\frac{1}{2\omega}(x_t-\phi_0-\phi_1 x_{t-1}-\ldots-\phi_k x_{t-k})^2,$$

and $h(\boldsymbol{\theta})=R\boldsymbol{\phi}-r$, *with* $rank(R')=s$, *where* $\boldsymbol{\phi}=(\phi_0,\phi_1,\ldots,\phi_k)$, *and* $\boldsymbol{\theta}=(\boldsymbol{\phi}',\omega)'$.

Note that such a result does not really depend on the nature of the true model, whose parameters only enter through the mean of x_t, $\nu(\boldsymbol{\rho})=E\left(x_t|\,\boldsymbol{\rho}\right)$, and its first $k+1$ theoretical "autocovariances", $\gamma_j(\boldsymbol{\rho})=E\left(x_t x_{t-j}|\,\boldsymbol{\rho}\right)$ $(j=0,\ldots,k)$, but rather on the particular form of the auxiliary model used. Intuitively, the reason is that from the point of indirect estimation, the sample mean $\bar{x}_T$ and the first $k+1$ sample "autocovariances" $\bar{\gamma}_{jT}^r$ $(j=0,\ldots,k)$ 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 $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]$ and $\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]$ can be derived under general circumstances. For the sake of clarity, though, we shall only present a formal result in the case of a single restriction:

Proposition 10 *Under Assumptions 1, 2 and 3*

$$\begin{aligned}\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_0^i)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]\right\} &= o_p(1) \quad \textit{if } h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]>\mathbf{0},\\ \sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_0^i)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]\right\} &= o_p(1) \quad \textit{if } h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]<\mathbf{0},\end{aligned}$$

and

$$\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_0^i)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]\right\}=o_p(1)=\sqrt{T}\left\{\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_0^i)^{-1}\right]-\tilde{\boldsymbol{\rho}}_T^e\left[(\mathcal{I}_0^e)^{-1}\right]\right\} \quad \textit{if } h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]=\mathbf{0}.$$

In other words, the optimal inequality restricted indirect estimator of $\boldsymbol{\rho}$ will converge to the optimal unrestricted indirect estimator if the inequality restriction is "correct", in the sense that $h[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)]>0$, where $\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)$ is the unrestricted pseudo true value of $\boldsymbol{\theta}$, while it will converge to the equality restricted one when the constraint is "incorrect", by which we mean that $h[\boldsymbol{\theta}^u(\boldsymbol{\rho})]<0$.

In fact, the *inequality constrained* and *unconstrained* indirect procedures will yield numerically identical results if the inequality restriction is not binding in a given sample, since in that case $\hat{\boldsymbol{\theta}}_T^i$ coincides with the unconstrained PML estimator, $\hat{\boldsymbol{\theta}}_T^u$ (and $\hat{\boldsymbol{\mu}}_T^i$ with $\hat{\boldsymbol{\mu}}_T^u = \mathbf{0}$). Similarly, the *inequality* and *equality constrained* procedures will yield numerically identical results if the inequality restriction is binding in a given sample, because in that case $\hat{\boldsymbol{\theta}}_T^i$ coincides with the *equality constrained* PML estimator, $\hat{\boldsymbol{\theta}}_T^e$, and consequently $\hat{\boldsymbol{\mu}}_T^i$ with $\hat{\boldsymbol{\mu}}_T^e$. In the case of multiple inequality constraints, $\tilde{\boldsymbol{\rho}}_T^i\left[(\mathcal{I}_0^i)^{-1}\right]$ will numerically coincide with either the unrestricted estimator, or an equality restricted estimator that imposes the subset of the s constraints that happen to be satisfied with equality by $\hat{\boldsymbol{\theta}}_T^i$. Therefore, it is not surprising that the *inequality constrained* optimal indirect estimator will be asymptotically equivalent to $\tilde{\boldsymbol{\rho}}_T^u\left[(\mathcal{I}_0^u)^{-1}\right]$ if $h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right] > \mathbf{0}$, or to some equality restricted estimator otherwise.

2.3 Extensions

One approach commonly followed by users of indirect 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 thereby, coming close to the idealised situation of $\mathcal{D}_0^u = 0$ discussed before. Unfortunately, such attempts often encounter numerical optimisation problems (see Andersen, Chung, and Sorensen (1999) (ACS)). 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 indirect estimation 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 $k_{\max} \geq 0$.

Let us initially consider the case of $k_{\max} = 0$, so that no optimisation whatsoever takes place. If the initial value $\hat{\boldsymbol{\theta}}_T^{(0)}$ is non-stochastic, $\boldsymbol{\theta}^{(0)}$ say, we simply have a special case of the equality constrained GMM-based indirect estimator, with the restrictions $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$. In effect, this transforms the GMM indirect estimation procedure in a CMD indirect estimation procedure in which we match the values of the multipliers $\hat{\boldsymbol{\mu}}_T^{(0)}$ 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 2 and 3. Typically, however, $\hat{\boldsymbol{\theta}}_T^{(0)}$ 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 Calzolari, Fiorentini and Sentana (2004) for an example). If that is the case, the results in section 2.1 imply that the *fully non-optimised* GMM indirect estimator of $\boldsymbol{\rho}$ based on $\hat{\boldsymbol{\theta}}_T^{(0)}$ and $\hat{\boldsymbol{\mu}}_T^{(0)}$, $\tilde{\boldsymbol{\rho}}_T^{(0)}$ say, will be consistent and asymptotically normal, as long as the regularity conditions in Assumptions 2 and 3 (with $\partial h'(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{I}_c$) remain valid when (i) $\hat{\boldsymbol{\theta}}_T^r$ is replaced by $\hat{\boldsymbol{\theta}}_T^{(0)}$, (ii) $\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)$ by the pseudo-true value of $\hat{\boldsymbol{\theta}}_T^{(0)}$, $\boldsymbol{\theta}^{(0)}(\boldsymbol{\rho}^0)$ say, (iii) $\hat{\boldsymbol{\mu}}_T^r$ by $\hat{\boldsymbol{\mu}}_T^{(0)}$, which are the Lagrange multipliers required to satisfy the sample first-order conditions (3) at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_T^{(0)}$, and (iv) $\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)$ by the corresponding pseudo-true value, $\boldsymbol{\mu}^{(0)}(\boldsymbol{\rho}^0)$.

Let us now consider the more interesting case of $k_{\max} = 1$. It is then clear that $\hat{\boldsymbol{\theta}}_T^{(1)}$ and $\hat{\boldsymbol{\mu}}_T^{(1)}$ will also be stochastic, with pseudo-true values given by $\boldsymbol{\theta}^{(1)}(\boldsymbol{\rho}^0) = \boldsymbol{\theta}^{(0)}(\boldsymbol{\rho}^0) + \mathcal{J}_0^{(0)}\boldsymbol{\mu}^{(0)}(\boldsymbol{\rho}^0)$ and $\boldsymbol{\mu}^{(1)}(\boldsymbol{\rho}^0) = -E\left\{\partial\bar{l}_T\left[\boldsymbol{\theta}^{(1)}(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\middle|\boldsymbol{\rho}^0\right\}$. If, *mutatis mutandi*, the regularity conditions in Assumptions 2 and 3 remain valid, then the *one-step optimised* GMM estimator of $\boldsymbol{\rho}$ based on $\hat{\boldsymbol{\theta}}_T^{(1)}$ and $\hat{\boldsymbol{\mu}}_T^{(1)}$, $\tilde{\boldsymbol{\rho}}_T^{(1)}$ say, will also be consistent and asymptotically normal. But since the above argument does not really depend on $k_{\max}$ being 1, or the way in which $\hat{\boldsymbol{\theta}}_T^{(0)}$ was obtained, it remains valid for any $k_{\max}$. 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, and (ii) to cater for an increasing number of practitioners who use the SNP auxiliary model with an 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 indirect 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.

For analogous reasons, an empirical researcher may alternatively decide to conduct a specification test to assess if there is any evidence for an additional feature of the data that she has not yet incorporated in her auxiliary model. Since most existing specification tests are of the LM form, a numerically sensible strategy could be to base the indirect 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 she exploits the information in the corresponding Lagrange multiplier. 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 indirect estimator of $\boldsymbol{\rho}$ is the same as the limiting distribution of the fully optimised unconstrained indirect 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 unconstrained estimators would have the same distribution under the (pseudo) null from Corollary 2, then the pre-test estimator will share the same asymptotically normal distribution. We shall look at the empirical performance of such a pre-test estimator in the next section.

3 An illustrative example: Stochastic volatility estimated as GARCH(1,1) with Gaussian and Student's t innovations

Consider the following log-normal stochastic volatility process

$$\begin{aligned} x_t &= \sqrt{h_t} u_t \\ \ln h_t &= \alpha + \delta \ln h_{t-1} + \sigma_v v_t \end{aligned} \tag{12}$$

where $|\delta| < 1, 0 < \sigma_v < \infty$, and $(u_t, v_t)'|\mathbf{X}_{t-1} \sim N(0, \mathbf{I}_2)$. This model was originally proposed as an alternative to the ARCH class of volatility models, 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. In this context, the autoregressive parameter δ is typically interpreted as a measure of the persistence of shocks to the volatility process, while the standard deviation parameter σ_v reflects its instantaneous sensitivity to those shocks. Finally, α is a simple scaling parameter that determines the average volatility level.

Unfortunately, it is impossible to find analytical expressions for the conditional distribution of x_t based on its own past values alone, despite the fact that its distribution conditional on $h_t, x_{t-1}, \ldots$ is Gaussian, with zero mean and variance h_t. Given its importance, though, it is not surprising that a voluminous collection of research papers has been devoted to the estimation of the parameters of interest $\boldsymbol{\rho} = (\alpha, \delta, \sigma_v)'$ (see ACS for a recent survey). In an influential such paper, Kim, Shephard and Chib (1998) (KSC) consider likelihood-based estimators of (12), 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 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

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

where $\varepsilon_t|\mathbf{X}_{t-1}$ follows a standardised Student's t distribution with η^{-1} degrees of freedom, so that $\boldsymbol{\theta} = (\psi, \varphi, \pi, \eta)'$. As is well known, the standardised t distribution nests the standard normal for $\eta = 0$, but has otherwise fatter tails. Broadly speaking, ψ and $(\varphi + \pi)$ play in the GARCH model the roles that α and δ have in the stochastic volatility model (12), while there is no such a close counterpart to σ_v. Nevertheless, it is important to note that the auxiliary and true models are non-nested, except in the trivial case in which x_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), the conditional variance λ_t will be nonnegative with probability one if $\psi \geq 0$, $\varphi \geq 0$ and $\pi \geq 0$.

2. The PML estimators of $\boldsymbol{\theta}$ may not be well behaved when $\varphi + \pi > 1$ (see Lumsdaine, 1996).

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

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

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

$$\psi \geq 0, \quad \varphi \geq \varphi_{\min}, \quad \pi \geq 0, \quad \varphi + \pi \leq 1, \quad 0 \leq \eta \leq \eta_{\max} \tag{13}$$

where $\varphi_{\min}$, and $1/2 - \eta_{\max}$ are arbitrarily chosen small values.[5]

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. For that reason, we also consider an 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. We also compute a third estimator along the lines described in section 2.3, which alternates between the previous two depending on whether or not the value of the one-sided LM normality test proposed by Fiorentini, Sentana and Calzolari (2003) exceeds the relevant 5% critical level. Finally, we consider a fourth 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. For the sake of brevity, we shall refer to the estimator that allows η to vary freely within its bounds as the "inequality restricted" indirect estimator, to the one that sets η to 0 as the "equality restricted" indirect estimator, to the mixed one as the "pre-test" indirect estimator, and to the fourth one as the "unrestricted" indirect estimator. In all cases, though, the remaining auxiliary parameters are always estimated subject to the other bounds in (13).

We assess the performance of our proposed procedures by means of an extended Monte Carlo analysis, with the same experimental designs 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 is the unconditional coefficient of variation of the unobserved volatility level h_t, κ say, where

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

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 $\kappa^{2\prime}s$ (around .5) are the rule, rather than the exception (see JPR and the references therein). JPR considered nine Monte Carlo designs, arranged in three rows and columns. The

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

rows are defined in terms of κ^2, and the columns by the autocorrelation coefficient for log volatility, δ. Finally, the remaining parameter α is chosen for scaling purposes so that the unconditional mean of the volatility level equals .0009. Although most of their reported results correspond to a sample size of $T = 500$ observations, we have also considered $T = 1,000$ and $2,000$. For the sake of brevity, though, we only report the results for smallest and largest sample sizes, and two designs: $\rho^0 = (-.736, .9, .363)'$ ($\kappa^2 = 1$) and $\rho^0 = (-.141, .98, .0614)'$ ($\kappa^2 = .1$), which roughly match what we tend to see in the empirical literature with weekly and daily data, respectively.

For convenience, we first optimise the pseudo log-likelihood function in terms of some unrestricted parameters $\boldsymbol{\theta}^*$, where $\psi = \theta_1^{*2}$, $\varphi = \varphi_{\min} + (1-\varphi_{\min})\sin^2(\theta_2^*)$, $\pi = (1-\varphi)\sin^2(\theta_3^*)$ and $\eta = \sin^2(\theta_4^*)\eta_{\max}$. Then, we compute the score in terms of the original parameters $\boldsymbol{\theta} = (\psi, \varphi, \pi, \eta)'$ using the analytical expressions derived by Fiorentini, Sentana and Calzolari (2003) 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 TH, with $H = 10$, as explained in the appendix. A larger value of H should in theory reduce the Monte Carlo variability of the indirect estimators according to the relation $(1 + H^{-1})$, but at the cost of a significant increase in the computational burden. Finally, we minimise numerically the GMM criterion function in terms of some unrestricted parameters $\boldsymbol{\rho}^*$, with $\alpha = \rho_1^*$, $\delta = \delta_{\max}\sin(\rho_2^*)$ and $\sigma_v = \rho_3^{*2}$, where $\delta_{\max} = .9999$, so as to ensure that $|\delta| < 1$ and $\sigma_v \geq 0$. In order to avoid the biases that an infeasible choice of initial values could induce on our Monte Carlo results, we decided to follow a sensible estimation strategy that an empirical researcher could repeat with real data. Specifically, we considered three different sets of initial values for the stochastic volatility parameters $\boldsymbol{\rho}$:

1) a method of moments estimator based on $E(x_t^2)$, $V(x_t^2)$ and $cov(x_t^2, x_{t-1}^2)$, with δ restricted to be between .01 and .99,

2) another method of moments estimator based on $E(\ln x_t^2)$, $V(\ln x_t^2)$ and $cov(\ln x_t^2, \ln x_{t-1}^2)$, with δ restricted again to be in the same range, and

3) an average of the previous two.

Then, we looked at the minimum value of the GMM criterion function, and reported the parameter estimates that corresponded to the *minimum minimorum*. Our procedure has two additional advantages that are important in practice: it uses consistent initial values, and it performs a sensitivity analysis of the convergence of the numerical optimisation procedure.

Table 1 contains the proportion of inequality and equality restricted PML estimators of $\boldsymbol{\theta}$ that satisfy with equality the different restrictions in (13). In this respect, note that the auxiliary model estimated by the unrestricted procedure coincides with the model estimated by the equality restricted one. When

κ^2 is 1, such restrictions are almost never binding, especially for $T = 2,000$. However, when κ^2 is .1, parameter configurations in which $\varphi + \pi = 1$ (i.e. IGARCH) are hardly ever estimated, but the estimates of the ARCH and GARCH coefficients φ and π, respectively, and the reciprocal of the degrees of freedom parameter η, reach their lower bounds fairly often, especially for the smaller sample size. In particular, when $T = 500$, 40% of the simulations have PML estimators based on the normal distribution for which at least one of the inequality restrictions on the ARCH and GARCH coefficients is binding, a percentage that rises to almost 60% in the case of the student t. 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 indirect estimation procedures are not only of theoretical interest, but also highly relevant in practice.

Figures 1 and 2 display kernel estimates of the sampling distributions of the "unrestricted", "equality restricted", "inequality restricted", and "pre-test" GMM-based indirect estimators of the structural parameters δ and σ_v for the case in which the optimal weighting matrix is estimated using the variance in the original data of the modified score of the auxiliary model evaluated at the PML parameter estimates.[6] In this respect, note that by including a multiplier in each first order condition, we automatically centre the scores around their sample mean. Given that the auxiliary model tends to fit the simulated data rather well, 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. GT96). As for bandwidth, we have used the automatic choice given in expression (3.29) in Silverman (1986).

In line with the existing literature (see ACS), we find that the sampling distributions of the different estimators of the autoregressive parameter δ are systematically skewed to the left. This is particularly so when δ^0 is high and σ_v^0 low, which mimics the behaviour of a PML estimator of the autoregressive parameter of an AR(1) process observed subject to measurement error. Intuitively, the reason for such a similarity is that the first equation in (12) can be transformed into the measurement equation $\ln x_t^2 = \ln h_t + \ln u_t^2$. And exactly like in that situation, the downward bias in the estimator of δ is transmitted into an upward bias in the absolute value of the estimates of the mean constant, α, and the standard deviation of the log-volatility innovations σ_v, whose sampling distribution is skewed to the right. Therefore, it is not surprising that the most important determinant of the performance of the estimators is precisely κ^2, which effectively plays the role of a signal to noise ratio.

If we now compare the "unrestricted" indirect estimators with the equality restricted estimator, the

[6] Note that since the "unrestricted" indirect estimator is effectively using a just-identified auxiliary model, it is generally invariant to the weighting matrix. Nevertheless, by using the optimal weighting matrix, we ensure that the objective function is evenly scaled across parameters, which improves the numerical properties of the optimisation algorithm.

most noticeable effect of taking into account the information in the score for η evaluated at $\eta = 0$ is that the precision with which we estimate the volatility of volatility parameter σ_v increases substantially, the more so the smaller the signal to noise ratio. This is due to the fact that σ_v is the parameter that most directly influences the degree of leptokurtosis of the conditional distribution of x_t, which is mainly captured in the GARCH model through the value of η, or its associated multiplier. As for the autoregressive parameter δ, the reported simulation evidence also confirms the efficiency gains stated in Proposition 8, although for $\kappa = .1$ large sample sizes seem to be required for normal asymptotics to apply.

In contrast, the "equality", "inequality" and "pre-test" versions of the indirect estimator are quite close to each other for these two Monte Carlo designs. Although the equality restricted indirect estimator typically outperforms the inequality restricted one, with the "pre-test" estimator being somewhere in between, the differences are minor. In this respect, it is important to point out that the pseudo-true values of η reported in Table 1, which were computed on the basis of $500,000$ observations, are different from zero, especially for $\kappa = 1$, which means that the sufficient condition for asymptotic equivalence stated in Corollary 2 and Proposition 10 does not apply.

Finally, a comparison of our results with the ones reported by JPR and ACS suggests that our indirect estimation procedures tend to outperform the PML and GMM estimators described in those papers for the realistic parametric configurations that we consider. In contrast, our indirect estimators are dominated by the Bayesian estimators proposed by JPR and KSC, 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 sense, it is important to mention that the relatively good performance of the Bayesian estimators in small samples is partly due to the imposition of priors that assign low density near the boundary values of the domains of δ and especially σ_v.

4 Conclusions

In this paper, we generalise the indirect estimation approaches of GT96 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 GT96 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 this respect, we consider not only the usual two-step GMM method proposed by GT96, but also a continuously updated one (à la Hansen, Heaton and Yaron, 1996). In addition, we combine the "constrained" parameter estimators and Lagrange/Kuhn-Tucker multipliers to extend the original class of CMD indirect estimators of GMR to the possibly restricted case, and prove that one can find "restricted" CMD indirect estimators that are asymptotically equivalent to the GMM estimators by an appropriate choice of weighting matrix.

Inequality restrictions must often be considered in practice because the pseudo log-likelihood function may not be well defined when certain parameter restrictions are violated, some of the auxiliary parameters may become (almost) underidentified in certain regions of the auxiliary parameter space, and/or some of the implications of the auxiliary model could be unacceptable from an economic viewpoint. 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, our second contribution is an extensive discussion of the impact of the constraints on the efficiency of the resulting indirect estimators. To do so, we first relate the asymptotic efficiency of our indirect estimators to the usually infeasible ML estimator. Then, we show that the asymptotic efficiency of indirect 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. 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 indirect estimators, and others in which false constraints enable the restricted indirect 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.

Finally, we also introduce indirect estimators based on partially optimised unconstrained estimators, as well as those that impose the constraints depending on the significance of some preliminary specification test. Such estimators are particularly useful in practice, especially when the auxiliary model is rather complex (e.g. an SNP specification with a large number of terms in the Hermite expansion) because they can successfully transform an unsuccessful attempt to optimise the pseudo-log likelihood function into a consistent estimator of the parameters of interest.

For illustrative purposes, 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 PML estimators are quite often at the boundary of the parameter space, which confirms the practical relevance of our proposed procedures. We also document that when the auxiliary model is estimated under Gaussianity, we can substantially increase the efficiency of the usual indirect estimators by including the information in the multiplier corresponding to the reciprocal of the degrees of freedom.

Further work is required in at least three main directions. From a numerical point of view, sequential indirect estimators may be the only feasible alternative in large-scale multivariate models involving many parameters (see Calzolari, Fiorentini and Sentana (2004) for further details and an application).

From a statistical point of view, the finite sample reliability of the asymptotic results obtained in this paper may sometimes be questionable. For instance, we may need fairly large sample sizes for the asymptotic distributions of the indirect estimators presented in Proposition 4 to approximate their finite sample distributions, as we saw in some of the Monte Carlo experiments reported in section 3. And although our simulations suggest otherwise, the asymptotic efficiency gains promised by Proposition 8, which result from considering additional Lagrange multipliers, may not always materialise in small samples. In this context, there are several research avenues that would be worth exploring. In particular, the recent GMM literature suggests that the continuously updated GMM-like methods suggested in section 2.1 may improve the finite sample performance of our indirect estimators. In addition, constrained versions of the implicit bias adjustment procedures discussed by Gouriéroux, Renault and Touzi (2000), either on their own (see Arvanitis and Demos (2003)), or together with the control variates techniques developed by Calzolari, Di Iorio and Fiorentini (1998) may also prove useful in this respect. Similarly, the Laplace-type procedures recently proposed by Chernozhukov and Hong (2003) may result in estimators with better finite sample properties, particularly in dynamic latent variable models, like the SV example discussed in section 3, in which the role of prior information on MCMC-based Bayesian estimators is non-negligible.

Finally, from a modelling point of view, the application of constrained indirect estimation procedures and other simulation-based inference methods to dynamic models with predetermined (as opposed to strictly exogenous) conditioning variables remains an important unresolved issue that merits further investigation.

References

Andersen, T.G., Chung, H-J., and Sorensen, B.E. (1999): "Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study", *Journal of Econometrics* 91, 61-87.

Andrews, D.W.K. (1999): "Estimation when a parameter is 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.

Arvanitis, S. and Demos, A. (2003): "Aspects of the indirect inference theory", mimeo, Athens University of Economics and Business.

Breusch, T., Qian, H., Schmidt, P. and Wyhowski, D. (1999): "Redundancy of moment conditions", *Journal of Econometrics* 91, 89-111.

Calzolari G., Di Iorio, F. and Fiorentini, G. (1998): "Control variates for variance reduction in indirect inference: interest rate models in continuous time", *Econometrics Journal* 1, C100-C112.

Calzolari, G., Fiorentini, G. and Sentana, E. (2000): "Constrained EMM and indirect inference estimation", CEMFI Working Paper 0005, available from <ftp://ftp.cemfi.es/wp/00/0005.pdf>.

Calzolari, G., Fiorentini, G. and Sentana, E. (2004): "Indirect estimation of conditionally heteroskedastic factor models", mimeo, CEMFI.

Chernozhukov, V. and Hong, H. (2003): "An MCMC approach to classical estimation", *Journal of Econometrics* 115, 293-346.

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.

Fiorentini, G., Sentana, E. and Calzolari, G. (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with student t innovations", *Journal of Business and Economic Statistics* 21, 532-546.

Gallant, A.R. and Long, J. (1997): "Estimating stochastic differential equations efficiently by minimum chi-squared", *Biometrika* 84, 125-141.

Gallant, A.R. and Nychka, D.W. (1987): "Seminonparametric maximum likelihood estimation", *Econometrica* 55, 363-390.

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

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

Godfrey, L.G. (1988): *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Cambridge.

Gouriéroux, C. and Monfort, A. (1995): *Statistics and econometric models*, vols. I and II, Cambridge.

Gouriéroux, C. and Monfort, A. (1996): *Simulation-based econometric methods*, Oxford.

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

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

Gouriéroux, C., Renault, E. and Touzi, N. (2000): "Calibration by simulation for small sample bias correction", in R. Mariano, T. Shuermann and M. Weeks (eds.) *Simulation-based inference in econometrics. Methods and applications*, 328-358, Cambridge.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029-1054.

Hansen, L.P., Heaton J. and Yaron A. (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.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

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

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*, 1-67. Chapman and Hall.

Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

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

Tauchen, G. (1996): "New minimum chi-square methods in empirical finance", in D. Kreps and K.

Wallis (eds.), *Advances in Economic Theory and Econometrics, VII World Congress*, 279-317, Cambridge.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1-28.

Appendix

Proofs of results

Proposition 1

If we linearise the complementary slackness conditions

$$h(\hat{\boldsymbol{\theta}}_T^r) \odot \hat{\boldsymbol{\mu}}_T^r = 0$$

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

$$\boldsymbol{\mu}_T^* \odot \frac{\partial h(\boldsymbol{\theta}_T^*)}{\partial \boldsymbol{\theta}'}\sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] + h(\boldsymbol{\theta}_T^*) \odot \sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right] = 0$$

where $\boldsymbol{\beta}_T^* = (\boldsymbol{\theta}_T^{*\prime}, \boldsymbol{\mu}_T^{*\prime})'$ is an "intermediate" value (in fact, a different one for each row). Then, given that in view of our high level assumptions, $\boldsymbol{\mu}_T^* - \boldsymbol{\mu}^r(\boldsymbol{\rho}^0) = o_p(1)$, $h(\boldsymbol{\theta}_T^*) - h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] = o_p(1)$, and $\partial h(\boldsymbol{\theta}_T^*)/\partial\boldsymbol{\theta} - \partial h\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta} = o_p(1)$, the result follows. □

Proposition 2

If we linearise the first-order conditions

$$\sqrt{T}\bar{m}_T(\hat{\boldsymbol{\theta}}_T^r) = \mathbf{0}$$

around $\boldsymbol{\beta}_T^r(\boldsymbol{\rho}^0)$, we obtain:

$$\sqrt{T}\bar{m}_T\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] + \left\{\frac{\partial^2 \bar{l}_T(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + (\boldsymbol{\mu}_T^\star \otimes I_c)\frac{\partial vec\left[\partial h'(\boldsymbol{\theta}_T^\star)/\partial\boldsymbol{\theta}\right]}{\partial\boldsymbol{\theta}'}\right\}\sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] + \frac{\partial h'(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}}\sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right] = \mathbf{0}$$

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

$$\begin{aligned}
\frac{\partial^2 \bar{l}_T(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= \mathcal{J}_0^r + o_p(1) \\
(\boldsymbol{\mu}_T^\star \otimes I_c)\frac{\partial vec\left[\partial h'(\boldsymbol{\theta}_T^\star)/\partial\boldsymbol{\theta}\right]}{\partial\boldsymbol{\theta}'} &= \left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^0) \otimes I_c\right]\frac{\partial vec\left\{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\right\}}{\partial\boldsymbol{\theta}'} + o_p(1) \\
\frac{\partial h'(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}} &= \frac{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}} + o_p(1)
\end{aligned}$$

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

Proposition 3

Let us now linearise the sample moments $\mathfrak{m}(\boldsymbol{\rho}^0; \hat{\boldsymbol{\beta}}_T^r)$ around $\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)$ to obtain

$$\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0; \hat{\boldsymbol{\beta}}_T^r) = \sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] + \frac{\partial\mathfrak{m}(\boldsymbol{\rho}^0; \boldsymbol{\beta}_T^\diamond)}{\partial\boldsymbol{\theta}'}\sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] + \frac{\partial\mathfrak{m}(\boldsymbol{\rho}^0; \boldsymbol{\beta}_T^\diamond)}{\partial\boldsymbol{\mu}'}\sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]$$

where $\boldsymbol{\beta}_T^\diamond$ is yet another "intermediate" value. Given that $\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right] = \mathbf{0}$, this implies that under our assumptions, $\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0; \hat{\boldsymbol{\beta}}_T^r)$ has the same asymptotic distribution as

$$\frac{\partial\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}'}\sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r - \boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right] + \frac{\partial\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\mu}'}\sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^r - \boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]$$

where

$$\begin{aligned}
\frac{\partial\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}'} &= \mathcal{J}_0^r + \left[\boldsymbol{\mu}^r(\boldsymbol{\rho}^0) \otimes I_c\right]\frac{\partial vec\left\{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\theta}\right\}}{\partial\boldsymbol{\theta}'} = \mathcal{K}_{11,0}^r \\
\frac{\partial\mathfrak{m}\left[\boldsymbol{\rho}^0; \boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\mu}'} &= \frac{\partial h'\left[\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}} = \mathcal{K}_{12,0}^r
\end{aligned}$$

But then, Proposition 2 directly yields the required result □

Proposition 4

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

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

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

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

where $\boldsymbol{\rho}_T^*$ is some "intermediate" value. But since $\mathfrak{m}\left(\boldsymbol{\rho}_T^*;\hat{\boldsymbol{\beta}}_T^r\right)$ 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

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

as required. □

Proposition 5

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

$$\sqrt{T}\mathfrak{m}(\boldsymbol{\rho}^0;\hat{\boldsymbol{\beta}}_T^r)-\left\{\mathcal{K}_{11,0}^r\sqrt{T}\left[\hat{\boldsymbol{\theta}}_T^r-\boldsymbol{\theta}^r(\boldsymbol{\rho}^0)\right]+\mathcal{K}_{12,0}^r\sqrt{T}\left[\hat{\boldsymbol{\mu}}_T^r-\boldsymbol{\mu}^r(\boldsymbol{\rho}^0)\right]\right\}=o_p(1)$$

□

Proposition 6

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

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

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

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

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

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

the first order conditions that characterise the binding functions imply that

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

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

Proposition 7

The fact that $\mathcal{D}_0^r$ is the asymptotic residual covariance matrix in the limiting least squares projection of $\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)$ onto $\sqrt{T}m_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ follows from (11) and the second part of Assumption 3. But then, since the projection error will be asymptotically orthogonal to the "regressors" $\sqrt{T}m_T\left[\boldsymbol{\beta}^r(\boldsymbol{\rho}^0)\right]$ by the usual first order condition of least squares projections, it trivially follows that $\mathcal{B}_0=\mathcal{C}_0^r+\mathcal{D}_0^r$. □

Corollary 1

Let $\boldsymbol{\theta} = (\delta, \psi)'$ denote the auxiliary model parameters, and $\boldsymbol{\mu} = (\mu_\delta, \mu_\psi)'$ the multipliers associated with the constraints $\delta = 0$ and $\psi \geq 0$, respectively. The average pseudo log-likelihood function of the MA(1) model for a sample of size T (ignoring initial conditions) will be given by:

$$\bar{l}_T(\boldsymbol{\theta}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\psi - \frac{1}{2\psi}\frac{1}{T}\sum_t [x_t - \nu_t(\delta)]^2,$$

and the (scaled) Lagrangian function by

$$Q_T(\boldsymbol{\beta}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\psi - \frac{1}{2\psi}\frac{1}{T}\sum_t [x_t - \nu_t(\delta)]^2 + \delta\mu_\delta + \psi\mu_\psi,$$

where

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

Since

$$\begin{aligned}
\bar{m}_{\delta T}(0, \psi, \mu_\delta, \mu_\psi) &= \frac{1}{\psi}\frac{1}{T}\sum_t u_t(0)\frac{\partial\nu_t(0)}{\partial\delta} + \mu_\delta = -\frac{1}{\psi}\frac{1}{T}\sum_t x_t x_{t-1} + \mu_\delta, \\
\bar{m}_{\psi T}(0, \psi, \mu_\delta, \mu_\psi) &= \frac{1}{2\psi}\frac{1}{T}\sum_t \left[\frac{u_t^2(0)}{\psi} - 1\right] + \mu_\psi = \frac{1}{2\psi}\frac{1}{T}\sum_t \left[\frac{x_t^2}{\psi} - 1\right] + \mu_\psi,
\end{aligned}$$

where

$$u_t(\delta) = \sum_{j=0}^{\infty}\delta^j x_{t-j}, \quad \frac{\partial\nu_t(\delta)}{\partial\delta} = -\sum_{j=1}^{\infty} j\delta^{j-1}x_{t-j},$$

it is easy to see that

$$\begin{aligned}
\mathfrak{m}_\delta[(\phi, \omega);(0, \psi, \mu_\delta, \mu_\psi)] &= E\left[-\frac{1}{\psi}x_t x_{t-1} + \mu_\delta \middle| \phi, \omega\right] = \frac{-\phi\omega}{\psi(1-\phi^2)} + \mu_\delta, \\
\mathfrak{m}_\psi[(\phi, \omega);(0, \psi, \mu_\delta, \mu_\psi)] &= E\left[\frac{1}{\psi}\left(\frac{x_t^2}{\psi} - 1\right) + \mu_\psi \middle| \phi, \omega\right] = \frac{1}{2\psi^2}\left(\frac{\omega}{1-\phi^2} - \psi\right) + \mu_\psi.
\end{aligned}$$

From here, it is clear that the binding functions $\boldsymbol{\beta}^e(\boldsymbol{\rho})$ that satisfy the moment conditions $\mathfrak{m}[\boldsymbol{\rho}; \boldsymbol{\beta}^e(\boldsymbol{\rho})] = \mathbf{0}$, together with the exclusion restriction $\psi^e(\boldsymbol{\rho}) \cdot \mu_\psi^e(\boldsymbol{\rho}) = 0$, plus the original parametric restrictions $\delta^e(\boldsymbol{\rho}) = 0$ and $\psi^e(\boldsymbol{\rho}) \geq 0$, will be given by

$$\delta^e(\boldsymbol{\rho}) = 0, \quad \mu_\delta^e(\boldsymbol{\rho}) = \phi, \quad \psi^e(\boldsymbol{\rho}) = \frac{\omega}{1-\phi^2} \geq 0, \quad \mu_\psi(\boldsymbol{\rho}) = 0.$$

As a result,

$$\sqrt{T}\bar{m}_{\delta T}[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)] = -\frac{1-(\phi^0)^2}{\omega^0}\frac{\sqrt{T}}{T}\sum_t x_t x_{t-1} + \phi^0,$$

$$\sqrt{T}\bar{m}_{\psi T}[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)] = \frac{1}{2}\frac{1-(\phi^0)^2}{\omega^0}\frac{\sqrt{T}}{T}\sum_t \left\{\frac{[1-(\phi^0)^2]x_t^2}{\omega^0} - 1\right\}.$$

But since the (scaled) average score of the true log-likelihood is given by

$$\begin{aligned}
\sqrt{T}\bar{q}_{\phi T}(\boldsymbol{\rho}^0) &= \frac{1}{\omega^0}\frac{\sqrt{T}}{T}\sum_t (x_t - \phi^0 x_{t-1})x_{t-1}, \\
\sqrt{T}\bar{q}_{\omega T}(\boldsymbol{\rho}^0) &= \frac{1}{2\omega^0}\frac{\sqrt{T}}{T}\sum_t \left[\frac{(x_t - \phi^0 x_{t-1})^2}{\omega^0} - 1\right],
\end{aligned}$$

which can be written as a (limiting) linear combination of $\bar{m}_T\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right]$, then the result immediately follows from Proposition 7 regardless of the value of $\boldsymbol{\rho}^0$. □

Proposition 8

Part 1: By the usual chain rule for first derivatives,

$$\begin{aligned}\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right] &= \frac{\partial g_1'\left[\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}}\sqrt{T}\frac{\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]}{\partial\pi_1}\\ &+\frac{\partial g_2'\left[\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}}\sqrt{T}\frac{\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]}{\partial\pi_2}+\frac{\partial h'\left[\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)\right]}{\partial\boldsymbol{\theta}}\boldsymbol{\mu}^e(\boldsymbol{\rho}^0).\end{aligned}$$

Then, given our assumptions about $g(\boldsymbol{\theta})$, and the fact that $\boldsymbol{\pi}_2=g_2(\boldsymbol{\theta})=h(\boldsymbol{\theta})$, so that the Lagrange multipliers are unaffected by the reparametrisation, it is clear that $\sqrt{T}\bar{m}_T\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right]$ spans exactly the same linear space as $\sqrt{T}\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\pi_1$ and $\sqrt{T}\left\{\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\pi_2+\boldsymbol{\mu}^e(\boldsymbol{\rho}^0)\right\}$ together. Hence, the equality restricted indirect estimators of $\boldsymbol{\rho}$ based on $\hat{\boldsymbol{\theta}}_T^e$ and $\hat{\boldsymbol{\mu}}_T^e$ simultaneously must be at least as efficient as unrestricted indirect estimator based on $\hat{\boldsymbol{\pi}}_{1T}^u$ alone in view of Proposition 7.

Part 2: Under the stated condition, the asymptotic residual covariance matrix in the limiting least squares projection of $\sqrt{T}\bar{q}_T(\boldsymbol{\rho}^0)$ on $\sqrt{T}\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\pi_1$ will be unaffected by the addition of $\sqrt{T}\{\partial\bar{l}_T\left[\boldsymbol{\pi}_1^u(\boldsymbol{\rho}^0);\mathbf{0}\right]/\partial\pi_2$ $+\boldsymbol{\mu}^e(\boldsymbol{\rho}^0)\}$ as extra "regressors". □

Proposition 9

It follows immediately from Proposition 7. □

Corollary 2

If the equality constraints are satisfied by the unrestricted pseudo-true values of $\boldsymbol{\theta}$, in the sense that $h\left[\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)\right]=\mathbf{0}$, then $\boldsymbol{\theta}^u(\boldsymbol{\rho}^0)=\boldsymbol{\theta}^e(\boldsymbol{\rho}^0)$, $\boldsymbol{\mu}^u(\boldsymbol{\rho}^0)=\boldsymbol{\mu}^e(\boldsymbol{\rho}^0)=\mathbf{0}$, and $m_t\left[\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right]=m_t\left[\boldsymbol{\beta}^u(\boldsymbol{\rho}^0)\right]$ $\forall t$. As a result, $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right]=\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^u(\boldsymbol{\rho}^0)\right]$ for all $\boldsymbol{\rho}$ in an open neighbourhood of $\boldsymbol{\rho}^0$, so that $\partial\mathfrak{m}\left[\boldsymbol{\rho}^0,\boldsymbol{\beta}^e(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\rho}=\partial\mathfrak{m}\left[\boldsymbol{\rho}^0,\boldsymbol{\beta}^u(\boldsymbol{\rho}^0)\right]/\partial\boldsymbol{\rho}$. For analogous reasons, $\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^0,\hat{\boldsymbol{\beta}}_T^e\right]-\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^0,\hat{\boldsymbol{\beta}}_T^u\right]=o_p(1)$ in view of Proposition 3, so that $\mathcal{I}_0^u=\mathcal{I}_0^e$. The required result then follows from Proposition 9. □

Corollary 3

For simplicity of notation, let us define $\mathbf{z}_t=(1,x_{t-1},\ldots,x_{t-k})'$, $\sigma_{xx}(\boldsymbol{\rho})=E\left(x_t^2\middle|\boldsymbol{\rho}\right)$, $\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})=E\left(\mathbf{z}_t x_t\middle|\boldsymbol{\rho}\right)$ 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

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

from where we can obtain the following binding functions

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

and

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

Therefore, we will have that

$$\begin{aligned}\mathfrak{m}_\phi\left[\boldsymbol{\rho},\boldsymbol{\beta}^u(\boldsymbol{\rho}^0)\right] &= \frac{\boldsymbol{\gamma}_{zx}(\boldsymbol{\rho})-\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{\rho}^0)\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho}^0)}{\omega^u(\boldsymbol{\rho}^0)}\\ \mathfrak{m}_\omega\left[\boldsymbol{\rho},\boldsymbol{\beta}^u(\boldsymbol{\rho}^0)\right] &= \frac{\sigma_{xx}(\boldsymbol{\rho})+\boldsymbol{\phi}^u(\boldsymbol{\rho}^0)'\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^u(\boldsymbol{\rho}^0)-2\boldsymbol{\sigma}_{zx}'(\boldsymbol{\rho})\boldsymbol{\phi}^u(\boldsymbol{\rho}^0)-\omega^u(\boldsymbol{\rho}^0)}{2\left[\omega^u(\boldsymbol{\rho}^0)\right]^2}\end{aligned}$$

and

$$
\begin{aligned}
\mathfrak{m}_{\phi}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right] &= \frac{\left[\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})-\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{0})\right]}{\omega^{e}(\boldsymbol{\rho}^{0})}+\omega^{e}(\boldsymbol{\rho}^{0})R'\boldsymbol{\mu}^{e}(\boldsymbol{\rho}^{0}) \\
\mathfrak{m}_{\omega}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right] &= \frac{\sigma_{xx}(\boldsymbol{\rho})+\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{0})'\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{0})-2\boldsymbol{\sigma}_{zx}'(\boldsymbol{\rho})\boldsymbol{\phi}^{e}(\boldsymbol{\rho}^{0})-\omega^{e}(\boldsymbol{\rho}^{0})}{2\left[\omega^{e}(\boldsymbol{\rho}^{0})\right]^{2}}
\end{aligned}
$$

Let us now define the $k+1$ vector of functions $\boldsymbol{\gamma}(\boldsymbol{\rho})=[\gamma_0(\boldsymbol{\rho}),\gamma_1(\boldsymbol{\rho}),\ldots,\gamma_k(\boldsymbol{\rho})]'$, where $\gamma_j(\boldsymbol{\rho})=E\left(x_t x_{t-j}|\,\boldsymbol{\rho}\right)$, and also $\nu(\boldsymbol{\rho})=E\left(x_t|\,\boldsymbol{\rho}\right)$, so that all the elements of $\gamma_{xx}(\boldsymbol{\rho})$, $\boldsymbol{\sigma}_{zx}(\boldsymbol{\rho})$ and $\boldsymbol{\Sigma}_{zz}(\boldsymbol{\rho})$ can be trivially written as functions of $\boldsymbol{\gamma}(\boldsymbol{\rho})$ and $\nu(\boldsymbol{\rho})$. Then, tedious but otherwise straightforward algebra shows that both $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]$ and $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]$ can be written as homeomorphic functions of $\boldsymbol{\gamma}(\boldsymbol{\rho})$ and $\nu(\boldsymbol{\rho})$. As a result, the estimators of $\boldsymbol{\rho}$ based on minimising the optimal norm of $\mathfrak{m}\left[\boldsymbol{\rho},\hat{\boldsymbol{\beta}}_T^{u}\right]$ or $\mathfrak{m}\left[\boldsymbol{\rho},\hat{\boldsymbol{\beta}}_T^{e}\right]$, will be asymptotically equivalent to the CMD estimators based on minimising the optimal norm of $\left[\nu(\boldsymbol{\rho})-\bar{x}_T,\boldsymbol{\gamma}'(\boldsymbol{\rho})-\bar{\boldsymbol{\gamma}}_T'\right]'$, where $\bar{\boldsymbol{\gamma}}_T$ contains the first $k+1$ sample (uncentred) autocovariances of x_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 Corollary 1.

In the first case, we have that $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})=\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})$, so that $m_t\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=m_t\left[\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]\ \forall t$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]$ for all $\boldsymbol{\rho}$ in an open neighbourhood of $\boldsymbol{\rho}^{0}$, so that $\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}=\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}$. In addition, $\sqrt{T}\hat{\boldsymbol{\mu}}_T^{i}=o_p(1)$ and $\sqrt{T}(\hat{\boldsymbol{\theta}}_T^{i}-\hat{\boldsymbol{\theta}}_T^{u})=o_p(1)$ from Propositions 1 and 2 respectively, and $\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{i}\right]-\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{u}\right]=o_p(1)$ in view of Proposition 3, so that $\mathcal{I}_0^{i}=\mathcal{I}_0^{u}$.

In the second case, $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})=\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})$, so that $m_t\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=m_t\left[\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]\ \forall t$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]$ for all $\boldsymbol{\rho}$ in an open neighbourhood of $\boldsymbol{\rho}^{0}$, so that $\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}=\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}$. Similarly, Propositions 1 to 3 also imply that $\sqrt{T}(\hat{\boldsymbol{\mu}}_T^{i}-\hat{\boldsymbol{\mu}}_T^{e})=o_p(1)$, $\sqrt{T}(\hat{\boldsymbol{\theta}}_T^{i}-\hat{\boldsymbol{\theta}}_T^{e})=o_p(1)$, and $\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{i}\right]-\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{e}\right]=o_p(1)$, so that $\mathcal{I}_0^{i}=\mathcal{I}_0^{e}$.

In the last case, of course, $\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})=\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})=\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})$, so that $m_t\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=m_t\left[\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=m_t\left[\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]\ \forall t$. Hence, $\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]=\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]=\mathfrak{m}\left[\boldsymbol{\rho},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]$ for all $\boldsymbol{\rho}$ in an open neighbourhood of $\boldsymbol{\rho}^{0}$, which implies that $\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{i}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}=\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{u}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}=\partial\mathfrak{m}\left[\boldsymbol{\rho}^{0},\boldsymbol{\beta}^{e}(\boldsymbol{\rho}^{0})\right]/\partial\boldsymbol{\rho}$. But in contrast, even if T is large, $\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{i}\right]$ will only coincide with $\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{u}\right]$ approximately half the time, while it will coincide with $\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{e}\right]$ the other half. Nevertheless, since in this case $\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{e}\right]-\sqrt{T}\mathfrak{m}\left[\boldsymbol{\rho}^{0},\hat{\boldsymbol{\beta}}_T^{u}\right]=o_p(1)$ from Corollary 1, all three estimators are asymptotically equivalent. □

Simulation-based estimators

For the clarity of exposition, we have assumed throughout that analytical expressions for the population objective function (5) and its first order conditions (6) can be readily obtained. However, in many 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 approximate the required expectations by means of ensemble averages of the levels and derivatives of the Lagrangian function (1) across H realizations of size T of the true process simulated with parameter values equal to $\boldsymbol{\rho}$. Specifically, if $\{x_t^h(\boldsymbol{\rho}),t=1,\ldots,T\}$ denotes the h^{th} such realization ($h=1,\ldots,H$), then

$$
\begin{aligned}
\mathcal{L}_T(\boldsymbol{\rho};\boldsymbol{\theta}) &= E(\bar{l}_T(\boldsymbol{\theta})|\boldsymbol{\rho})\simeq\frac{1}{H}\sum_{h=1}^{H}\frac{1}{T}\sum_{t=1}^{T}\ln f[x_t^h(\boldsymbol{\rho})|\mathbf{X}_{t-1}^h(\boldsymbol{\rho});\boldsymbol{\theta}]=\bar{\mathcal{L}}_T(\boldsymbol{\rho};\boldsymbol{\theta}), \\
\mathfrak{m}_T(\boldsymbol{\rho};\boldsymbol{\beta}) &= E\left(\left.\frac{\partial\bar{l}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|\boldsymbol{\rho}\right)+\frac{\partial h'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\mu}\simeq\frac{1}{H}\sum_{h=1}^{H}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ln f[x_t^h(\boldsymbol{\rho})|\mathbf{X}_{t-1}^h(\boldsymbol{\rho});\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}}+\frac{\partial h'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\mu}=\bar{\mathfrak{m}}_T(\boldsymbol{\rho};\boldsymbol{\beta}),
\end{aligned}
$$

where we can make the right hand side terms arbitrarily close in a numerical sense to the left hand side ones as $H\to\infty$. 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 x_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 stochastic volatility example in section 3, in which we relied on simulations to compute the required moments, all three conditions were fulfilled.

Since we are assuming that x_t is strictly stationary and ergodic, there is, in fact, an alternative simulation scheme, which replaces the required expectations by their sample analogues in a single but very large realization of the process, $\{x_n(\boldsymbol{\rho}), n = 1, \ldots, T \cdot H\}$. In particular, we will have:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\rho};\boldsymbol{\theta}) &= E(\bar{l}_T(\boldsymbol{\theta})|\boldsymbol{\rho}) \simeq \frac{1}{T\cdot H}\sum_{n=1}^{T\cdot H}\ln f[x_n(\boldsymbol{\rho})|\mathbf{X}_{n-1}(\boldsymbol{\rho});\boldsymbol{\theta}] = \mathcal{L}_{TH}(\boldsymbol{\rho};\boldsymbol{\theta}),\\
\mathfrak{m}(\boldsymbol{\rho};\boldsymbol{\beta}) &= E\left(\left.\frac{\partial\bar{l}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|\boldsymbol{\rho}\right) + \frac{\partial h'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\mu} \simeq \frac{1}{T\cdot H}\sum_{n=1}^{T\cdot H}\frac{\partial\ln f[x_n(\boldsymbol{\rho})|\mathbf{X}_{n-1}(\boldsymbol{\rho});\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}} + \frac{\partial h'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\mu} = \mathfrak{m}_{TH}(\boldsymbol{\rho};\boldsymbol{\beta}).
\end{aligned}$$

In this case, we can again make left and right hand sides arbitrarily close in a numerical sense as $H \to \infty$.

Finally, we can approximate the different binding functions $\boldsymbol{\beta}^r(\boldsymbol{\rho})$ by means of either $\bar{\boldsymbol{\beta}}^r_T(\boldsymbol{\rho})$ or $\boldsymbol{\beta}^r_{TH}(\boldsymbol{\rho})$, which are the appropriately constrained pseudo ML estimators and associated multipliers computed on the basis of $\bar{\mathcal{L}}_T(\boldsymbol{\rho};\boldsymbol{\theta})$ and $\mathcal{L}_{TH}(\boldsymbol{\rho};\boldsymbol{\theta})$, respectively. The main attraction of the first procedure is that it may sometimes improve the small sample properties of the estimators of $\boldsymbol{\rho}$ (see e.g. Gouriéroux, Renault and Touzi (2000) and Arvanitis and Demos (2003)).

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 $\hat{\boldsymbol{\mu}}^u_T = 0$, we can always regard the GMM-based indirect estimation 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 $\left[\partial h'(\hat{\boldsymbol{\theta}}^r_T)/\partial\boldsymbol{\theta}\right] \cdot \hat{\boldsymbol{\mu}}^r_T$ 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 $S_\tau(\boldsymbol{\rho};\boldsymbol{\beta}_T)$ used in the computation of the optimal weighting matrix for the continuously updated GMM-based indirect estimators can also be arbitrarily approximated by replacing the required expected values by their sample counterparts in a long simulation of length $T \cdot H$. In any case, it is important to bear in mind that since H is finite in practice, the asymptotic covariance matrix of the GMM and CMD indirect estimators in Proposition 4 must be multiplied by the scalar quantity $(1 + H^{-1})$ (see GMR).

Table 1

Auxiliary model characteristics:

Pseudo-true values, proportion of auxiliary model parameter estimates at the boundary (Equality/Inequality), and rejection frequencies of normality test

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

κ^2	α^0	δ^0	σ_v^0
1	*-.736*	*.9*	*.363*
$\psi^e(\rho^0)/\psi^i(\rho^0)$	$\varphi^e(\rho^0)/\varphi^i(\rho^0)$	$\pi^e(\rho^0)/\pi^i(\rho^0)$	$\mu_\eta^e(\rho^0)/\eta^i(\rho^0)$
$7.8/7.9\times10^{-3}$	.166/.177	.754/.746	.418/.156

	$T = 500$	$T = 2000$
$\hat{\varphi}_T^r = \varphi_{\min}$	.001/.002	0/0
$\hat{\pi}_T^r = 0$	0/0	0/0
$\hat{\varphi}_T^r + \hat{\pi}_T^r = 1$	.007/.013	0/0
$\hat{\eta}_T = 0$	*1*/0	*1*/0
total	.008/.015	0/0
LM rejections	.977	1

κ^2	α^0	δ^0	σ_v^0
.1	*-.141*	*.98*	*.0614*
$\psi^e(\rho^0)/\psi^i(\rho^0)$	$\varphi^e(\rho^0)/\varphi^i(\rho^0)$	$\pi^e(\rho^0)/\pi^i(\rho^0)$	$\mu_\eta^e(\rho^0)/\eta^i(\rho^0)$
$1.7/1.7\times10^{-3}$	.026/.026	.955/.955	.046/.027

	$T = 500$	$T = 2000$
$\hat{\varphi}_T^r = \varphi_{\min}$	.323/.302	.190/.191
$\hat{\pi}_T^r = 0$	.137/.139	.015/.019
$\hat{\varphi}_T^r + \hat{\pi}_T^r = 1$	.005/.006	.009/.001
$\hat{\eta}_T = 0$	*1*/.248	*1*/.065
total	.400/.577	.203/.256
LM rejections	.206	.504

Figure 1A: Sampling distribution of unrestricted, equality restricted, inequality restricted and pretest indirect estimators of δ

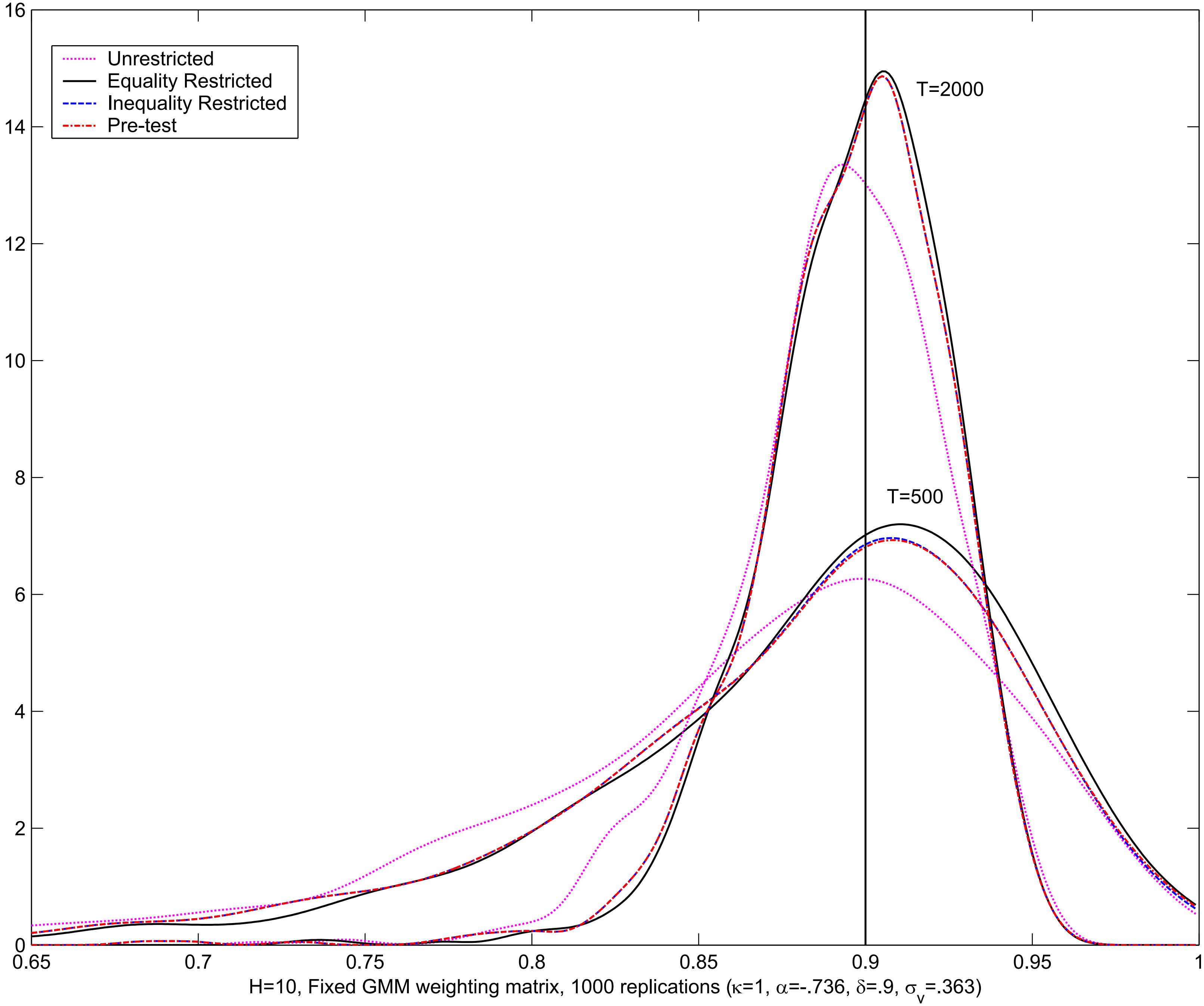


Figure 1B: Sampling distribution of unrestricted, equality restricted, inequality restricted and pretest indirect estimators of σ_v

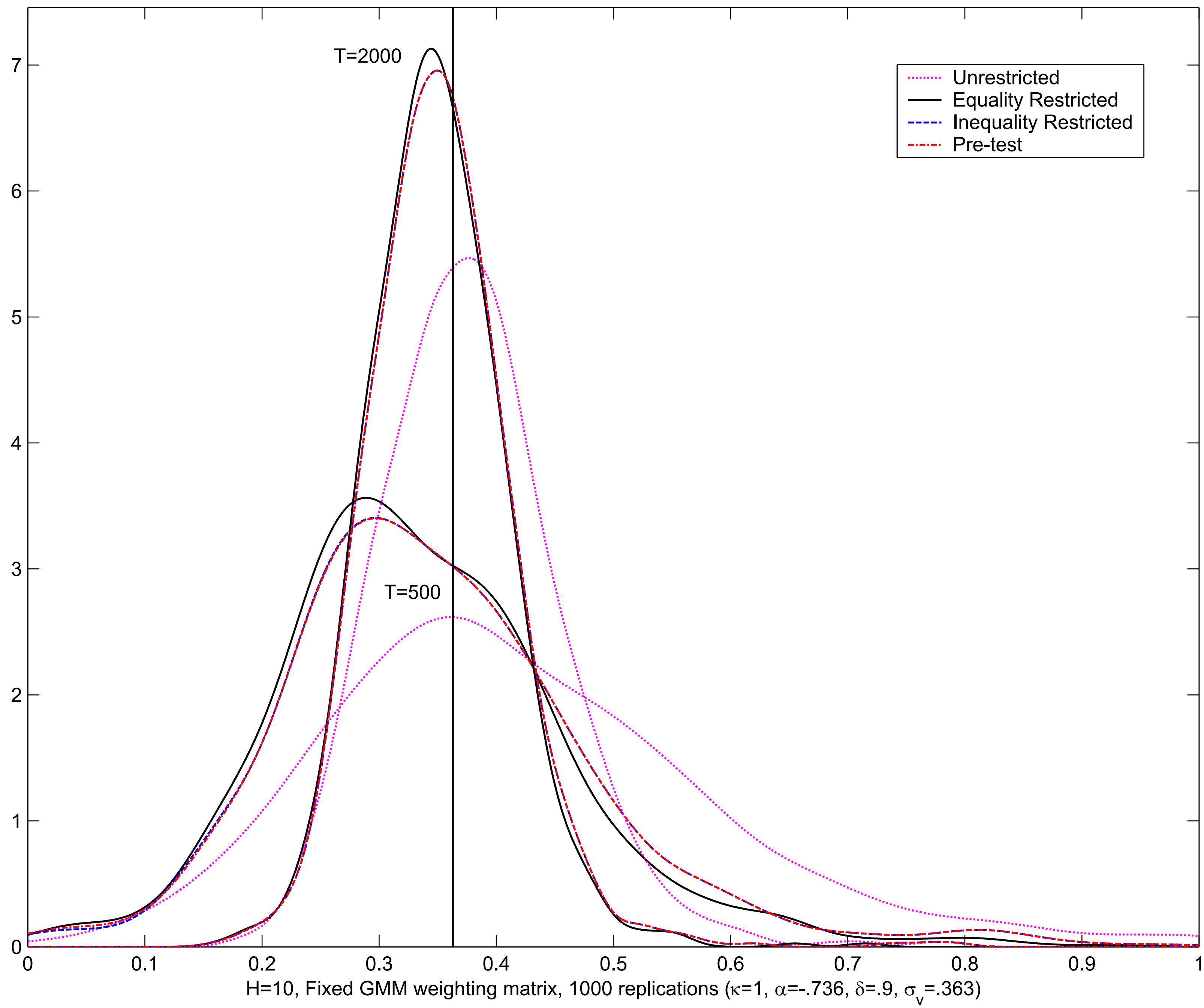
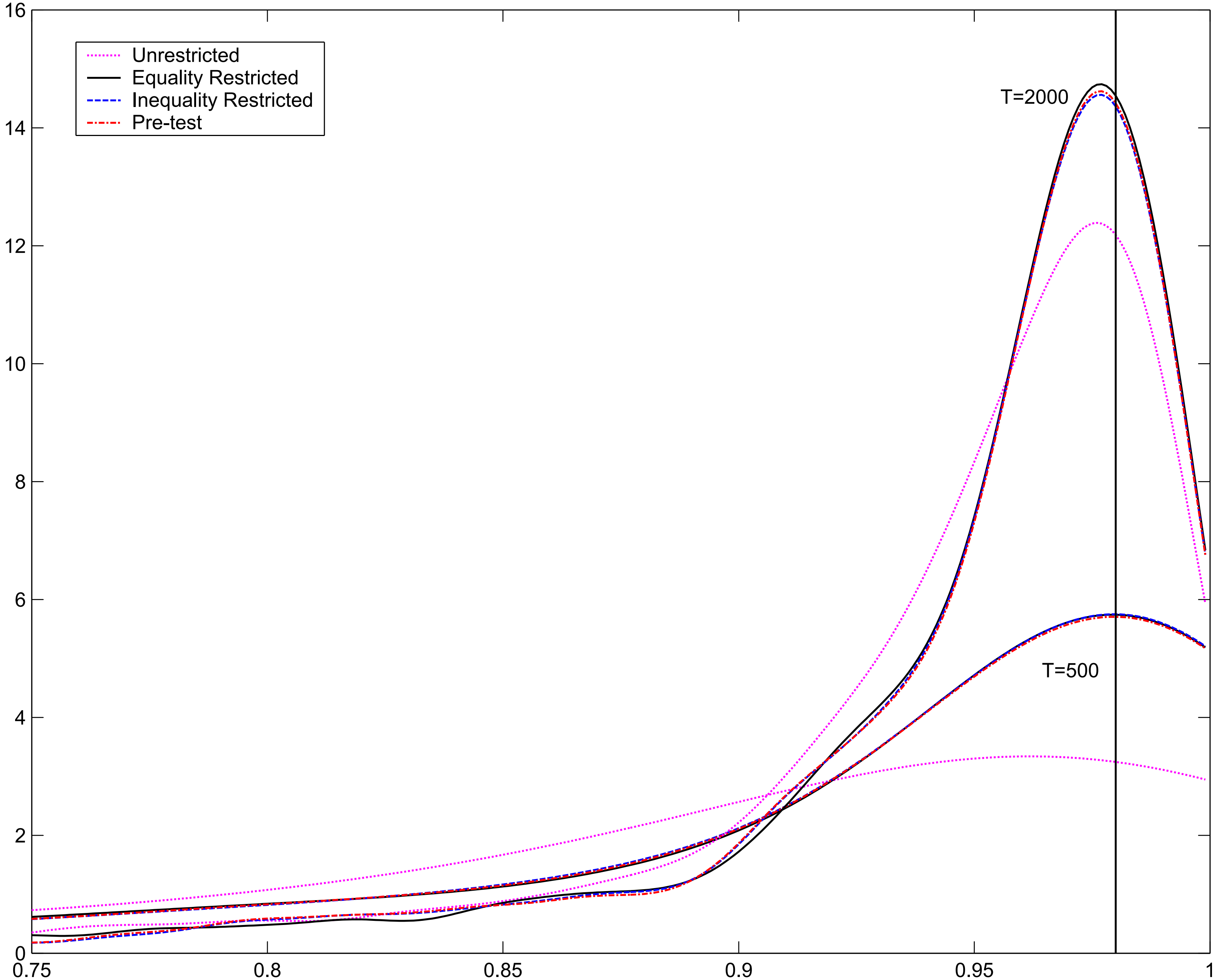
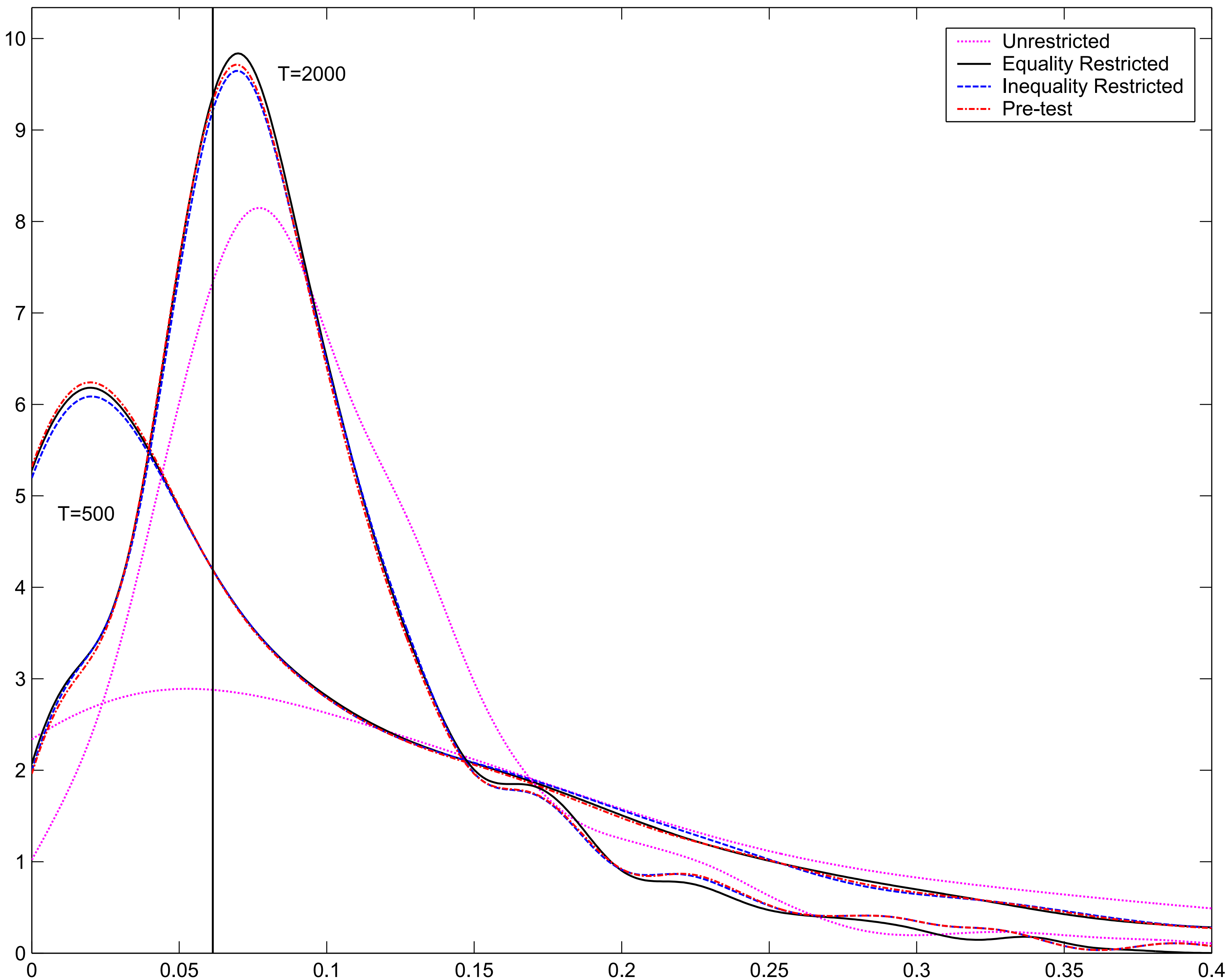


Figure 2A: Sampling distribution of unrestricted, equality restricted, inequality restricted and pretest indirect estimators of δ



H=10, Fixed GMM weighting matrix, 1000 replications (κ=.1, α=-.141, δ=.98, σ_v=.0614)

Figure 2B: Sampling distribution of unrestricted, equality restricted, inequality restricted and pretest indirect estimators of σ_v



H=10, Fixed GMM weighting matrix, 1000 replications (κ=.1, α=-.141, δ=.98, σ_v=.0614)