The relative efficiency of pseudo maximum likelihood estimation and inference in conditionally heteroskedastic dynamic regression models*

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>

February 2003
Revised: June 2006

Incomplete

Abstract

We compare the efficiency of several likelihood-based parametric and semiparametric estimators of conditional mean and variance parameters in multivariate dynamic models with *i.i.d.* spherical innovations, and show that pseudo maximum likelihood estimators are inefficient except under normality. We also provide conditions for partial adaptivity of semiparametric procedures, and relate them to the consistency of maximum likelihood estimators under distributional misspecification. We propose Hausman tests that compare pseudo maximum likelihood estimators with their more efficient but less robust competitors. We also study the efficiency of sequential estimators of the shape parameters. Finally, we provide finite sample results through Monte Carlo simulations.

Keywords: Adaptivity, ARCH, Elliptical Distributions, Financial Returns, Hausman tests, Semiparametric Estimators, Sequential Estimators.

JEL: C13, C14, C12, C51, C52

*We would like to thank Dante Amengual, Manuel Arellano, Nour Meddahi, Javier Mencía, Paolo Zaffaroni, participants at the European Meeting of the Econometric Society (Stockholm, 2003), the Symposium on Economic Analysis (Seville, 2003), the CAF Conference on Multivariate Modelling in Finance and Risk Management (Sandbjerg, 2006), as well as audiences at Bocconi, Cass Business School, CEMFI, CREST, EUI, Roma La Sapienza and Queen Mary for useful comments and suggestions. Of course, the usual caveat applies.

1 Introduction

Many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering effects. Nevertheless, the Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) remain consistent for the conditional mean and variance parameters in those circumstances, so long as those moments are correctly specified.

However, a non-normal distribution may be indispensable when one is interested in features of the distribution of asset returns beyond its conditional mean and variance. For instance, empirical researchers and financial market practitioners are often interested in the so-called Value at Risk of an asset, which is the positive threshold value V such that the probability of the asset suffering a reduction in wealth larger than V equals some pre-specified level $\varkappa < 1/2$. In addition, they are sometimes interested in the probability of the joint occurrence of several extreme events, which is regularly underestimated by the multivariate normal distribution, especially in larger dimensions. This naturally leads one to specify a parametric leptokurtic distribution for the standardised innovations, such as the multivariate student t analysed in Fiorentini, Sentana and Calzolari (2003), and estimate the conditional mean and variance parameters jointly with the parameters characterising the shape of the assumed distribution by maximum likelihood (ML). However, while ML will often yield more efficient estimators of the conditional mean and variance parameters than Gaussian PML if the assumed conditional distribution is correct, it may end up sacrificing consistency when it is not, as shown by Newey and Steigerwald (1997).

If one were mostly interested in the first two conditional moments, the semiparametric (SP) estimators of Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999) would offer an attractive solution because they are sometimes both consistent and partially efficient, as proved by Linton (1993), Drost and Klaassen (1997), Drost, Klaassen and Werker (1997), or Linton and Steigerwald (2000). However, they suffer from the curse of dimensionality, which severely limits their use in multivariate models. To avoid this problem, Hodgson and Vorkink (2001) and Hafner and Rombouts (2006) have recently discussed elliptically symmetric semiparametric (SSP) estimators, which retain univariate rates for their nonparametric part regardless of the cross-sectional dimension of the data, but which are unfortunately more fragile.

One of the main objectives of our paper is to study in detail the trade-offs between efficiency and consistency of the conditional mean and variance parameters that arise in this context. While many of the aforementioned papers provide detailed analyses of some of these issues, especially in univariate models, or in models with no mean, to our knowledge we are the first to simultaneously analyse all the hard choices than an empirical researcher faces in practice.

Furthermore, we do so in a multivariate framework with non-zero means, in which some of the earlier results seem overly simple. Moreover, we explicitly look at the efficiency ranking of the feasible ML procedure that jointly estimates the shape parameters, as well as the infeasible ML, SSP, SP and PML estimators considered in the existing literature. We also provide conditions for partial adaptivity of the SSP and SP procedures, which we relate to the conditions for the consistency of the corresponding parametric ML estimators when the conditional distribution is misspecified. Finally, we propose simple Hausman tests that compare the feasible ML and SSP estimators to the Gaussian PML ones to assess the validity of the distributional assumptions.

But given that practitioners often want to go beyond the first two conditional moments, one cannot simply treat the shape parameters as nuisance parameters. For that reason, we also consider sequential estimators of the shape parameters, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators, and assess their asymptotic efficiency relative to their feasible ML counterpart. In particular, we consider a sequential ML estimator, as well as sequential method of moments (MM) estimators based on higher order moment parameters such as the coefficient of multivariate excess kurtosis.

The rest of the paper is organised as follows. In section 2, we present closed-form expressions for the score vector and the conditional information matrix of a log-likelihood function based on a spherically symmetric assumption for the innovations, and derive the asymptotic distributions of the Gaussian PML estimator and both SP estimators, as well as the sequential estimators of the shape parameters. Then, in section 3 we compare the efficiency of the different estimators of the conditional mean and variance parameters, discuss two specific models of practical interest, and obtain some general results on partial adaptivity. In section 4, we compare the relative efficiency of the different estimators of the shape parameters, while in section 5 we first study the consistency of the conditional mean parameters when the conditional distribution is misspecified, and then introduce the Hausman tests. A Monte Carlo evaluation of the different parameter estimators and testing procedures can be found in section 6. Finally, we present our conclusions in section 7. Proofs and auxiliary results are gathered in the appendix.

2 Theoretical background

2.1 The model

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N dependent variables, $\mathbf{y}_t$, is typically assumed to be generated as:

$$\begin{gathered}
\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}_0)\varepsilon_t^*, \\
\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}), \\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}),
\end{gathered}$$

where $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are N and $N(N+1)/2$-dimensional vectors of functions known up to the $p\times 1$ vector of true parameter values $\boldsymbol{\theta}_0$, $\mathbf{z}_t$ are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and $\mathbf{z}_t$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular $N\times N$ "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t^*$ is a vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0)=\mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0)=\mathbf{I}_N$. As a consequence,

$$\left.\begin{array}{l} E(\mathbf{y}_t|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0)=\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)\\ V(\mathbf{y}_t|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0)=\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\end{array}\right\}. \tag{1}$$

To complete the model, we need to specify the distribution of $\boldsymbol{\varepsilon}_t^*$. We shall initially assume that, conditional on $\mathbf{z}_t$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as some particular member of the spherical family with a well defined density (see the appendix), or $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\eta}_0\sim i.i.d.\ s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ for short, where $\boldsymbol{\eta}$ are some q additional parameters that determine the shape of the distribution of $\varsigma_t=\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. The most prominent example is, of course, the spherical normal distribution, which we assume corresponds to $\boldsymbol{\eta}_0=\mathbf{0}$. For illustrative purposes, though, we shall also look in some detail at the special case in which $\boldsymbol{\varepsilon}_t^*$ follows a standardised multivariate t with ν_0 degrees of freedom, or $i.i.d.\ t(\mathbf{0},\mathbf{I}_N,\nu_0)$ for short. As is well known, the multivariate student t approaches the multivariate normal as $\nu_0\to\infty$, but has generally fatter tails. For that reason, we shall define η as $1/\nu$, which will always remain in the finite range $0\le\eta_0<1/2$ under our assumptions.

2.2 The log-likelihood function, score vector and information matrix

Let $\boldsymbol{\phi}=(\boldsymbol{\theta}',\boldsymbol{\eta})'$ denote the $p+q$ parameters of interest, which we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T based on a particular parametric spherical assumption will take the form $L_T(\boldsymbol{\phi})=\sum_{t=1}^T l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi})=d_t(\boldsymbol{\theta})+c(\boldsymbol{\eta})+g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$, where $d_t(\boldsymbol{\theta})=-1/2\ln|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})=\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})$. Fiorentini, Sentana and Calzolari (2003) provide expressions for $c(\eta)$ and $g\,[\varsigma_t(\boldsymbol{\theta}),\eta]$ in the multivariate student t case, which are obviously such that $L_T(\boldsymbol{\theta},0)$ collapses to a conditionally Gaussian log-likelihood.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. Then, it is straightforward to show that if $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank, and $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ are differentiable

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})=\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+\frac{\partial g_t\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\boldsymbol{\theta}}=[\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}),$$

and

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})=\frac{\partial c(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}+\frac{\partial g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\boldsymbol{\eta}}=\mathbf{e}_{rt}(\boldsymbol{\phi}),$$

where

$$\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) = -2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varsigma}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \tag{2}$$

$$\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) = -vec\left\{\mathbf{I}_N + 2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varsigma}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}, \tag{3}$$

$$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}),$$

$$\mathbf{Z}_{st}(\boldsymbol{\theta}) = \frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\right],$$

and the Jacobian matrices $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}'$ depend on the particular specification adopted.[1]

Given that $-2\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma$ is equal to $(N\eta+1)/[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})]$ in the student t case, and to 1 under Gaussianity, it is straightforward to check that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\eta)$ coincides with the expression in Fiorentini, Sentana and Calzolari (2003), while $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

$$\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) = \begin{bmatrix}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{bmatrix} = \left\{\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right]\end{array}\right\}.$$

As for $\mathbf{e}_{rt}(\boldsymbol{\theta},\mathbf{0})$, Fiorentini, Sentana and Calzolari (2003) show that in the multivariate student t case it is proportional to the second Laguerre polynomial:

$$e_{rt}(\boldsymbol{\theta},0) = \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4}.$$

Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}_{dt}'(\boldsymbol{\phi}),\mathbf{e}_{rt}(\boldsymbol{\phi})]'$ evaluated at $\boldsymbol{\phi}_0$ follows a vector martingale difference, and therefore, the same is true of the score vector $\mathbf{s}_t(\boldsymbol{\phi})$. His results also imply that, under suitable regularity conditions, which in particular require that $\boldsymbol{\phi}_0$ belongs to the interior of the parameter space, the asymptotic distribution of the feasible ML estimator will be $\sqrt{T}(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0)\to N\left[\mathbf{0},\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right]$, where

$$\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0],$$

$$\mathcal{I}_t(\boldsymbol{\phi}) = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right] = V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\eta})\mathbf{Z}_t'(\boldsymbol{\theta}),$$

$$\mathbf{h}_t(\boldsymbol{\phi}) = \begin{pmatrix}\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) & \mathbf{h}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi})\\ \mathbf{h}_{\boldsymbol{\theta}\eta t}'(\boldsymbol{\phi}) & \mathbf{h}_{\eta\eta t}(\boldsymbol{\phi})\end{pmatrix} = \frac{\partial\mathbf{s}_t(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'} = \frac{\partial^2 l_t(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'},$$

$$\mathbf{Z}_t(\boldsymbol{\theta}) = \begin{pmatrix}\mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = \begin{pmatrix}\mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{pmatrix},$$

[1] Note that while both $\mathbf{Z}_t(\boldsymbol{\theta})$ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ depend on the specific choice of square root matrix $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ does not, a property that inherits from $l_t(\boldsymbol{\phi})$. The same result is not generally true for non-elliptical distributions (see Mencía and Sentana (2005)), in which case one should redefine $\mathbf{Z}_{st}(\boldsymbol{\theta})$ as $\{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]$, as in the proofs of Propositions 13 and 17.

and

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{bmatrix} \mathcal{M}_{dd}(\boldsymbol{\eta}) & \mathcal{M}_{dr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{dr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{bmatrix} = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{lr}(\boldsymbol{\eta}) & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}$$

$$= V\left\{ \left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{array} \right] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right\} = V\left\{ \left[\begin{array}{c} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{array} \right] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right\} = V\left[\mathbf{e}_t(\boldsymbol{\phi}) | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right].$$

The following result generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 4.2 in Hafner and Rombouts (2006):

Proposition 1 *If $\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\eta} \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with density function $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$, then*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix},$$

where

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = \mathrm{M}_{ll}(\boldsymbol{\eta}) \mathbf{I}_N,$$
$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = \mathrm{M}_{ss}(\boldsymbol{\eta}) \left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]) vec(\mathbf{I}_N) vec'(\mathbf{I}_N),$$
$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = vec(\mathbf{I}_N) \mathrm{M}_{sr}(\boldsymbol{\eta}_0),$$

$$\mathrm{M}_{ll}(\boldsymbol{\eta}) = E\left\{ \left[\frac{2\partial g(\varsigma_t, \boldsymbol{\eta})}{\partial \varsigma} \right]^2 \frac{\varsigma_t}{N} \middle| \boldsymbol{\eta} \right\},$$

$$\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2} \left\{ 1 + V\left[\frac{\varsigma_t}{N} \frac{2\partial g(\varsigma_t, \boldsymbol{\eta})}{\partial \varsigma} \middle| \boldsymbol{\eta} \right] \right\},$$

$$\mathrm{M}_{sr}(\boldsymbol{\eta}) = -E\left\{ \left[\frac{\varsigma_t}{N} \frac{2\partial g(\varsigma_t, \boldsymbol{\eta})}{\partial \varsigma} + 1 \right] \mathbf{e}'_{rt}(\boldsymbol{\phi}) \middle| \boldsymbol{\eta} \right\},$$

and $\mathbf{K}_{NN}$ is the commutation matrix of orders N, N.

In the multivariate standardised student t case, in particular:

$$\mathrm{M}_{ll}(\eta) = \frac{\nu (N+\nu)}{(\nu-2)(N+\nu+2)}, \ \mathrm{M}_{ss}(\eta) = \frac{(N+\nu)}{(N+\nu+2)}, \ \mathrm{M}_{sr}(\eta) = -\frac{2(N+2)\nu^2}{(\nu-2)(N+\nu)(N+\nu+2)},$$

and

$$\mathcal{M}_{rr}(\eta) = \frac{\nu^4}{4} \left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{N+\nu}{2}\right) \right] - \frac{N\nu^4 \left[\nu^2 + N(\nu-4) - 8\right]}{2(\nu-2)^2 (N+\nu)(N+\nu+2)},$$

where $\psi(x)$ is the di-gamma function (see Abramowitz and Stegun (1964)), which under normality reduce to $\mathrm{M}_{ll}(0) = 1$, $\mathrm{M}_{ss}(0) = 1$, $\mathrm{M}_{sr}(0) = 0$ and $\mathcal{M}_{rr}(0) = N(N+2)/2$. In this sense, it is interesting to note that as N increases, $\mathrm{M}_{ll}(\eta)$, $\mathrm{M}_{ss}(\eta)$ and $\mathrm{M}_{sr}(\eta)$ converge to $\nu/(\nu-2)$, 1 and 0, respectively. This is due to the fact that the multivariate student t can be written as a scale mixture of normals, with a positive mixing variable which can be filtered out with a precision that increases in N (see Mencía and Sentana (2005)). Thus, the marginal log-likelihood function of the observed data will become arbitrarily close to the sum of the conditional log-likelihood of the observed data given the mixing variable, which is multivariate Gaussian and only depends on $\boldsymbol{\theta}$, plus the marginal log-likelihood of the mixing variable, which only depends on η.

2.3 Pseudo maximum likelihood estimators of $\boldsymbol{\theta}$

If the interest of the researcher lied exclusively in $\boldsymbol{\theta}$, which are the parameters characterising the conditional mean and variance functions, then one attractive possibility would be to estimate an equality restricted version of the model in which $\boldsymbol{\eta}$ is set to zero. Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote such a PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. The proof is based on the fact that in those circumstances, the pseudo log-likelihood score, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$, is a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from $\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})$. Importantly, this property is preserved even when the standardised innovations, $\boldsymbol{\varepsilon}_t^*$, are not stochastically independent of $\mathbf{z}_t$ and I_{t-1}. The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result:

Proposition 2 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)]$, where*

$$\begin{gathered}
\mathcal{C}(\boldsymbol{\phi}) = \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}), \\
\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{A}_t(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}), \\
\mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{B}_t(\boldsymbol{\phi}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),
\end{gathered}$$

and

$$\mathcal{K}(\kappa) = V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa \cdot vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}, \quad (4)$$

which only depends on the shape parameters $\boldsymbol{\eta}$ through the population coefficient of multivariate excess kurtosis

$$\kappa = \frac{E(\varsigma_t^2|\boldsymbol{\eta})}{N(N+2)} - 1. \quad (5)$$

But if κ_0 is infinite then $\mathcal{B}(\boldsymbol{\phi}_0)$ will be unbounded, and the asymptotic distribution of some or all the elements of $\tilde{\boldsymbol{\theta}}_T$ will be non-standard, unlike that of $\hat{\boldsymbol{\theta}}_T$ (see Hall and Yao (2003)).

The following result, which specifies the covariance between the Gaussian pseudo score and the true score, will repeatedly prove useful below:

Proposition 3 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $(\mathbf{0}, \mathbf{I}_N)$, then*

$$E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\left[\mathbf{e}'_{dt}(\boldsymbol{\phi}), \mathbf{e}'_{rt}(\boldsymbol{\phi})\right]\middle|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right\} = [\mathcal{K}(0)\,|\mathbf{0}]. \quad (6)$$

Importantly, note that (6) holds regardless of whether or not the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is spherical, provided we interpret $\mathbf{e}_{rt}(\boldsymbol{\phi})$ as the gradient of the relevant shape parameters.

2.4 Sequential estimators of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$

In practice, we will often be interested in features of the distribution of asset returns, such as its quantiles, which go beyond its conditional mean and variance. For that purpose, we can use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of $\boldsymbol{\eta}$ as $\tilde{\boldsymbol{\eta}}_T = \arg\max_{\eta} L_T(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$, possibly subject to some inequality constraints on $\boldsymbol{\eta}$. In the student t case, for instance, $\tilde{\eta}_T$ will be characterised by the first-order Kuhn-Tucker (KT) conditions

$$\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, \tilde{\eta}_T) + \tilde{\lambda}_{\eta T} = 0; \quad \tilde{\eta}_T \geq 0; \quad \tilde{\lambda}_{\eta T} \geq 0; \quad \tilde{\lambda}_{\eta T} \cdot \tilde{\eta}_T = 0,$$

where $\bar{s}_{\eta T}(\boldsymbol{\theta}, \eta)$ is the sample mean of $s_{\eta t}(\boldsymbol{\theta}, \eta)$, and λ_η the KT multiplier associated with the constraint $\eta \geq 0$.

Such a sequential ML estimator of $\boldsymbol{\eta}$ can be given a rather intuitive interpretation. If $\boldsymbol{\theta}_0$ were known, then the squared Euclidean norm of the standardised innovations, $\varsigma_t(\boldsymbol{\theta}_0)$, would be independently and identically distributed over time, with density function $h(\varsigma; \boldsymbol{\eta})$.[2] Therefore, we could obtain the infeasible ML estimator of $\boldsymbol{\eta}$ by maximising with respect to $\boldsymbol{\eta}$ the log-likelihood function of the observed $\varsigma_t(\boldsymbol{\theta}_0)'s$, $\sum_{t=1}^{T} \ln h\left[\varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\eta}\right]$. Although in practice the standardised residuals are usually unobservable, it turns out that $\tilde{\boldsymbol{\eta}}_T$ is the estimator so obtained when we treat $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$ as if they were really observed.

The asymptotic distribution of the sequential ML estimator of $\boldsymbol{\eta}$, which reflects the sample uncertainty in $\tilde{\boldsymbol{\theta}}_T$, is stated in the following result:

Proposition 4 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_0) \to N\left[0, \mathcal{F}(\boldsymbol{\phi}_0)\right]$, where*

$$\mathcal{F}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0) + \mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0).$$

Importantly, since $\mathcal{C}(\boldsymbol{\phi}_0)$ will become unbounded as $\kappa_0 \to \infty$, the asymptotic distribution of $\tilde{\boldsymbol{\eta}}_T$ will also be non-standard in that case, unlike that of the feasible ML estimator $\hat{\boldsymbol{\eta}}_T$.

If we can obtain closed-form expressions for at least q functions of ς_t, $\boldsymbol{\upsilon}(.)$ say, then we can also compute a sequential method of moments (MM) estimator of $\boldsymbol{\eta}$, $\check{\boldsymbol{\eta}}_T(\boldsymbol{\Omega})$ say, by minimising with respect to $\boldsymbol{\eta}$ the quadratic form $\bar{\mathbf{n}}_{\boldsymbol{\eta}T}'(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})\boldsymbol{\Omega}\bar{\mathbf{n}}_{\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$, where $\boldsymbol{\Omega}$ is a positive definite weighting matrix, and $\mathbf{n}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \boldsymbol{\upsilon}[\varsigma_t(\boldsymbol{\theta})] - E\{\boldsymbol{\upsilon}[\varsigma_t(\boldsymbol{\theta})]|\boldsymbol{\eta}\}$. Given that $E[\varsigma_t(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{\eta}] = N$, the most obvious moment to use is (5), which suffices to identify η in the multivariate student t case through the theoretical relationship $\kappa = 2/(\nu - 4)$ (see Fiorentini, Sentana and Calzolari (2003)). In this context, if we define the influence function

$$n_{\eta t}(\boldsymbol{\theta}, \eta) = \frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)} - \frac{1 - 2\eta}{1 - 4\eta},$$

[2] For instance, when $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu)$, the distribution of ς_t will be that of either an F variate with N and ν_0 degrees of freedom multiplied by $N(\nu_0 - 2)/\nu_0$ if $\nu_0 < \infty$, or a chi-square random variable with N degrees of freedom under Gaussianity (see e.g. Lemma 1 in Fiorentini, Sentana and Calzolari (2003)).

we obtain

$$\breve{\eta}_T = \frac{\max\left[0, \bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)\right]}{4\max\left[0, \bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)\right]+2},$$

where

$$\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T) = \frac{T^{-1}\sum_{t=1}^T \varsigma_t^2(\tilde{\boldsymbol{\theta}}_T)}{N(N+2)} - 1$$

is Mardia's (1970) sample coefficient of multivariate excess kurtosis of the estimated standardised residuals. We can obtain a closely related estimator, $\mathring{\eta}_T$ say, from the modified influence function

$$\mathring{n}_{\eta t}(\boldsymbol{\theta}, \eta) = \frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)} - \frac{2(1-2\eta)\varsigma_t(\boldsymbol{\theta})}{N(1-6\eta)} + \frac{(1-2\eta)^2}{(1-4\eta)(1-6\eta)},$$

which is the relevant second-order orthogonal polynomial when ς_t is proportional to an $F_{N,\nu}$ random variable. The asymptotic distributions of these two sequential MM estimators of η are stated in the following result:

Proposition 5 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ is i.i.d. $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$, with $\nu_0 > 8$, then under the regularity conditions A.1 in Bollerslev and Wooldridge (1992) we have that $\sqrt{T}(\breve{\eta}_T - \eta_0) \to N[0, \mathcal{G}(\boldsymbol{\phi}_0)]$ and $\sqrt{T}(\mathring{\eta}_T - \eta_0) \to N[0, \mathcal{H}(\boldsymbol{\phi}_0)]$, where*

$$\begin{aligned}
\mathcal{G}(\boldsymbol{\phi}_0) &= [\mathcal{E}(\boldsymbol{\phi}_0) + \mathcal{R}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{R}(\boldsymbol{\phi}_0) - 2\mathcal{R}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}(\boldsymbol{\phi}_0)]/\mathcal{N}^2(\boldsymbol{\phi}_0), \\
\mathcal{H}(\boldsymbol{\phi}_0) &= [\mathcal{L}(\boldsymbol{\phi}_0) + \mathcal{Q}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{Q}(\boldsymbol{\phi}_0)]/\mathcal{N}^2(\boldsymbol{\phi}_0),
\end{aligned}$$

$$\begin{aligned}
\mathcal{D}(\boldsymbol{\phi}_0) &= cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0), n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \frac{4(\nu_0-2)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}\mathbf{W}_s(\boldsymbol{\phi}_0), \\
\mathcal{E}(\boldsymbol{\phi}_0) &= V[n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \frac{(\nu_0-2)^2}{(\nu_0-4)^4}\left[\frac{(N+6)(N+4)}{N(N+2)}\frac{(\nu_0-2)(\nu_0-4)}{(\nu_0-6)(\nu_0-8)} - 1\right], \\
\mathcal{L}(\boldsymbol{\phi}_0) &= V[\mathring{n}_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \mathcal{E}(\boldsymbol{\phi}_0) - \frac{8(\nu_0-2)^2(N+\nu_0-2)}{N(\nu_0-6)^2(\nu_0-4)}, \\
\mathcal{R}(\boldsymbol{\phi}_0) &= cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \eta_0), n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \frac{4(\nu_0-2)}{N(\nu_0-4)}\mathbf{W}_s(\boldsymbol{\phi}_0), \\
\mathcal{Q}(\boldsymbol{\phi}_0) &= cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \eta_0), \mathring{n}_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = -\frac{8(\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}\mathbf{W}_s(\boldsymbol{\phi}_0), \\
\mathcal{N}(\boldsymbol{\phi}_0) &= cov[s_{\eta t}(\boldsymbol{\theta}_0, \eta_0), n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \frac{2\nu_0^2}{(\nu_0-4)^2},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{W}_s(\boldsymbol{\phi}_0) &= \mathbf{Z}_d(\boldsymbol{\phi}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]' \\
&= E\left\{\left.\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)]\right| \boldsymbol{\phi}_0\right\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0].
\end{aligned}$$

Note that since both $\mathcal{G}(\boldsymbol{\phi}_0)$ and $\mathcal{H}(\boldsymbol{\phi}_0)$ will diverge to infinity as ν_0 converges to 8 from above, $\breve{\eta}_T$ and $\mathring{\eta}_T$ will not be root-T consistent for $4 \le \nu_0 \le 8$. Moreover, since κ is infinite for $2 < \nu_0 \le 4$, $\breve{\eta}_T$ and $\mathring{\eta}_T$ will not even be consistent in the interior of this range.

More generally, we could consider the higher order moment parameters of spherical random variables introduced by Berkane and Bentler (1986), $\tau_k(\boldsymbol{\eta})$, which Maruyama and Seo (2003) relate to the higher order moments of ς_t as $E(\varsigma_t^k|\boldsymbol{\eta}) = [\tau_k(\boldsymbol{\eta})+1]E(\varsigma_t^k|\mathbf{0})$, where

$$E(\varsigma_t^k|\mathbf{0}) = 2^k(N/2)(1+N/2)\cdots(k-2+N/2)(k-1+N/2),$$

whence we can also obtain the higher-order orthogonal polynomials of ς_t.[3] By using these additional moments, we can in principle improve the efficiency of the sequential MM estimators, although the precision with which we can estimate $\tau_k(\boldsymbol{\eta})$ rapidly decreases with k (see Newey and Powell (1998) for a characterisation of efficient sequential estimators).

Finally, if we were to iterate the sequential ML procedure, and achieved convergence, then we would obtain fully efficient ML estimators of all model parameters. In fact, a single scoring iteration without line searches that started from $\tilde{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\eta}}_T$ (or any other root-T consistent estimators) would suffice to yield an estimator of $\boldsymbol{\phi}$ that would be asymptotically equivalent to the full-information ML estimator $\hat{\boldsymbol{\phi}}_T$, at least up to terms of order $O_p(T^{-1/2})$. Specifically,

$$\begin{pmatrix} \ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T \\ \ddot{\boldsymbol{\eta}}_T - \tilde{\boldsymbol{\eta}}_T \end{pmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0) \\ \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0) \end{bmatrix}^{-1} \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T) \\ \mathbf{s}_{\eta t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T) \end{bmatrix}.$$

If we use the partitioned inverse formula, then it is easy to see that

$$\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T = \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\right]^{-1}$$
$$\times\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T) - \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T)\right] = \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T),$$

where

$$\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) = \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\right]^{-1},$$

and

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}|\eta t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) &= \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\eta t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \\ &= \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}^{-1}_{rr}(\boldsymbol{\eta}_0)\mathbf{e}_{rt}(\boldsymbol{\phi}_0)\right] \end{aligned} \tag{7}$$

is the residual from the unconditional theoretical regression of the score corresponding to $\boldsymbol{\theta}$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$, on the score corresponding to $\boldsymbol{\eta}$, $\mathbf{s}_{\eta t}(\boldsymbol{\phi}_0)$. The residual score $\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ is sometimes called the parametric efficient score of $\boldsymbol{\theta}$, and its variance,

$$\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)$$
$$= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}^{-1}_{rr}(\boldsymbol{\eta}_0)\mathrm{M}'_{sr}(\boldsymbol{\eta}_0)\right],$$

[3] In the standardised multivariate student t, for instance,
$\tau_k(\boldsymbol{\eta})+1 = (1-2\eta)^{k-1}/\{(1-2k\eta)[1-2(k-1)\eta]\cdots(1-4\eta)\}$ for $2\le k<\nu/2$.

the marginal information matrix of $\boldsymbol{\theta}$, or the feasible parametric efficiency bound. In this respect, note that $\mathcal{I}^{\theta\theta}(\boldsymbol{\phi}_0)$, which is the inverse of $\mathcal{P}(\boldsymbol{\phi}_0)$, coincides with the first block of $\mathcal{I}^{-1}(\boldsymbol{\phi}_0)$, and therefore it gives us the asymptotic variance of the feasible ML estimator, $\hat{\boldsymbol{\theta}}_T$.

2.5 Semiparametric estimators of $\boldsymbol{\theta}$

It is worth noting that the last summand of (7) coincides with $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on (the linear span of) $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ from Proposition 3. Such an interpretation immediately suggests alternative estimators of $\boldsymbol{\theta}$ that replace our parametric assumption on the shape of the distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ by nonparametric or semiparametric alternatives. In this section, we shall consider two such estimators.

The first one is fully nonparametric, and therefore replaces the linear span of $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ by the so-called unrestricted tangent set, which is the Hilbert space generated by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$. The following proposition, which generalises the univariate results of Gonzalez-Rivera and Drost (1999) and Propositions 2, 3 and 4.2 in Hafner and Rombouts (2006) to multivariate models in which the conditional mean vector is not identically zero, describes the resulting semiparametric efficient score and the corresponding efficiency bound:

Proposition 6 *When $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, the semiparametric efficient score is given by the following expression:*

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\phi}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathcal{K}(0)\,\mathcal{K}^+(\kappa_0)\,\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})\right], \tag{8}$$

where $+$ denotes Moore-Penrose inverses, while the semiparametric efficiency bound is

$$\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\phi}_0)\left[\mathcal{M}_{dd}(\boldsymbol{\eta}_0) - \mathcal{K}(0)\,\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}_0). \tag{9}$$

In practice, however, $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced by a nonparametric estimator obtained from the density of the standardised innovations $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, which suffers from the curse of dimensionality.

For this reason, Hodgson and Vorkink (2001), Hafner and Rombouts (2006) and other authors have suggested to limit the admissible distributions to the class of spherically symmetric ones. As a consequence, the restricted tangent set in this case becomes the Hilbert space generated by all time-invariant functions of $\varsigma_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$. The following proposition, which corrects and extends Proposition 7 in Hafner and Rombouts (2006), provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 7 *When $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}, \phi_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, the elliptically symmetric semiparametric efficient score is given by the following expression:*

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left[-\left\{\frac{2\partial g(\varsigma_t, \boldsymbol{\eta}_0)}{\partial \varsigma}\frac{\varsigma_t}{N} + 1\right\} - \frac{2}{(N+2)\kappa_0 + 2}\left(\frac{\varsigma_t}{N} - 1\right)\right], \tag{10}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]}\right\}. \tag{11}$$

Once again, $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced in practice by a semiparametric estimate obtained from the joint density of ε_t^*. However, the elliptical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h(\varsigma_t; \boldsymbol{\eta})$, avoiding in this way the curse of dimensionality.

3 The relative efficiency of the different estimators of $\boldsymbol{\theta}$

3.1 General ranking and full efficiency conditions

In the previous section we have effectively considered five different estimators of $\boldsymbol{\theta}$: (1) the infeasible ML estimator, whose computation requires knowledge of $\boldsymbol{\eta}_0$; (2) the feasible ML estimator, which simultaneously estimates $\boldsymbol{\eta}$; (3) the elliptically symmetric semiparametric estimator, which restricts ε_t^* to have an *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ conditional distribution, but does not impose any additional structure on the distribution of ς_t; (4) the unrestricted semiparametric estimator, which only assumes that the conditional distribution of ε_t^* is *i.i.d.*$(\mathbf{0}, \mathbf{I}_N)$; and (5) the PML estimator, which imposes $\boldsymbol{\eta} = \mathbf{0}$ even though the true conditional distribution of ε_t^* may not be Gaussian. The following proposition ranks the inverses of the asymptotic variances of those five estimators:

Proposition 8 *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \phi_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, then*

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) \geq \mathcal{P}(\boldsymbol{\phi}_0) \geq \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) \geq \mathcal{S}(\boldsymbol{\phi}_0) \geq \mathcal{C}^{-1}(\boldsymbol{\phi}_0).$$

In general, the above matrix inequalities are strict, at least in part. However, there is one instance in which all the above inequalities become equalities: when the true conditional distribution is Gaussian. In that case, the PML estimator is obviously fully efficient, which implies that all the other estimators of $\boldsymbol{\theta}$ must also be efficient. Moreover, normality is the only one such instance within the spherical family:

Proposition 9 *1. If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \phi_0$ is i.i.d. $N(\mathbf{0}, \mathbf{I}_N)$, then*

$$\mathcal{I}_t(\boldsymbol{\theta}_0, \mathbf{0}) = V\left[\mathbf{s}_t(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \mathbf{0}\right] = \begin{bmatrix} V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \mathbf{0}\right] & \mathbf{0} \\ \mathbf{0}' & \mathcal{M}_{rr}(\mathbf{0}) \end{bmatrix}$$

where

$$V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \mathbf{0}\right] = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \mathbf{0}\right] = \mathcal{A}_t(\boldsymbol{\theta}_0, \mathbf{0}) = \mathcal{B}_t(\boldsymbol{\theta}_0, \mathbf{0}).$$

2. *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $-2/(N+2) < \kappa_0 < \infty$, and $\mathbf{W}_s(\boldsymbol{\phi}_0) \neq \mathbf{0}$, then $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\varsigma_t|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. Gamma with mean N and variance $N[(N+2)\kappa_0 + 2]$.*

3. *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and $\mathbf{Z}_l(\boldsymbol{\phi}_0) \neq \mathbf{0}$, then $\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\boldsymbol{\eta}_0 = \mathbf{0}$.*

The first part of this proposition, which generalises Proposition 2 in Fiorentini, Sentana and Calzolari (2003), implies that as far as $\boldsymbol{\theta}$ is concerned, there is no asymptotic efficiency loss in estimating $\boldsymbol{\eta}$ when $\boldsymbol{\eta}_0 = \mathbf{0}$.[4] The second part generalises the results in Gonzalez-Rivera (1997), and it implies that the SSP estimator can be fully efficient only if ε_t^* has a conditional Kotz distribution (see Kotz (1975)), which is a sufficient but not necessary condition for $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = \mathbf{0}$, which in turn implies $\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$. Finally, the last part of Proposition 9 generalises Result 2 in Drost and Gonzalez-Rivera (1999) and Proposition 6 in Haffner and Rombouts (2006).

Unfortunately, it is virtually impossible to obtain closed-form expressions for the different efficiency bounds in conditionally heteroskedastic dynamic non-Gaussian models, as one has to resort to Monte Carlo integration methods to compute the expected values of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ or $\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}_{dt}'(\boldsymbol{\theta})$ (see e.g. Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999)). In the next subsection, though, we shall obtain closed-form expressions in two situations of practical interest.

3.2 Examples

3.2.1 Univariate conditionally heteroskedastic autoregressive models

Consider the following univariate, covariance stationary AR(h)-ARCH(q) model:

$$\left.\begin{array}{c} y_t = \mu_t(\pi_0, \boldsymbol{\rho}_0) + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \mu_t(\pi, \boldsymbol{\rho}) = \pi(1 - \sum_{j=1}^h \rho_j) + \sum_{j=1}^h \rho_j y_{t-j}, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega(1 - \sum_{j=1}^q \alpha_j) + \sum_{j=1}^q \alpha_j [y_{t-j} - \mu_{t-j}(\pi, \boldsymbol{\rho})]^2, \\ \varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ s(0, 1, \boldsymbol{\eta}_0). \end{array}\right\} \tag{12}$$

Define $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_h)'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$, so that $\boldsymbol{\theta} = (\pi, \boldsymbol{\rho}', \omega, \boldsymbol{\alpha}')'$. We can establish the following result:

Proposition 10 *If in model (12) $\boldsymbol{\alpha}_0 = \mathbf{0}$, and all the roots of $1 - \sum_{j=1}^h \rho_{j0} L^j = 0$ are outside the unit circle, then the feasible ML estimators of π, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ are as efficient as the infeasible ML estimators, which require knowledge of $\boldsymbol{\eta}_0$. If in addition $\kappa_0 < \infty$, then the elliptically symmetric semiparametric estimators of π, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ are also fully efficient. The same is true of the semiparametric estimators of $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$, but not of π. In contrast, the inefficiency ratio of the Gaussian PML estimators is $\mathrm{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$ for π and $\boldsymbol{\rho}$, and $4/\{[3\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1](3\kappa_0 + 2)\}$ for $\boldsymbol{\alpha}$.*

[4] In the multivariate student t case, in fact, the feasible ML estimator of $\boldsymbol{\theta}$ will be numerically identical to the PML estimator approximately half the time in large samples because $\eta = 0$ lies at the boundary of the admissible parameter space (see e.g. Andrews (1999)).

Not surprisingly, we can also show that these inefficiency ratios coincide with the ratios of the non-centrality parameters of the corresponding tests of conditional homoskedasticity against local alternatives of the form $\boldsymbol{\alpha}_{0T} = \boldsymbol{\alpha}_0/\sqrt{T}$ in model (12) (see Linton and Steigerwald (2000)).

3.2.2 Multivariate conditionally heteroskedastic autoregressive models

Consider a single factor version of the conditionally heteroskedastic factor model in Sentana and Fiorentini (2001) augmented with covariance stationary diagonal VAR(1) dynamics:

$$\left.\begin{array}{c}\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\pi}_0, \boldsymbol{\rho}_0) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*, \\ \mu_{it}(\pi_i, \rho_i) = \pi_i(1-\rho_i) + \rho_i y_{it-1}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{cc}'\lambda_t(\boldsymbol{\theta}) + \boldsymbol{\Gamma}, \\ \lambda_t(\boldsymbol{\theta}) = 1 + \sum_{j=1}^q \alpha_j[f_{kt-j}^2(\boldsymbol{\theta}) + \omega_{t-j}(\boldsymbol{\theta}) - 1], \\ \varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0), \end{array}\right\} \tag{13}$$

where $f_{kt}(\boldsymbol{\theta})$ is the conditionally linear Kalman filter estimator of the underlying common factor, and $\omega_t(\boldsymbol{\theta})$ the corresponding conditional mean square error (see Sentana (2004) for details). Define $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)'$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_N)'$, $\boldsymbol{\gamma} = vecd(\boldsymbol{\Gamma})$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$, so that $\boldsymbol{\theta} = (\boldsymbol{\pi}', \boldsymbol{\rho}', \mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$. We can establish the following result:

Proposition 11 *If in model (13) $\boldsymbol{\alpha}_0 = 0$, $\gamma_{i0} > 0\ \forall i$, and $|\rho_{i0}| < 1\ \forall i$, then the feasible ML estimators of $\boldsymbol{\pi}$, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ are as efficient as the infeasible ML estimators, which require $\boldsymbol{\eta}_0$ to be known. If in addition $\kappa_0 < \infty$, then the elliptically symmetric semiparametric estimators of $\boldsymbol{\pi}$, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ are also fully efficient. The same is also true of the semiparametric estimators of $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$, but not of $\boldsymbol{\pi}$. In contrast, the inefficiency ratio of the Gaussian PML estimators is $\mathrm{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$ for $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$, and $4/\{[3\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1](3\kappa_0 + 2)\}$ for $\boldsymbol{\alpha}$.*

These inefficiency ratios coincide with the corresponding ratios in the univariate example of Proposition 10. In the multivariate student t case with $\nu_0 > 4$, in particular, they become $(\nu_0-2)(\nu_0+N+2)/[\nu_0(\nu_0+N)]$ and $(\nu_0+N+2)\,(\nu_0-4)/[(\nu_0-1)(\nu_0+N-1)]$, respectively. For any given N, these ratios are monotonically increasing in ν_0, and approach 1 from below as $\nu_0 \to \infty$ in accordance to Proposition 9, and 0 from above as $\nu_0 \to 2^+$ or $\nu_0 \to 4^+$. For instance, for $N = 1$ and $\nu_0 = 9$, they take the value of $.9\mathring{3}$ and $.8\mathring{3}$, respectively, while for $\nu_0 = 5$, their values are only .8 and .4. At the same time, these ratios are decreasing in N for a given ν_0, which reflects the fact that the information matrix is "increasing" in N, as discussed after Proposition 1. For $\nu_0 = 9$ and $N = 3$, for instance, they take the value of .907 and .795, respectively, while for $\nu_0 = 5$, their values are only .75 and .357.

Furthermore, we can also show that these inefficiency ratios coincide with the ratios of the non-centrality parameters of the corresponding tests of conditional homoskedasticity against local alternatives of the form $\boldsymbol{\alpha}_{0T} = \boldsymbol{\alpha}_0/\sqrt{T}$ in model (13) (see Sentana and Fiorentini (2001)).

3.3 Some general results on partial adaptivity

In the previous subsection we have studied two situations in which some, but not all elements of $\boldsymbol{\theta}$ can be estimated as efficiently as if $\boldsymbol{\eta}_0$ were known (see also Lange, Little and Taylor (1989)), a fact that would be described in the semiparametric literature as partial adaptivity. Effectively, this requires that some elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ be orthogonal to the relevant tangent set after partiallying out the effects of the remaining elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ by regressing the former on the latter. Partial adaptivity, though, often depends on the model parametrisation. The following proposition, which generalises earlier results by Bickel (1982), Linton (1993), Drost, Klaassen and Werker (1997) and Hodgson and Vorkink (2003), introduces a general sufficient condition for partial adaptivity of the elliptically symmetric semiparametric estimators in multivariate dynamic models:

Proposition 12 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$, and we can find a homeomorphic transformation $\mathbf{r}_s(.) = [\mathbf{r}'_{1s}(.), r'_{2s}(.)]'$ of the conditional mean and variance parameters $\boldsymbol{\theta}$ into an alternative set of parameters $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_1, \vartheta'_2)'$, where ϑ_2 is a scalar, and $\mathbf{r}_s(\boldsymbol{\theta})$ is twice continuously differentiable with rank$[\partial\mathbf{r}'_s(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that*

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1) \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_2\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1) \end{array}\right\} \tag{14}$$

for all t, then the elliptically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_1$ is ϑ_2-adaptive.

In view of Proposition 8, the feasible ML estimator of $\boldsymbol{\vartheta}_1$ will also be ϑ_2-adaptive when the parametric conditional distribution of $\boldsymbol{\varepsilon}_t^*$ assumed for estimation purposes is correct.

In principle, it may seem that the two examples discussed in the previous sections cannot be rationalised in terms of this proposition because their parametrisations do not satisfy condition (14). In particular, the ARCH parameters $\boldsymbol{\alpha}$ are not generally scale-invariant. However, as explained by Linton and Steigerwald (2000) in the context of model (12), condition (14) will be effectively satisfied under the maintained hypothesis of $\boldsymbol{\alpha}_0 = \mathbf{0}$.

It is also possible to find an analogous result for the unrestricted semiparametric estimator, but at the cost of restricting further the set of parameters that can be estimated in a partially adaptive manner. The following proposition, which does not require sphericity, generalises Theorem 3.1 in Drost and Klaassen (1997):

Proposition 13 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}$ is i.i.d.$(\mathbf{0}, \mathbf{I}_N)$, and we can find a homeomorphic transformation $\mathbf{r}_g(.) = [\mathbf{r}'_{1g}(.), \mathbf{r}'_{2g}(.), \mathbf{r}'_{3g}(.)]'$ of the conditional mean and variance parameters $\boldsymbol{\theta}$ into an alternative set of parameters $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, \boldsymbol{\delta}'_3)'$, where $\boldsymbol{\delta}_2 = vech(\boldsymbol{\Delta}_2)$, $\boldsymbol{\Delta}_2$ is an unrestricted positive (semi)definite matrix of order N, $\boldsymbol{\delta}_3$ is $N \times 1$, and $\mathbf{r}_g(\boldsymbol{\theta})$ is twice continuously differentiable with rank$[\partial\mathbf{r}'_g(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that*

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t^*(\boldsymbol{\delta}_1) + \boldsymbol{\Sigma}_t^{*1/2}(\boldsymbol{\delta}_1)\boldsymbol{\delta}_3 \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{*1/2}(\boldsymbol{\delta}_1)\boldsymbol{\Delta}_2\boldsymbol{\Sigma}_t^{*1/2\prime}(\boldsymbol{\delta}_1) \end{array}\right\} \tag{15}$$

for all t, then the semiparametric estimator of $\boldsymbol{\delta}_1$ is $(\boldsymbol{\delta}_2, \boldsymbol{\delta}_3)$-adaptive.

Such a reparametrisation implicitly requires that $\boldsymbol{\Sigma}_t^{*-1/2}(\boldsymbol{\delta}_1)[\mathbf{y}_t - \boldsymbol{\mu}_t^*(\boldsymbol{\delta}_1)] = \boldsymbol{\delta}_3 + \boldsymbol{\Delta}_2^{1/2}\boldsymbol{\varepsilon}_t^*$ is *i.i.d. elliptical*$(\boldsymbol{\delta}_3, \boldsymbol{\Delta}_2)$. Unfortunately, the constant conditional correlation model of Bollerslev (1990), which assumes that $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{D}_t(\boldsymbol{\theta}_1)\mathbf{R}\mathbf{D}_t(\boldsymbol{\theta}_1)$, where $\mathbf{D}_t$ is a positive diagonal matrix, $\boldsymbol{\theta}_2 = vecl(\mathbf{R})$ and $\mathbf{R}$ a correlation matrix, seems to be the only multivariate GARCH specification proposed so far that can be parametrised in this way as long as we additionally assume that $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mathbf{0}\ \forall t$, in which case $\boldsymbol{\delta}_3$ is unnecessary. And even in that case, we could only adaptively estimate the parameters of $\boldsymbol{\Sigma}_t^{*1/2}(\boldsymbol{\delta}_1) = \mathbf{D}_t(\boldsymbol{\theta}_1)\{E[\mathbf{D}_t(\boldsymbol{\theta}_1)]\}^{-1}$, which will typically correspond to the relative scale parameters of the N univariate ARCH models for the elements of $\mathbf{y}_t$. In most other models, we may need to artificially augment the original parametrisation with $\boldsymbol{\delta}_2$ and $\boldsymbol{\delta}_3$ even though we know that $\boldsymbol{\delta}_{20} = vech(\mathbf{I}_N)$ and $\boldsymbol{\delta}_{30} = \mathbf{0}$, which could be associated with a substantial efficiency cost. Furthermore, in doing so, we must guarantee that the parameters $\boldsymbol{\delta}_1$ remain identified (see Newey and Steigerwald (1997) for a detailed discussion of these issues in univariate models). In this sense, the main difference between Propositions 12 and 13 is that in the elliptically symmetric case we can restrict $\boldsymbol{\Delta}_2$ to be a scalar matrix, and $\boldsymbol{\delta}_3$ to $\mathbf{0}$ regardless of the mean specification, which reduces the number of parameters by a factor of $N(N+3)/2$.

4 The relative efficiency of ML and sequential estimators of $\boldsymbol{\eta}$

The asymptotic variance of the feasible ML estimator of $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}_T$, is

$$\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) = \left[\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\right]^{-1},$$

which coincides with the inverse of the variance of the efficient parametric score of $\boldsymbol{\eta}$, $\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$, which is the residual in the theoretical regression of $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$ on $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$. As a result, this residual variance, or marginal information matrix, will generally be smaller than $\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)$, which corresponds to the infeasible ML estimator of $\boldsymbol{\eta}$ that we could compute if the $\varsigma_t(\boldsymbol{\theta}_0)'s$ were directly observed. The following proposition characterises the ranking of the asymptotic covariance matrices of the five estimators of $\boldsymbol{\eta}$ that we have considered:

Proposition 14 *1. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, then $\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) \le \mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) \le \mathcal{F}(\boldsymbol{\phi}_0)$.*

2. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ with $\nu_0 > 8$, then $\mathcal{F}(\boldsymbol{\phi}_0) \le \mathcal{H}(\boldsymbol{\phi}_0)$. If in addition

$$\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{W}_s(\boldsymbol{\phi}_0) = \frac{(N+\nu_0-2)}{(\nu_0-4)}\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathbf{W}_s(\boldsymbol{\phi}_0), \tag{16}$$

then $\mathcal{H}(\boldsymbol{\phi}_0) \le \mathcal{G}(\boldsymbol{\phi}_0)$, with equality if and only if

$$\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2(N+\nu_0-2)}{N(\nu_0-4)}\mathbf{W}'_s(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}\mathbf{t}}(\boldsymbol{\theta}_0, 0) = 0\ \forall t. \tag{17}$$

Condition (16) is trivially satisfied in Gaussian models, and in dynamic univariate models with no mean. Also, it is worth mentioning that (17), which in turn implies (16), is satisfied by most dynamic univariate GARCH-M models (see Fiorentini, Sentana and Calzolari (2004)).

Given that $\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) = \mathbf{0}$ under normality from Proposition 9, it is clear that $\tilde{\boldsymbol{\eta}}_T$ will be as asymptotically efficient as the feasible ML estimator $\hat{\boldsymbol{\eta}}_T$ when $\boldsymbol{\eta}_0 = \mathbf{0}$, which in turn is as efficient as the infeasible ML estimator in that case. Moreover, if we use a multivariate student t log-likelihood function, these estimators will share the same half normal asymptotic distribution under conditional normality, although they would not necessarily be equal when they are not zero. Similarly, the asymptotic distributions of $\breve{\eta}_T$ and $\mathring{\eta}_T$ will also tend to be half normal as the sample size increases when $\eta_0 = 0$, since $\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)$ is root-T consistent for κ, which is 0 in the Gaussian case. However, while $\mathring{\eta}_T$ will always be as efficient as $\hat{\eta}_T$ under normality because $\mathring{n}_{\eta t}(\boldsymbol{\theta}_0, 0)$ is proportional to $s_{\eta t}(\boldsymbol{\theta}_0, 0)$, $\breve{\eta}_T$ will be less efficient unless condition (17) is satisfied.

5 Distributional misspecification and parameter consistency

5.1 Pseudo-true values of the parameters

So far, we have maintained the assumption that the conditional distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ is either *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ or sometimes *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$. However, one of the most important reasons for the popularity of the Gaussian pseudo-ML estimator of $\boldsymbol{\theta}$ despite its inefficiency is that it remains root-T consistent and asymptotically normally distributed under fairly weak distributional assumptions provided that (1) is true. In contrast, the efficient spherically-based ML estimator may become inconsistent if the true distribution of $\boldsymbol{\varepsilon}_t^*$ given $\mathbf{z}_t$ and I_{t-1} does not coincide with the assumed one, even though (1) holds, as forcefully argued by Newey and Steigerwald (1997) in the univariate case. To focus our discussion, in the remaining of this section we shall assume that (1) is true, and that we specifically decide to use the student t log-likelihood function for estimation purposes. Nevertheless, our results can be easily extended to any other spherically-based likelihood estimators. In this sense, the main advantage of the student t likelihood four our purposes is the fact that its limiting relationship to the Gaussian distribution can be made explicit. For simplicity, we shall also define the pseudo-true values of $\boldsymbol{\theta}$ and η as consistent roots of the expected t pseudo log-likelihood score, which under appropriate regularity conditions will maximise the expected value of the t pseudo log-likelihood function.

Two important points to bear in mind in studying the potential inconsistencies in $\hat{\boldsymbol{\theta}}_T$ are (i) that the spherical distribution assumed for estimation purposes will often nest the Gaussian distribution as a limiting case, and (ii) that $\hat{\boldsymbol{\theta}}_T = \tilde{\boldsymbol{\theta}}_T$ whenever $\hat{\boldsymbol{\eta}}_T = \mathbf{0}$. For instance, the t distribution is estimated subject to the inequality constraint $\eta \geq 0$. The following proposition explains the consequences of this inequality restriction:

Proposition 15 *1. Let $\boldsymbol{\phi}_\infty$ denote the pseudo-true values of the parameters $\boldsymbol{\theta}$ and η implied by a multivariate student t log-likelihood function. If the unconditional coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$ is not positive, where the expectation in (5) is taken with respect to the true unconditional distribution of the data, then $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ and $\eta_\infty = 0$.*

2. If the unconditional coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$ is strictly negative, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}\hat{\eta}_T = o_p(1)$ and $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$.

3. If the unconditional coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$ is exactly 0, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}\hat{\eta}_T$ will have an asymptotic normal distribution censored from below at 0, and $\tilde{\boldsymbol{\theta}}_T$ will be identical to $\hat{\boldsymbol{\theta}}_T$ with probability approaching 1/2. If in addition

$$E\left\{\left[(N+2-\varsigma_t)\boldsymbol{\varepsilon}_t^{*\prime}, (N+2-\varsigma_t)\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}\right]\mathbf{Z}_{dt}'|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\right\} = \mathbf{0}, \tag{18}$$

then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$ the rest of the time.

In the rest of this section we will concentrate on those distributions for which the condition in Proposition 15 is violated. The following proposition generalises the first part of Theorem 1 in Newey and Steigerwald (1997) to a broad class of multivariate dynamic models.

Proposition 16 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ but not t, with $\kappa_0 > 0$, and we can find a homeomorphic transformation $\mathbf{r}_s(.) = [\mathbf{r}_{1s}'(.), r_{2s}'(.)]'$ of the conditional mean and variance parameters $\boldsymbol{\theta}$ into an alternative set of parameters $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1', \vartheta_2')'$, where ϑ_2 is a scalar and $\mathbf{r}_s(\boldsymbol{\theta})$ is twice continuously differentiable with rank$[\partial\mathbf{r}_s'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that (14) holds, then the pseudo-true value of feasible student-t based ML estimator of $\boldsymbol{\vartheta}_1$, $\boldsymbol{\vartheta}_{1\infty}$, is equal to the true value $\boldsymbol{\vartheta}_{10}$.*

Importantly, note that the transformed parameters that we can estimate in a partially adaptive manner by means of the SSP estimator coincide with the parameters that we continue to estimate consistently with a misspecified student t-based pseudo-ML estimator.

If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ is not *i.i.d. spherical*, and $\kappa_0 > 0$, then in general the feasible student t-based ML estimator will be inconsistent, and the same applies to the SSP estimator. However, it may still be possible to estimate consistently some parameters:

Proposition 17 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}$ is i.i.d. $(\mathbf{0}, \mathbf{I}_N)$ but not spherical, with $\kappa_0 > 0$, and we can find a homeomorphic transformation $\mathbf{r}_g(.) = [\mathbf{r}_{1g}'(.), \mathbf{r}_{2g}'(.), \mathbf{r}_{3g}'(.)]'$ of the conditional mean and variance parameters $\boldsymbol{\theta}$ into an alternative set of parameters $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \boldsymbol{\delta}_3')'$, where $\boldsymbol{\delta}_2 = vech(\boldsymbol{\Delta}_2)$, $\boldsymbol{\Delta}_2$ is an unrestricted positive (semi)definite matrix of order N, $\boldsymbol{\delta}_3$ is $N \times 1$, and $r_g(\boldsymbol{\theta})$ is twice continuously differentiable with rank$\left[\partial\mathbf{r}_g'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that (15) holds, then the pseudo-true value of feasible student-t based ML estimator of $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_{1\infty}$, is equal to the true value $\boldsymbol{\delta}_{10}$.*

This proposition is the multivariate generalisation of Theorem 2 in Newey and Steigerwald (1997).[5] In simple terms, it says that we cannot estimate consistently either the mean or the covariance matrix of the *i.i.d.* pseudo-standardised residuals $\boldsymbol{\Sigma}_t^{*-1/2}(\boldsymbol{\delta}_{10})[\mathbf{y}_t - \boldsymbol{\mu}_t^*(\boldsymbol{\delta}_{10})] = \boldsymbol{\delta}_3 + \boldsymbol{\Delta}_2^{1/2}\boldsymbol{\varepsilon}_t^*$. As discussed at the end of section 3.3, though, we may only be able to write the conditional mean and covariance functions as in (15) at the cost of augmenting the model with a large number of additional parameters, which will generally lead to an inefficiency loss.

Importantly, note that the transformed parameters that we can estimate in a partially adaptive manner by means of the unrestricted semiparametric estimator coincide with the parameters that we continue to estimate consistently with a misspecified student-t based ML estimator. However, if the *i.i.d.* assumption does not hold, the semiparametric estimator may also become inconsistent.[6] In this sense, it is important to bear in mind that in non-elliptical models the conditional distribution of $\mathbf{y}_t$ is not invariant to the specific choice of $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ assumed to generate the data, a choice that could conceivably change over time (see Mencía and Sentana (2005)).

5.2 Hausman tests

There are several ways in which we can test the validity of the multivariate t assumption. One possibility is to nest that distribution within a more flexible parametric family, which allows us to conduct an LM test of the nesting restrictions. This is the approach in Mencía and Sentana (2005), who use the generalised hyperbolic family as the nesting distribution. But we can also consider a Hausman specification test. The rationale is that the feasible elliptical ML estimator $\hat{\boldsymbol{\theta}}_T$ is efficient under correct specification of the conditional distribution of $\mathbf{y}_t$. In contrast, if the conditional mean and variance of $\mathbf{y}_t$ are correctly specified, but the conditional distribution of ε_t^* is not *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \eta)$, then $\tilde{\boldsymbol{\theta}}_T$ will remain root-T consistent as long as κ_0 is bounded, while $\hat{\boldsymbol{\theta}}_T$ will probably not, as Propositions 16 and 17 illustrate. More formally

Proposition 18 *Let*

$$H^W_{\hat{\boldsymbol{\theta}}_T} = T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)' \left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right]^+ (\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T),$$

and

$$H^s_{\hat{\boldsymbol{\theta}}_T} = T\bar{\mathbf{s}}'_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0}) \left[\mathcal{B}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)\right]^+ \bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0}),$$

where $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0})$ is the sample average of the Gaussian PML score evaluated at the feasible ML estimator $\hat{\boldsymbol{\theta}}_T$. If the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $H^W_{\hat{\boldsymbol{\theta}}_T} \xrightarrow{d} \chi^2_s$ and $H^W_{\hat{\boldsymbol{\theta}}_T} - H^s_{\hat{\boldsymbol{\theta}}_T} = o_p(1)$ under correct specification of the conditional distribution of $\mathbf{y}_t$, where $s = rank\left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right]$.

[5] It is also possible to generalise the second part of their Theorem 1, in the sense that if the true conditional mean of $\mathbf{y}_t$ is $\mathbf{0}$, and we impose this restriction in estimation, then $\boldsymbol{\delta}_3$ is unnecessary.

[6] Hodgson (2000) shows that consistency of the conditional mean parameters is preserved in conditionally homoskedastic non-linear regression models when the innovations are not *i.i.d.* if certain conditions are satisfied.

In practice, we must replace $\mathcal{A}(\boldsymbol{\phi}_0)$, $\mathcal{B}(\boldsymbol{\phi}_0)$ and $\mathcal{I}(\boldsymbol{\phi}_0)$ by consistent estimators to make $H^W_{\hat{\boldsymbol{\theta}}_T}$ and $H^s_{\hat{\boldsymbol{\theta}}_T}$ operational. In order to guarantee the positive semidefiniteness of their weighting matrices, it is convenient to estimate all these matrices as the sample averages of the corresponding conditional expressions in Propositions 1 and 2 evaluated at a common estimator of $\boldsymbol{\phi}$, such as $\hat{\boldsymbol{\phi}}_T$, $(\tilde{\boldsymbol{\theta}}_T, \tilde{\eta}_T)$ or $(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$, the latter being such that $\mathcal{B}(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$ is always bounded.

In view of Proposition 9, though, such feasible Hausman tests will become numerically unstable when $\hat{\eta}_T > 0$ but $\eta_0 = 0$ even though in theory they should be identically 0 because $\left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right] = \mathbf{0}$ in that case.

Given that the power of these Hausman tests depends on the asymptotic biases of $\hat{\boldsymbol{\theta}}_T$ under misspecification of the conditional distribution of the standardised innovations, it may be convenient to concentrate on those parameters which may be more affected by such distributional misspecification. For instance, in the situation discussed in Proposition 16 power would be maximised if we based our Hausman test on ϑ_2 exclusively, and the same will be true of $\boldsymbol{\delta}_2$ and $\boldsymbol{\delta}_3$ in the context of Proposition 17.

Given that the SSP estimator is also efficient relative to the PML estimator under sphericity, but it may lose its consistency otherwise, we can consider alternative specification tests as follows:

Proposition 19 *Let*

$$H^W_{\mathring{\boldsymbol{\theta}}_T} = T(\tilde{\boldsymbol{\theta}}_T - \mathring{\boldsymbol{\theta}}_T)' \left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\right]^+ (\tilde{\boldsymbol{\theta}}_T - \mathring{\boldsymbol{\theta}}_T),$$

and

$$H^s_{\mathring{\boldsymbol{\theta}}_T} = T\bar{\mathbf{s}}'_{\boldsymbol{\theta}T}(\mathring{\boldsymbol{\theta}}_T, \mathbf{0}) \left[\mathcal{B}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\phi}_0)\mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)\right]^+ \bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\mathring{\boldsymbol{\theta}}_T, \mathbf{0}),$$

where $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\mathring{\boldsymbol{\theta}}_T, \mathbf{0})$ is the sample average of the Gaussian PML score evaluated at the SSP estimator $\mathring{\boldsymbol{\theta}}_T$. If the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $H^W_{\mathring{\boldsymbol{\theta}}_T} \xrightarrow{d} \chi^2_s$ and $H^W_{\mathring{\boldsymbol{\theta}}_T} - H^s_{\mathring{\boldsymbol{\theta}}_T} = o_p(1)$ under correct specification of the conditional distribution of $\mathbf{y}_t$, where $s = rank\left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\right]$.

Once again, it may be convenient to concentrate on the parameters that are more likely to reflect the misspecification.

Finally, the difference between $\tilde{\eta}_T$ and $\hat{\eta}_T$ suggests yet another Hausman specification test of the model, which will be given by the following expression:

$$H^W_{\eta T} = T(\tilde{\eta}_T - \hat{\eta}_T)^2 \left[\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)\right]^+,$$

where the Moore-Penrose generalised inverse in this scalar case is simply the reciprocal of $\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ if $\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ is positive, and 0 otherwise. Under correct specification of the conditional distribution of $\boldsymbol{\varepsilon}^*_t$, $H^W_{\eta T}$ will be asymptotically distributed as a chi-square with one degree of freedom when $\eta_0 > 0$. But again, feasible versions of $H^W_{\eta T}$ may become numerically unstable when $\hat{\eta}_T > 0$ or $\tilde{\eta}_T > 0$ but $\eta_0 = 0$, even though the infeasible version would be identically 0 because $[\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)] = 0$ in that case. Note that the power of this third Hausman test depends on the difference between the pseudo true values of $\tilde{\eta}_T$ and $\hat{\eta}_T$ when the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is not multivariate t, which will depend in turn on the asymptotic bias in $\hat{\boldsymbol{\theta}}_T$.

6 Monte Carlo comparisons

6.1 Univariate results (to be completed)

In this section, we assess the finite sample performance of the different estimators and testing procedures discussed above by means of an extensive Monte Carlo exercise, with an experimental design borrowed from Bollerslev and Wooldridge (1992). Specifically, we simulate and estimate model (12), where $\pi_0 = 0$, $\rho_0 = .5$, $\omega_0 = .05$, $\alpha_0 = .15$ and $\beta_0 = .8$. As for η_0, we consider a Gaussian distribution, and two student t's with 8 and 4 degrees of freedom respectively. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 1,000$ observations based on 10,000 Monte Carlo replications.

Following Fiorentini, Sentana and Calzolari (2003), our estimation procedure employs the following mixed approach: initially, we use a scoring algorithm with a fairly large tolerance criterion; then, after "convergence" is achieved, we switch to a Newton-Raphson algorithm to refine the solution. Both stages are implemented by means of the NAG Fortran 77 Mark 19 library E04LBF routine (see Numerical Algorithm Group 2001 for details), with the analytical expressions derived in Section 2 of that paper.

Figures 1a, 1b and 1c display kernel estimates of the sampling distributions of the ML (solid), SP (), SSP () and PML (dashed) estimators of the autoregressive coefficient ρ for $\nu_0 = \infty$, 8 and 4, respectively, constructed with the automatic bandwidth choice given in expression (3.28) of Silverman (1986). As expected from Proposition 9, the distribution of these four estimators is almost identical under normality, which is not very surprising given that the ML and PNL are numerically identical over half the time. However, they progressively differ as the degrees of freedom decrease. In this respect, it is important to mention that the distribution of the SP, SSP and PML estimator of ρ remain Gaussian even when $\nu_0 = 4$ because the different asymptotic variances turn out to be block diagonal matrices in this model (see Theorem 4 in Engle (1982)).

Figures 2a-c and 3a-c display the corresponding kernel estimates of the sampling distributions of the estimators of the ARCH and GARCH parameters α and β, respectively. Again, there is no noticeable differences in the Gaussian case, but the differences become apparent as the distribution of the standardised innovations becomes more leptokurtic. In fact, when $\nu_0 = 4$

the shape of the distribution of the PML estimators $\tilde{\alpha}_T$ and $\tilde{\beta}_T$ is clearly non-standard even for $T = 1,000$, as discussed in Section 3.

Finally, Figures 4a, 4b and 4c display kernel estimates of the sampling distributions of the ML (solid), sequential ML (dashed) and sequential MM (dash-dotted) estimators of η when $\nu_0 = \infty$, 8 and 4, respectively, together with the fraction of parameter values estimated at the lower bound of 0. Given that there is considerable probability mass on or near the origin, we have used the reflection methods discussed by Silverman (1986) to construct those densities in order to guarantee that they integrate to 1. As can be seen, the proportions of zero estimates of η exceed the theoretical value of 1/2 for $\eta_0 = 0$. Although the three estimators behave similarly under Gaussianity, they are radically different in the other two cases. As explained in Section 4, while $\hat{\eta}_T$ is asymptotically normally distributed in those two cases, $\breve{\eta}_T$ is not root-T when $\nu_0 = 8$ or $\nu_0 = 4$, neither is $\tilde{\eta}_T$ in the latter case.

In order to assess the effects of misspecification on the estimators, we have consider three additional conditional distributions for ε_t^*: an *i.i.d.* symmetric normal-gamma mixture with the same kurtosis coefficient as the t_8, an *i.i.d.* asymmetric student t distribution, also with the same kurtosis coefficient, but with the largest negative asymmetry possible; and the student t with time-varying degrees of freedom considered by Demos and Sentana (1998). We generate the standardised normal-gamma mixture as . Similarly, we generate the standardised asymmetric student t distribution as . Finally, the process for the degrees of freedom that we have considered is , with . Figure 5a displays the density functions of the six distributions that we have considered, while Figure 5b shows a zoom of their left tail.

Given Propositions 13 and 17, Figures 5a to

Finally, Figures presents

Standard errors are computed using analytical derivatives based on the expressions in Proposition 1 in the case of the t, and Proposition 2 in the Gaussian case.

6.2 Multivariate example (to be completed)

In this section, we assess the finite sample performance of the different estimators and testing procedures discussed above by means of an extensive Monte Carlo exercise, with an experimental design that augments one in Sentana and Fiorentini (2001) with diagonal VAR(1) dynamics, as in model (13). Specifically, we simulate and estimate a 6-variate version of the model (13), where $\boldsymbol{\pi}_0 = 0$, $\boldsymbol{\rho}_0 = .01 \cdot \boldsymbol{\iota}_6$, $\mathbf{c}_0 = \boldsymbol{\iota}_6$, $\boldsymbol{\gamma}_0 = 2 \cdot \boldsymbol{\iota}_6$, $\alpha_0 = .15$ and $\beta_0 = .8$. As for η_0, we also consider a Gaussian distribution, and two student t's with 8 and 4 degrees of freedom respectively. We also consider an *i.i.d.* normal-gamma distribution with the same coefficient of multivariate excess kurtosis as the t_8, an *i.i.d.* asymmetric student t with the maximum negative skewness possible for the same excesss kurtosis, and a symmetric student t distribution with time-varying degrees of freedom. Again, we only report the results for $T = 1,000$ observations based on 10,000 Monte Carlo replications.

7 Conclusions (to be completed)

In the context of the general multivariate dynamic regression model with time-varying variances and covariances considered by Bollerslev and Wooldridge (1992), we compare the efficiency of the feasible ML procedure that jointly estimates the shape parameters with the efficiency of the infeasible ML, SSP, SP and PML estimators of the conditional mean and variance parameters considered in the existing literature. In this respect, we show that if the distribution of the standardised innovations is *i.i.d.* spherical, the ranking is infeasible ML, feasible ML, SSP, SP and PML, with equality if and only if the spherical distribution is in fact Gaussian, in which case there is no efficiency loss in simultaneously estimating the shape parameters. In this respect, our results generalise earlier findings by Gonzalez-Rivera and Drost (1999), Fiorentini, Sentana and Calzolari (2003) and Hafner and Rombouts (2006).

Furthermore, we study in detail two popular examples of conditionally heteroskedastic models, one univariate and the other one multivariate, and obtain closed-formed expressions for the inefficiency ratios of different subsets of parameters under the assumption of constant variances. Importantly, those inefficiency ratios coincide with the ratios of the non-centrality parameters of the tests of conditional homoskedasticity associated with the different estimators.

More generally, we show that the SSP estimator is adaptive for all but one global scale parameter in an appropriate reparametrisation of the model in which the conditional covariance matrix is proportional to this global scale parameter. We also show that the general SP estimator is adaptive for a much more restricted set of parameters in an alternative reparametrisation that only seems to fit the constant conditional correlation model of Bollerslev (1987) with a zero conditional mean. These results directly generalise the ones obtained for univariate GARCH models by Linton (1993) and Drost and Klaassen (1997), respectively, as well as the results in Hodgson and Vorkink (2003) for a specific multivariate GARCH-M model.

We also thoroughly analyse the effects of distributional misspecification on the consistency of the conditional mean and variance parameters. In particular, we initially show that when the conditional distribution is platykurtic, so that the coefficient of multivariate excess kurtosis is negative, the feasible ML estimators based on the multivariate student distribution converge to the Gaussian PML estimators. On the other hand, we show that when the conditional distribution is spherical and leptokurtic, but neither t nor Gaussian, the feasible student t-

based ML estimator is consistent for exactly the same parameters for which the SSP estimator is adaptive, which are effectively all but a global scale factor. Furthermore, we show that when the conditional distribution is leptokurtic but not spherical, the feasible ML estimator is consistent for exactly the same restricted subset of parameters for which the general SP estimator is adaptive, which excludes both the mean and the covariance matrix of the *i.i.d.* pseudo-standardised innovations. These results generalise Newey and Steigerwald's (1997) Theorems 1 and 2, respectively, which they obtained for univariate models. In this respect, we would like to emphasise that our inconsistency results apply not only to the multivariate student t log-likelihood, but also to any other spherically-based likelihood estimators. The main advantage of the student t for our purposes is that we can make explicit its limiting relationship to the Gaussian distribution.

In view of the importance of the distributional assumptions, we propose simple Hausman tests that compare the feasible ML and SSP estimators to the Gaussian PML ones.

Finally, we also consider sequential estimators of the shape parameters, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators. In particular, we consider a sequential ML estimator, as well as sequential MM estimators based on the coefficient of multivariate excess kurtosis. The main advantage of such estimators is that they preserve the consistency of the conditional mean and variance functions, but at the same time allow for a more realistic conditional distribution. We show that the usual efficiency ranking of the estimators of the shape parameters is infeasible ML, feasible ML, sequential ML and sequential MM. These results are important in practice because empirical researchers often want to go beyond the first two conditional moments, which implies that one cannot simply treat the shape parameters as if they were nuisance parameters. We also propose an alternative Hausman test that compares the feasible ML estimator of those parameters to the sequential ML one.

In a detailed Monte Carlo experiment we find that

Further work is required in at least four directions. First, from a modelling point of view, the assumption of *i.i.d.* innovations in non-spherical multivariate models seems rather strong, for it forces the conditional distribution of the observed variables to depend on the choice of square root matrix used to obtain the underlying innovations from the observations. Secondly, from an estimation point of view, the development of semiparametric estimators that do not require the assumption of *i.i.d.* innovations remains an important unresolved issue that merits further investigation. Thirdly, the availability of analytical finite sample results would probably make the choice between bias and efficiency look more balanced than what standard root-T asymptotics suggests. Finally, the existing literature, including our paper, places too much emphasis on parameter estimation, while practitioners are often more interested in functionals of the conditional distribution, such as the forecasting intervals required in value at risk calculations. An evaluation of the consequences that the different estimation procedures that we have considered have for such objects constitutes a fruitful avenue for future research.

Appendix

Proofs and auxiliary results

Some useful distribution results

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\circ$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\circ = e_t\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is a non-negative random variable independent of $\mathbf{u}_t$, whose distribution determines the distribution of $\boldsymbol{\varepsilon}_t^\circ$. The variables e_t and $\mathbf{u}_t$ are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^\circ$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\circ$ is distributed as a standardised multivariate student t random vector of dimension N with ν_0 degrees of freedom, then $e_t = \sqrt{(\nu_0 - 2)\zeta_t/\xi_t}$, where ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $\nu_0 > 2$ and variance $2\nu_0$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis κ_0, which is given by $E(e_t^4)/[N(N+2)] - 1$, will also be bounded. For instance, $\kappa_0 = 2/(\nu_0 - 4)$ in the student t case with $\nu_0 > 4$, and $\kappa_0 = 0$ under normality. In this respect, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^\circ$ is proportional to $\mathbf{u}_t$, then $\kappa_0 \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

Then, it is easy to combine the representation of elliptical distributions above with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of a spherically symmetric distribution with $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}, \tag{A1}$$

and

$$E(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime})vec'(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^\circ)] = (\kappa_0+1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)], \tag{A2}$$

respectively.

We shall also make use of the fact that in the student t case $\zeta_t/(\xi_t+\zeta_t)$ has a beta distribution with parameters $N/2$ and $\nu_0/2$, which is independent of $\mathbf{u}_t$. As is well known, if a random variable X defined over $[0,1]$ has a beta distribution with parameters (a,b), where $a > 0$, $b > 0$, then its density function is

$$f_X(x) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1},$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the usual beta function. Fortunately, it is often trivial to find apparently complex moments of a beta random variable from first principles. For instance,

$$E[X^p(1-X)^q] = \frac{1}{B(a,b)}\int_0^1 x^p(1-x)^q x^{a-1}(1-x)^{b-1}dx = \frac{B(a+p,b+q)}{B(a,b)}$$

for any real values of p and q such that $a+p>0$ and $b+q>0$. Another particularly convenient moment for our purposes is $E[X^p\ln(1-X)]$. But since

$$\int_0^1 \ln(1-x)x^{a+p-1}(1-x)^{b-1}dx = \frac{\partial}{\partial b}\int_0^1 x^{a+p-1}(1-x)^{b-1}dx = \frac{\partial}{\partial b}B(a+p,b),$$

then we can write

$$\begin{aligned} E[X^p(1-X)^q\ln(1-X)] &= \frac{B(a+p,b+q)}{B(a,b)}\frac{\partial \ln B(a+p,b+q)}{\partial b} \\ &= \frac{B(a+p,b+q)}{B(a,b)}\left[\psi(b+q)-\psi(a+p+b+q)\right], \end{aligned}$$

thanks to the definition of the beta function in terms of the gamma function above.

Proposition 1

For our purposes it is convenient to rewrite $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ as

$$\begin{aligned} \mathbf{e}_{lt}(\boldsymbol{\phi}_0) &= -2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0\right]}{\partial\varsigma}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) = -\sqrt{\varsigma_t}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t, \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0) &= -vec\left\{\mathbf{I}_N + 2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0\right]}{\partial\varsigma}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right\} = -vec\left[\mathbf{I}_N + \varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t\mathbf{u}_t'\right], \end{aligned}$$

where ς_t and $\mathbf{u}_t$ are mutually independent for any standardised spherical distribution, with $E(\mathbf{u}_t)=\mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$, $E(\varsigma_t)=N$ and $E(\varsigma_t^2)=N(N+2)(\kappa_0+1)$. Importantly, we only need to compute unconditional moments because ς_t and $\mathbf{u}_t$ are independent of $\mathbf{z}_t$ and I_{t-1} by assumption. Then, it easy to see that

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_0)] = E[-\sqrt{\varsigma_t}2\partial g(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma]\cdot E(\mathbf{u}_t) = \mathbf{0},$$

and that

$$E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)] = -vec\left\{\mathbf{I}_N + E\left[\varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]\cdot E(\mathbf{u}_t\mathbf{u}_t')\right\} = -vec(\mathbf{I}_N)\left\{1+E\left[\frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]\right\}.$$

In this context, we can use expression (2.21) in Fang, Kotz and Ng (1990) to write the density function of ς_t as

$$h(\varsigma_t;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})], \tag{A3}$$

whence

$$\left[\frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta})}{\partial\varsigma}+1\right] = \frac{2}{N}\left[1+\varsigma_t\frac{\partial\ln h(\varsigma_t;\boldsymbol{\eta})}{\partial\varsigma}\right]. \tag{A4}$$

On this basis, we can use the fact that

$$E(\varsigma_t) = \int_0^\infty \varsigma h(\varsigma;\boldsymbol{\eta})d\varsigma = N < \infty$$

to show that

$$E\left[\varsigma_t \frac{\partial \ln h(\varsigma_t;\boldsymbol{\eta})}{\partial \varsigma}\right] = \int_0^\infty \varsigma \frac{\partial h(\varsigma_t;\boldsymbol{\eta})}{\partial \varsigma} d\varsigma = -1, \tag{A5}$$

which implies that $E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)] = \mathbf{0}$, as required.

Similarly, we can also show that

$$\begin{aligned} E[\mathbf{e}_{lt}(\boldsymbol{\phi}_0)\mathbf{e}'_{lt}(\boldsymbol{\phi}_0)] &= E\left\{\varsigma_t\left[\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]^2 \cdot \mathbf{u}_t\mathbf{u}'_t\right\} = \mathbf{I}_N \cdot E\left\{\frac{\varsigma_t}{N}\left[\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]^2\right\}, \\ E[\mathbf{e}_{lt}(\boldsymbol{\phi}_0)\mathbf{e}'_{st}(\boldsymbol{\phi}_0)] &= -E\left\{\sqrt{\varsigma_t}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\mathbf{u}_t vec'\left[\mathbf{I}_N + \varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t\mathbf{u}'_t\right]\right\} = \mathbf{0} \end{aligned}$$

by virtue of (A1), and

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)\mathbf{e}'_{st}(\boldsymbol{\phi}_0)] &= E\left\{vec\left[\mathbf{I}_N + \varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t\mathbf{u}'_t\right] vec'\left[\mathbf{I}_N + \varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t\mathbf{u}'_t\right]\right\} \\ &= E\left[\varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]^2 \frac{1}{N(N+2)}[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N)] \\ &\quad + 2E\left[\frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right] vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N) + vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N) \\ &= \frac{N}{(N+2)}E\left[\frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]^2 (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) \\ &\quad + \left\{\frac{N}{(N+2)}E\left[\frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]^2 - 1\right\} vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N)] \end{aligned}$$

by virtue of (A2), (A4) and (A5).

Finally, given that

$$\mathbf{e}_{rt}(\boldsymbol{\phi}_0) = \frac{\partial c(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} + \frac{\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}},$$

it is clear $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ will be a function of ς_t but not of $\mathbf{u}_t$, which immediately implies that $E[\mathbf{e}_{lt}(\boldsymbol{\phi}_0)\mathbf{e}'_{rt}(\boldsymbol{\phi}_0)] = \mathbf{0}$, and that

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)\mathbf{e}'_{rt}(\boldsymbol{\phi}_0)] &= E\left\{-vec\left[\mathbf{I}_N + \varsigma_t\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\cdot\mathbf{u}_t\mathbf{u}'_t\right]\mathbf{e}'_{rt}(\boldsymbol{\phi}_0)\right\} \\ &= vec(\mathbf{I}_N)E\left\{-\left[1 + \frac{\varsigma_t}{N}\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\right]\mathbf{e}'_{rt}(\boldsymbol{\phi}_0)\right\}. \end{aligned}$$

□

Proposition 2

The proof is based on a straightforward application of Proposition 1 in Bollerslev and Wooldridge (1992) to the spherically symmetric case. Since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, and

$\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ is a vector martingale difference sequence, then to obtain $\mathcal{B}_t(\boldsymbol{\phi}_0)$ we only need to compute $V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0]$. But since

$$\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}_0, \mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0}) \end{array}\right] = \left(\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\,[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N] \end{array}\right) = \left[\begin{array}{c} \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\varsigma_t \mathbf{u}_t \mathbf{u}_t' - \mathbf{I}_N) \end{array}\right]$$

for any spherical distribution, with ς_t and $\mathbf{u}_t$ both mutually and serially independent, then (4) follows from (A1) and (A2). As for $\mathcal{A}_t(\boldsymbol{\phi}_0)$, we know that its formula, which is valid regardless of the exact nature of the true conditional distribution, coincides with $\mathcal{B}_t(\boldsymbol{\phi}_0)$ when $\kappa_0 = 0$ by the (conditional) information matrix equality. □

Proposition 3

We can use the conditional version of the generalised information matrix equality (see e.g. Newey and McFadden (1994)) to show that

$$\begin{aligned} E\left\{ \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}) \left[\mathbf{s}_{\boldsymbol{\theta}t}'(\boldsymbol{\phi}), \mathbf{s}_{\boldsymbol{\eta}t}'(\boldsymbol{\phi})\right] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right\} &= -E\left\{ \left[\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})}{\partial \boldsymbol{\theta}'} \middle| \frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})}{\partial \boldsymbol{\eta}'} \right] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right\} \\ &= -E\left\{ [\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\mathbf{0}] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi} \right\} = [\mathcal{A}_t(\boldsymbol{\phi})|\mathbf{0}] \end{aligned}$$

irrespective of the conditional distribution of $\boldsymbol{\varepsilon}_t^*$, where we have used the fact that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$ does not vary with $\boldsymbol{\eta}$ when regarded as the estimation function for $\tilde{\boldsymbol{\theta}}_T$. Then, the required result follows from the martingale difference nature of both $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ and $\mathbf{e}_t(\boldsymbol{\phi}_0)$. □

Proposition 4

We can use standard arguments (see e.g. Newey and McFadden (1994)) to show that the sequential ML estimator of $\boldsymbol{\eta}$ is asymptotically equivalent to a MM estimator based on the linearised influence function

$$s_{\eta t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathcal{I}_{\theta\eta}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \mathbf{0}).$$

Then, the expression for $\mathcal{F}(\boldsymbol{\phi}_0)$ follows from the definitions of $\mathcal{B}(\boldsymbol{\phi}_0)$, $\mathcal{C}(\boldsymbol{\phi}_0)$ and $\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ in Propositions 1 and 2, together with the martingale difference nature of $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ and $\mathbf{e}_t(\boldsymbol{\phi}_0)$.□

Proposition 5

In this case, the linearised influence functions corresponding to $\breve{\eta}_T$ and $\mathring{\eta}_T$ are

$$n_{\eta t}(\boldsymbol{\theta}_0, \eta) - \mathcal{R}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

and

$$\mathring{n}_{\eta t}(\boldsymbol{\theta}_0, \eta) - \mathcal{Q}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

respectively, whence we can directly obtain the formulae for $\mathcal{G}(\boldsymbol{\phi}_0)$ and $\mathcal{H}(\boldsymbol{\phi}_0)$. Therefore, the only remaining task is to obtain closed-form expressions for the required moments. In this respect, we can use the law of iterated expectations to show that

$$\begin{aligned} cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0),n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] &= \mathbf{Z}_d(\boldsymbol{\phi}_0)\cdot E[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)\cdot n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)|\varsigma_t;\boldsymbol{\phi}_0] \\ &= \mathbf{W}_s(\boldsymbol{\phi}_0)E\left[\left(\frac{\varsigma_t}{N}-1\right)n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)\right|\boldsymbol{\phi}_0\bigg], \end{aligned}$$

and

$$\begin{aligned} &cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\eta_0),n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] = \mathbf{Z}_d(\boldsymbol{\phi}_0)\cdot E[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\eta_0)\cdot n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)|\varsigma_t;\boldsymbol{\phi}_0] \\ &\qquad = \mathbf{W}_s(\boldsymbol{\phi}_0)E\left[\left(\frac{N+\nu_0}{\nu_0-2+\varsigma_t}\frac{\varsigma_t}{N}-1\right)n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)\right|\boldsymbol{\phi}_0\bigg]. \end{aligned}$$

Then, we can use the properties of the beta distribution discussed before to show that

$$\begin{aligned} E\left[\left(\frac{\varsigma_t^2}{N(N+2)}-\frac{\nu_0-2}{\nu_0-4}\right)^2\right] &= \frac{(\nu_0-2)^2}{(\nu_0-4)^4}\left[\frac{(N+6)(N+4)}{N(N+2)}\frac{(\nu_0-2)(\nu_0-4)}{(\nu_0-6)(\nu_0-8)}-1\right], \\ E\left[\left(\frac{\varsigma_t}{N}-1\right)\left(\frac{\varsigma_t^2}{N(N+2)}-\frac{\nu_0-2}{\nu_0-4}\right)\right] &= \frac{4(\nu_0-2)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}, \end{aligned}$$

and

$$E\left[\left(\frac{N+\nu_0}{\nu_0-2+\varsigma_t}\frac{\varsigma_t}{N}-1\right)\left(\frac{\varsigma_t^2}{N(N+2)}-\frac{\nu_0-2}{\nu_0-4}\right)\right] = \frac{4(\nu_0-2)}{N(\nu_0-4)}.$$

On the other hand, since $\mathring{n}_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ is the residual from the least squares projection of $n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ on $\varsigma_t/N-1$, we can obtain the relevant expressions for $\mathring{n}_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ from those of $n_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ by using the fact that

$$E\left[\left(\frac{\varsigma_t}{N}-1\right)^2\right] = \frac{2(N+\nu_0-2)}{N(\nu_0-4)}$$

and

$$E\left[\left(\frac{N+\nu_0}{\nu_0-2+\varsigma_t}\frac{\varsigma_t}{N}-1\right)\left(\frac{\varsigma_t}{N}-1\right)\right] = \frac{2}{N}.$$

□

Proposition 6

It trivially follows from (4) and (6) that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\phi})-\mathcal{K}(0)\,\mathcal{K}^{+}(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\,|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0}$$

for any spherically symmetric distribution. In addition, we also know that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\phi})-\mathcal{K}(0)\,\mathcal{K}^{+}(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0}.$$

Hence, the second summand of (8), which can be interpreted as $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the residual from the theoretical regression of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on a constant and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, belongs to the unrestricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with zero conditional means and bounded second moments that are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$.

Now, if we write (8) as

$$[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta})]\, \mathbf{e}_{dt}(\boldsymbol{\phi}) + \mathbf{Z}_d(\boldsymbol{\theta})\mathcal{K}\,(0)\, \mathcal{K}^+\,(\kappa)\, \mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0}),$$

we can use the law of iterated expectations to show that the semiparametric efficient score (8) evaluated at the true parameter values will be unconditionally orthogonal to the unrestricted tangent set because so is $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$, and $E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta})|\boldsymbol{\phi}\right] = \mathbf{0}$.

Finally, the expression for the semiparametric efficiency bound will be

$$E\left[\left.\begin{array}{c} \{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\right]\} \\ \times \{\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathbf{e}'_{dt}(\boldsymbol{\phi}) - \mathbf{e}'_{dt}(\boldsymbol{\theta}, 0)\mathcal{K}^+\,(\kappa)\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})\} \end{array}\right|\boldsymbol{\phi}\right]$$
$$= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right]$$
$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left[\mathbf{e}'_{dt}(\boldsymbol{\phi}) - \mathbf{e}'_{dt}(\boldsymbol{\theta}, \mathbf{0})\mathcal{K}^+\,(\kappa)\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})|\boldsymbol{\phi}\right\}$$
$$-E\left\{\mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\right]\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right\}$$
$$+E\left\{\mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta}, 0)\right]\left[\mathbf{e}'_{dt}(\boldsymbol{\phi}) - \mathbf{e}'_{dt}(\boldsymbol{\theta}, 0)\mathcal{K}^+\,(\kappa)\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})|\boldsymbol{\phi}\right\}$$
$$= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathcal{M}_{dd}\,(\boldsymbol{\eta}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\kappa)\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})$$

by virtue of (4), (6) and the law of iterated expectations. □

Proposition 7

First of all, it is easy to show that for any spherical distribution

$$\begin{aligned} \mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0, \mathbf{0}) &= E\left[\left.\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}_0, \mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0}) \end{array}\right| \varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\phi}_0\right] = E\left\{\left.\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{array}\right| \varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\phi}_0\right\} \\ &= E\left[\left.\begin{array}{c} \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N) \end{array}\right| \varsigma_t\right] = \left(\frac{\varsigma_t}{N} - 1\right)\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array}\right], \end{aligned} \tag{A5}$$

and

$$\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) = E\left[\left.\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}_0) \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0) \end{array}\right| \varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\phi}_0\right]$$
$$= -E\left\{\left.\begin{array}{c} 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\varsigma \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\mathbf{I}_N + 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\varsigma \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right] \end{array}\right| \varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\phi}_0\right\}$$
$$= -E\left\{\left.\begin{array}{c} 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\varsigma \cdot \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\mathbf{I}_N + 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\varsigma \cdot \varsigma_t\mathbf{u}_t\mathbf{u}'_t) \end{array}\right| \varsigma_t\right\} = -\left\{\frac{2\partial g(\varsigma_t, \boldsymbol{\eta}_0)}{\partial\varsigma}\frac{\varsigma_t}{N} + 1\right\}\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array}\right], \tag{A6}$$

where we have used again the fact that $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}'_t) = N^{-1}\mathbf{I}_N$, and $\varsigma_t(\boldsymbol{\theta}_0)$ and $\mathbf{u}_t$ are stochastically independent.

In addition, we can use the law of iterated expectations to show that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right]$$

and

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=E\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},0)|\boldsymbol{\phi}\right]=E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right].$$

Hence, to compute these matrices we simply need to obtain the scalar moments

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\frac{2\partial g(\varsigma_t,\boldsymbol{\eta})}{\partial\varsigma}\frac{\varsigma_t}{N}+1\right]\middle|\boldsymbol{\eta}\right\}$$

and

$$E\left[\left(\frac{\varsigma_t}{N}-1\right)^2\middle|\boldsymbol{\eta}\right].$$

In this respect, we can use (5) to show that the latter is simply $[(N+2)\kappa+2]/N$, so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=\frac{(N+2)\kappa+2}{N}\begin{pmatrix}\mathbf{0}&\mathbf{0}\\\mathbf{0}&vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{pmatrix}=\mathring{\mathcal{K}}(\kappa).$$

As for the former, we can use the fact that $E(\varsigma_t^2|\boldsymbol{\eta})=N(N+2)(\kappa+1)<\infty$ to show that

$$E\left[\varsigma_t^2\frac{\partial\ln h(\varsigma_t;\boldsymbol{\eta})}{\partial\varsigma}\middle|\boldsymbol{\eta}\right]=\int_0^\infty\varsigma^2\frac{\partial h(\varsigma;\boldsymbol{\eta})}{\partial\varsigma}d\varsigma=-2N.$$

If we then combine this result with (A4) and (A5), we will have that for any spherically symmetric distribution

$$-E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\frac{2\partial g\left(\varsigma_t,\boldsymbol{\eta}\right)}{\partial\varsigma}\frac{\varsigma_t}{N}+1\right]\middle|\boldsymbol{\eta}\right\}=\frac{2}{N},$$

so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=\mathring{\mathcal{K}}(0),$$

which coincides with the value of $E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]$ under normality.

Therefore, it trivially follows from the expressions for $\mathring{\mathcal{K}}(0)$ and $\mathring{\mathcal{K}}(\kappa_0)$ above that

$$\begin{aligned}&E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\,|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right\}\\&=E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\,|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right\}=\mathbf{0}\end{aligned}$$

for any spherically symmetric distribution. In addition, we also know that

$$E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right\}=\mathbf{0}.$$

Thus, even though $\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa_0)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\right]$ is the residual from the theoretical regression of $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})$ on a constant and $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$, it turns out that the second summand of (10) belongs to the restricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varsigma}_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$.

Now, if write (10) as

$$\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})-\mathbf{Z}_d(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})+\mathbf{Z}_d(\boldsymbol{\phi})\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the elliptically symmetric semiparametric efficient score is indeed unconditionally orthogonal to the restricted tangent set.

Finally, the expression for the semiparametric efficiency bound will be

$$E\left[\left.\begin{array}{c}\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})-\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\right\}\\ \times\left\{\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}_{dt}'(\boldsymbol{\theta})-\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\phi})\right\}\end{array}\right|\boldsymbol{\phi}\right]$$

$$=E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right]$$

$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\phi})|\boldsymbol{\phi}\right\}$$

$$-E\left\{\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}_{d}'(\boldsymbol{\phi})|\boldsymbol{\phi}\right\}$$

$$+E\left\{\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}\left(\kappa\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\phi})|\boldsymbol{\phi}\right\}$$

$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]-\frac{4}{N\left[(N+2)\kappa+2\right]}\right\}$$

by virtue of the law of iterated expectations. □

Proposition 8

The proof that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ is at least as large as $\mathcal{P}(\boldsymbol{\phi}_0)$ in the positive semidefinite matrix sense follows trivially from the fact that the latter is the residual variance in the multivariate theoretical regression of $\mathbf{s}_{\theta t}(\boldsymbol{\phi}_0)$ on $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, while the former is the unconditional variance of $\mathbf{s}_{\theta t}(\boldsymbol{\phi}_0)$. The fact that the residual variance of a multivariate regression cannot increase as we increase the number of regressors also explains why $\mathcal{P}(\boldsymbol{\phi}_0)$ is at least as large (in the positive semidefinite matrix sense) as $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$, and why the latter is at least as large as $\mathcal{S}(\boldsymbol{\phi}_0)$, reflecting the fact that the relevant tangent sets become increasing larger. Finally, the positive semidefiniteness of $\mathcal{S}(\boldsymbol{\phi}_0)-\mathcal{A}(\boldsymbol{\phi})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\phi})$ can be proved along the same lines as Result 2 in Gonzalez-Rivera and Drost (1999). □

Proposition 9

The proof of the first part is trivial, except perhaps for the fact that $\mathrm{M}_{sr}(\mathbf{0})=\mathbf{0}$, which follows from Proposition 3 because $\mathbf{e}_{st}(\boldsymbol{\theta}_0,\mathbf{0})$ coincides with $\mathbf{e}_{st}(\boldsymbol{\phi}_0)$ under normality.

To prove the second part, note that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathring{\mathcal{S}}(\boldsymbol{\phi})$ is $\mathbf{W}_d(\boldsymbol{\phi})\mathbf{W}_d'(\boldsymbol{\phi})$ times the residual variance in the theoretical regression of $-2\partial g(\varsigma_t,\boldsymbol{\eta})/\partial\varsigma\cdot(\varsigma_t/N)-1$ on $(\varsigma_t/N)-1$, which given that $\mathbf{W}_d(\boldsymbol{\phi})\neq\mathbf{0}$ can only be 0 if the regression residual is identically 0 for all t. The solution to the resulting differential equation is

$$g(\varsigma_t,\boldsymbol{\eta})=-\frac{N(N+2)\kappa}{2[(N+2)\kappa+2]}\ln\varsigma_t-\frac{1}{2[(N+2)\kappa+2]}\varsigma_t+C,$$

which in view of (A3) implies that

$$h(\varsigma_t;\boldsymbol{\eta})\propto\varsigma_t^{\frac{N}{(N+2)\kappa+2}-1}\exp\left\{-\frac{1}{2[(N+2)\kappa+2]}\varsigma_t\right\},$$

i.e. the density of Gamma random variable with mean N and variance $N[(N+2)\kappa_0+2]$.

Finally, to prove the third part we use the fact that after some tedious algebraic manipulations we can write $\mathcal{M}_{dd}(\boldsymbol{\eta})-\mathcal{K}(0)\mathcal{K}^{+}(\kappa)\mathcal{K}(0)$ as

$$\begin{bmatrix} (\mathrm{M}_{ll}(\boldsymbol{\eta})\text{-}1)\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \left[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}\frac{1}{\kappa+1}\right](\mathbf{I}_{N^2}+\mathbf{K}_{NN})+\left[\mathrm{M}_{ss}(\boldsymbol{\eta}_0)\text{-}1+\frac{2\kappa}{(\kappa+1)[(N+2)\kappa+2]}\right]vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}.$$

Therefore, given that $\mathbf{Z}_l(\boldsymbol{\phi}_0)\neq\mathbf{0}$, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathcal{S}(\boldsymbol{\phi})$ will be zero only if $\mathrm{M}_{ll}(\boldsymbol{\eta})=1$, which in turn requires that the residual variance in the multivariate regression of $-2\partial g(\varsigma_t,\boldsymbol{\eta})/\partial\varsigma\cdot\boldsymbol{\varepsilon}_t^*$ on $\boldsymbol{\varepsilon}_t^*$ is zero for all t, or equivalently, that $\partial g(\varsigma_t,\boldsymbol{\eta})/\partial\varsigma=-1/2$. But since the solution to this differential equation is $g(\varsigma_t,\boldsymbol{\eta})=-.5\varsigma_t+C$, then the result follows from (A3). □

Proposition 10

It is tedious but otherwise straightforward to prove that when $\boldsymbol{\alpha}_0=\mathbf{0}$

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)=\begin{bmatrix} \omega_0^{-1/2}(1-\sum_{j=1}^h\rho_{j0}) & 0 \\ \omega_0^{-1/2}(y_{t-1}-\pi_0,\ldots,y_{t-h}-\pi_0)' & \mathbf{0} \\ 0 & \frac{1}{2}\omega_0^{-1} \\ \mathbf{0} & \frac{1}{2}(\varepsilon_{t-1}^{*2}-1,\ldots,\varepsilon_{t-q}^{*2}-1)' \end{bmatrix},$$

so that

$$\mathbf{Z}_d(\boldsymbol{\phi}_0)=\begin{bmatrix} \omega_0^{-1/2}(1-\sum_{j=1}^h\rho_{j0}) & 0 \\ \mathbf{0} & \mathbf{0} \\ 0 & \frac{1}{2}\omega_0^{-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Proposition 1 then implies that the information matrix will have only four non-zero blocks along its diagonal, which correspond to π, $\boldsymbol{\rho}$, $\boldsymbol{\alpha}$ and $(\omega,\boldsymbol{\eta})$. The same proposition also implies that

$$\begin{aligned} \mathcal{I}_{\pi\pi}(\boldsymbol{\phi}_0) &= \mathrm{M}_{ll}(\boldsymbol{\eta}_0)\omega_0^{-1}(1-\sum_{j=1}^h\rho_{j0})^2, \\ \mathcal{I}_{\rho\rho}(\boldsymbol{\phi}_0) &= \mathrm{M}_{ll}(\boldsymbol{\eta}_0)\omega_0^{-1}\boldsymbol{\Sigma}_0, \end{aligned}$$

where $\boldsymbol{\Sigma}_0$ is the $h\times h$ autocovariance matrix of $(y_{t-1},\ldots,y_{t-h})'$, and

$$\begin{aligned} E\left[\mathcal{I}_{\alpha\alpha t}(\boldsymbol{\phi}_0)|\,\boldsymbol{\phi}_0\right] &= [3\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1]\cdot E\left[\frac{1}{4}(\varepsilon_{t-1}^{*2}-1,\ldots,\varepsilon_{t-q}^{*2}-1)'(\varepsilon_{t-1}^{*2}-1,\ldots,\varepsilon_{t-q}^{*2}-1)\right] \\ &= \frac{3\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1}{4}\cdot E[(\varepsilon_t^{*2}-1)^2|\boldsymbol{\eta}_0]\cdot\mathbf{I}_q=\frac{[3\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1](3\kappa_0+2)}{4}\mathbf{I}_q, \end{aligned}$$

in view of the fact that ε_t^* is serially independent when $\boldsymbol{\alpha}_0=\mathbf{0}$, and $E(\varepsilon_t^{*2}-1)^2=V(\varepsilon_t^{*2})=(3\kappa_0+2)$.

Given that $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ has a block-structure, the block-diagonality of $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ and $\mathcal{S}(\boldsymbol{\phi}_0)$ follows from expressions (11) and (9), respectively. Finally, we can use Proposition 1 in Demos and Sentana (1998) to show that $\mathcal{C}(\boldsymbol{\phi}_0)$ is also block-diagonal, with $\mathcal{C}_{\pi\pi}(\boldsymbol{\phi}_0)=\omega(1-\sum_{j=1}^h\rho_{j0})^{-2}$, $\mathcal{C}_{\boldsymbol{\rho\rho}}(\boldsymbol{\phi}_0)=\omega\boldsymbol{\Sigma}_0^{-1}$, and $\mathcal{C}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}_0)=\mathbf{I}_q$, although note that there is a missing scalar term in front of their expression for $\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}_0)$. □

Proposition 11

Using the results in appendix A.5 of Sentana and Fiorentini (2001), and appendices C and D in Sentana (2004), it is tedious but otherwise straightforward to prove that when $\boldsymbol{\alpha}_0 = \mathbf{0}$

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0) = \left\{ \begin{array}{cc} [\mathbf{I}_N - diag(\boldsymbol{\rho}_0)]\boldsymbol{\Sigma}_0^{-1/2\prime} & \mathbf{0} \\ diag[\mathbf{y}_{t-1} - \boldsymbol{\pi}]\boldsymbol{\Sigma}_0^{-1/2\prime} & \mathbf{0} \\ \mathbf{0} & (\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \boldsymbol{\Sigma}_0^{-1/2\prime}) \\ \mathbf{0} & \frac{1}{2}\mathbf{E}_N'(\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \boldsymbol{\Sigma}_0^{-1/2\prime}) \\ \mathbf{0} & \frac{1}{2}\left[\begin{array}{c} f_{kt-1}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \\ \vdots \\ f_{kt-q}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \end{array}\right](\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime}) \end{array} \right\}$$

where $\mathbf{E}_N'$ is the unique matrix that transforms $vec(\mathbf{A})$ in $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}_N' vec(\mathbf{A})$, and $\boldsymbol{\Sigma}$ is shorthand for $\mathbf{cc}' + \Gamma$. As a result,

$$\mathbf{Z}_d(\boldsymbol{\phi}_0) = \left[\begin{array}{cc} [\mathbf{I}_N - diag(\boldsymbol{\rho}_0)]\boldsymbol{\Sigma}_0^{-1/2\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \boldsymbol{\Sigma}_0^{-1/2\prime}) \\ \mathbf{0} & \frac{1}{2}\mathbf{E}_N'(\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \boldsymbol{\Sigma}_0^{-1/2\prime}) \\ \mathbf{0} & \mathbf{0} \end{array} \right],$$

where we have used the fact that $f_{kt}(\boldsymbol{\theta}_0) = \mathbf{c}_0'\Sigma_0^{-1}\boldsymbol{\varepsilon}_t$ and $\omega_t(\boldsymbol{\theta}_0) = 1 - \mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0 = (1 + \mathbf{c}_0'\Gamma_0^{-1}\mathbf{c}_0)^{-1} = \omega(\boldsymbol{\theta}_0)$ $\forall t$ when $\boldsymbol{\alpha}_0 = \mathbf{0}$ (see Sentana and Fiorentini (2001)), so that $E[f_{kt}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1|\boldsymbol{\phi}_0] = E(\mathbf{c}_0'\Sigma_0^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\Sigma_0^{-1}\mathbf{c}_0 - \mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0|\boldsymbol{\phi}_0) = \mathbf{0}$.

Proposition 1 then implies that the information matrix will have only four non-zero blocks along its diagonal, which correspond to $\boldsymbol{\pi}$, $\boldsymbol{\rho}$, $\boldsymbol{\alpha}$ and $(\mathbf{c}, \boldsymbol{\gamma}, \boldsymbol{\eta})$. The same proposition also implies that

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\pi\pi}}(\boldsymbol{\phi}_0) &= \mathrm{M}_{ll}(\boldsymbol{\eta}_0)[\mathbf{I}_N - diag(\boldsymbol{\rho}_0)]\boldsymbol{\Sigma}_0^{-1}[\mathbf{I}_N - diag(\boldsymbol{\rho}_0)], \\ \mathcal{I}_{\boldsymbol{\rho\rho}}(\boldsymbol{\phi}_0) &= \mathrm{M}_{ll}(\boldsymbol{\eta}_0)E[diag(\mathbf{y}_t - \boldsymbol{\pi}_0)\boldsymbol{\Sigma}_0^{-1}diag(\mathbf{y}_t - \boldsymbol{\pi}_0)|\boldsymbol{\phi}_0], \end{aligned}$$

and

$$\begin{aligned} E\left[\mathcal{I}_{\alpha\alpha t}(\boldsymbol{\phi}_0)|\,\boldsymbol{\phi}_0\right] = \frac{1}{4}E\Bigg\{ &\left(\begin{array}{c} f_{kt-1}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \\ \vdots \\ f_{kt-q}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \end{array}\right)(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime} \otimes \mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1/2\prime})\left[\mathrm{M}_{ss}(\boldsymbol{\eta}_0)\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right)\right. \\ &\left. + [\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\right](\boldsymbol{\Sigma}_0^{-1/2}\mathbf{c}_0 \otimes \boldsymbol{\Sigma}_0^{-1/2}\mathbf{c}_0)\left(\begin{array}{c} f_{kt-1}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \\ \vdots \\ f_{kt-q}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1 \end{array}\right)'\Bigg\} \\ = &\frac{3\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1}{4}(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)^2 E\{[f_{kt}^2(\boldsymbol{\theta}_0) + \omega(\boldsymbol{\theta}_0) - 1]^2|\boldsymbol{\phi}_0\}\mathbf{I}_q \end{aligned}$$

in view of the fact that f_{kt} is serially independent when $\boldsymbol{\alpha}_0 = 0$. In this respect, we can show that

$$E\{[f_{kt}^2(\boldsymbol{\theta}_0)+\omega(\boldsymbol{\theta}_0)-1]^2|\boldsymbol{\phi}_0\}=E(\mathbf{c}_0'\Sigma_0^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\Sigma_0^{-1}\mathbf{c}_0)^2-(\mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0)^2$$
$$=E[vec(\mathbf{c}_0'\Sigma_0^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\Sigma_0^{-1}\mathbf{c}_0)vec'(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)|\boldsymbol{\phi}_0]-(\mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0)^2$$
$$=(\mathbf{c}_0'\Sigma_0^{-1/2\prime}\otimes\mathbf{c}_0'\Sigma_0^{-1/2\prime})E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\boldsymbol{\phi}_0](\Sigma_0^{-1/2}\mathbf{c}_0\otimes\Sigma_0^{-1/2}\mathbf{c}_0)-(\mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0)^2$$
$$=(\mathbf{c}_0'\Sigma_0^{-1/2\prime}\otimes\mathbf{c}_0'\Sigma_0^{-1/2\prime})(\kappa_0+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)](\Sigma_0^{-1/2}\mathbf{c}_0\otimes\Sigma_0^{-1/2}\mathbf{c}_0)$$
$$-(\mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0)^2=(3\kappa_0+2)(\mathbf{c}_0'\Sigma_0^{-1}\mathbf{c}_0)^2=(3\kappa_0+2)(\mathbf{c}_0'\Gamma_0^{-1}\mathbf{c}_0)^2(1+\mathbf{c}_0'\Gamma_0^{-1}\mathbf{c}_0)^{-2}.$$

Given that $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ has a block-structure, the block-diagonality of $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ and $\mathcal{S}(\boldsymbol{\phi}_0)$ follows from expressions (11) and (9), respectively. Finally, it follows directly from Proposition 6 that $\mathcal{C}(\boldsymbol{\phi}_0)$ will also be block-diagonal, with

$$\begin{aligned}
\mathcal{A}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\phi}_0) &= \mathcal{B}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\phi}_0)=\mathcal{C}_{\boldsymbol{\pi}\boldsymbol{\pi}}^{-1}(\boldsymbol{\phi}_0)=[\mathbf{I}_N-diag(\boldsymbol{\rho}_0)]\boldsymbol{\Sigma}_0^{-1}[\mathbf{I}_N-diag(\boldsymbol{\rho}_0)],\\
\mathcal{A}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\phi}_0) &= \mathcal{B}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\phi}_0)=\mathcal{C}_{\boldsymbol{\rho}\boldsymbol{\rho}}^{-1}(\boldsymbol{\phi}_0)=E[diag(\mathbf{y}_t-\boldsymbol{\pi}_0)\boldsymbol{\Sigma}_0^{-1}diag(\mathbf{y}_t-\boldsymbol{\pi}_0)|\boldsymbol{\phi}_0],\\
\mathcal{A}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\phi}_0) &= .5(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)^2E\{[f_{kt}^2(\boldsymbol{\theta}_0)+\omega(\boldsymbol{\theta}_0)-1]^2|\boldsymbol{\phi}_0\}\mathbf{I}_q,\\
\mathcal{B}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\phi}_0) &= .25(3\kappa_0+2)(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)^2E\{[f_{kt}^2(\boldsymbol{\theta}_0)+\omega(\boldsymbol{\theta}_0)-1]^2|\boldsymbol{\phi}_0\}\mathbf{I}_q,
\end{aligned}$$

so that

$$\mathcal{C}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1}(\boldsymbol{\phi}_0)=\frac{(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)^2}{3\kappa_0+2}E\{[f_{kt}^2(\boldsymbol{\theta}_0)+\omega(\boldsymbol{\theta}_0)-1]^2|\boldsymbol{\phi}_0\}\mathbf{I}_q=(\mathbf{c}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{c}_0)^4=\frac{(\mathbf{c}_0'\Gamma_0^{-1}\mathbf{c}_0)^4}{(1+\mathbf{c}_0'\Gamma_0^{-1}\mathbf{c}_0)^4}.$$

□

Proposition 12

Given our assumptions on the mapping $\mathbf{r}_s(.)$, we can directly work in terms of the $\boldsymbol{\vartheta}$ parameters. In this sense, since the conditional covariance matrix of $\mathbf{y}_t$ is of the form $\vartheta_2\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)$, it is straightforward to show that

$$\mathbf{Z}_{dt}(\boldsymbol{\vartheta})=\left\{\begin{array}{cc}\vartheta_2^{-1/2}\,[\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_1)/\partial\boldsymbol{\vartheta}_1]\,\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1) & \frac{1}{2}\{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_1)]/\partial\boldsymbol{\vartheta}_1\}[\Sigma_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)\otimes\Sigma_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)]\\ 0 & \frac{1}{2}\vartheta_2^{-1}vec'(\mathbf{I}_N)\end{array}\right\}=\left[\begin{array}{cc}\mathbf{Z}_{\vartheta_1 lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\vartheta_1 st}(\boldsymbol{\vartheta})\\ 0 & \mathbf{Z}_{\vartheta_2 st}(\boldsymbol{\vartheta})\end{array}\right].\qquad\text{(A7)}$$

Thus, the elliptically symmetric score vector for $\boldsymbol{\vartheta}$ will be

$$\left[\begin{array}{c}\mathbf{s}_{\vartheta_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\ s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right]=\left[\begin{array}{c}\mathbf{Z}_{\vartheta_1 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})+\mathbf{Z}_{\vartheta_1 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\ \mathbf{Z}_{\vartheta_2 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right],\qquad\text{(A8)}$$

where $\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ are given in (2) and (3), respectively.

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\vartheta_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is

$$
\begin{aligned}
& E\left\{\left[\begin{array}{cc}\mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})\end{array}\right]\left[\begin{array}{cc}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0}\\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{array}\right]\left[\begin{array}{c}0\\ \mathbf{Z}'_{\vartheta_2 st}(\boldsymbol{\vartheta})\end{array}\right]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\Bigg\}\\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_2}E\left\{\frac{1}{2}\frac{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_1)]}{\partial\boldsymbol{\vartheta}_1}[\Sigma_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)\otimes\Sigma_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\Bigg\}vec(\mathbf{I}_N)\\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_2}\mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta},\boldsymbol{\eta})vec(\mathbf{I}_N),
\end{aligned}
$$

with $\mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta},\boldsymbol{\eta})=E[\mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})|\boldsymbol{\vartheta},\boldsymbol{\eta}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$, as well as the law of iterated expectations, together with the results in Proposition 1.

We can use the same arguments to show that the unconditional variance of $s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be given by

$$
\begin{aligned}
& E\left\{\left[\begin{array}{cc}0 & \mathbf{Z}_{\vartheta_2 st}(\boldsymbol{\vartheta})\end{array}\right]\left[\begin{array}{cc}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0}\\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{array}\right]\left[\begin{array}{c}0\\ \mathbf{Z}'_{\vartheta_2 st}(\boldsymbol{\vartheta})\end{array}\right]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\Bigg\}\\
= & \frac{1}{4\vartheta_2^2}vec'(\mathbf{I}_N)[\mathrm{M}_{ss}(\boldsymbol{\eta})\,(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+[\mathrm{M}_{ss}(\boldsymbol{\eta})-1])vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]vec(\mathbf{I}_N)\\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}N}{4\vartheta_2^2}.
\end{aligned}
$$

Hence, the residuals from the unconditional regression of $\mathbf{s}_{\vartheta_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ on $s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be:

$$
\begin{aligned}
\mathbf{s}_{\vartheta_1|\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) &= \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})+\mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\
-\frac{4\vartheta_2^2}{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}1]\}N} & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}1]\}}{2\vartheta_2}\mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta})vec(\mathbf{I}_N)\frac{1}{2\vartheta_2}vec'(\mathbf{I}_N)\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\
&= \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})+[\mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})-\mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta},\boldsymbol{\eta})]\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}).
\end{aligned}
$$

The first term of $\mathbf{s}_{\vartheta_1|\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is clearly conditionally orthogonal to any function of $\varsigma_t(\boldsymbol{\vartheta})$. In contrast, the second term is not conditionally orthogonal to functions of $\varsigma_t(\boldsymbol{\vartheta})$, but since the conditional covariance between any such function and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be time-invariant, it will be unconditionally orthogonal by the law of iterated expectations. As a result, $\mathbf{s}_{\vartheta_1|\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be unconditionally orthogonal to the elliptically symmetric tangent set, which in turn implies that the elliptically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_1$ will be ϑ_2-adaptive. □

Proposition 13

Given our assumptions on the mapping $\mathbf{r}_g(.)$, we can directly work in terms of the $\boldsymbol{\delta}$ parameters. Given the specification for the conditional mean and variance in (15), and the fact that $\boldsymbol{\varepsilon}_t^*$ is assumed to be *i.i.d.* conditional on $\mathbf{z}_t$ and I_{t-1}, it is tedious but otherwise straightforward to show that the score vector will be

$$
\left[\begin{array}{c}\mathbf{s}_{\delta_1 t}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{s}_{\delta_2 t}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{s}_{\delta_3 t}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{array}\right]=\left[\begin{array}{c}\mathbf{Z}_{\delta_1 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})+\mathbf{Z}_{\delta_1 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{Z}_{\delta_2 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{Z}_{\delta_3 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{array}\right], \tag{A9}
$$

where

$$\left.\begin{array}{l}\mathbf{Z}_{\delta_1 lt}(\boldsymbol{\delta})=\left\{\partial\boldsymbol{\mu}_t^{*\prime}(\boldsymbol{\delta}_1)/\partial\boldsymbol{\delta}_1+\partial vec'[\boldsymbol{\Sigma}_t^{*1/2}(\boldsymbol{\delta}_1)]/\partial\boldsymbol{\delta}_1\cdot(\boldsymbol{\delta}_3\otimes\mathbf{I}_N)\right\}\boldsymbol{\Sigma}_t^{*-1/2\prime}(\boldsymbol{\delta}_1)\boldsymbol{\Delta}_2^{-1/2\prime},\\ \mathbf{Z}_{\delta_1 st}(\boldsymbol{\delta})=\partial vec'[\boldsymbol{\Sigma}_t^{*1/2}(\boldsymbol{\delta}_1)]/\partial\boldsymbol{\delta}_1\cdot[\boldsymbol{\Delta}_2^{1/2}\otimes\boldsymbol{\Sigma}_t^{*-1/2\prime}(\boldsymbol{\delta}_1)\boldsymbol{\Delta}_2^{-1/2\prime}],\\ \mathbf{Z}_{\delta_2 st}(\boldsymbol{\delta})=\mathbf{D}_N'(\mathbf{I}_N\otimes\boldsymbol{\Delta}_2^{-1/2\prime})=\mathbf{Z}_{\delta_2 s}(\boldsymbol{\delta}),\\ \mathbf{Z}_{\delta_3 lt}(\boldsymbol{\delta})=\boldsymbol{\Delta}_2^{-1/2\prime}=\mathbf{Z}_{\delta_3 l}(\boldsymbol{\delta}),\end{array}\right\}\quad\text{(A10)}$$

$\mathbf{D}_N$ is the duplication matrix of order N (see Magnus and Neudecker (1988)),

$$\begin{aligned}\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho}) &= -\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*},\\ \mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho}) &= -vec\left\{\mathbf{I}_N+\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\delta})\right\},\end{aligned}$$

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta})=\boldsymbol{\Delta}_2^{-1/2}\boldsymbol{\Sigma}_t^{*-1/2}(\boldsymbol{\delta}_1)[\mathbf{y}_t-\boldsymbol{\mu}_t^*(\boldsymbol{\delta}_1)-\boldsymbol{\Sigma}_t^{*1/2}\boldsymbol{\delta}_3],\quad\text{(A11)}$$

and $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ is the conditional density of $\boldsymbol{\varepsilon}_t^*$ given $\mathbf{z}_t$, I_{t-1} and the shape parameters $\boldsymbol{\varrho}$.

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\delta_1 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ and the remaining elements of the score will be given by

$$\begin{bmatrix}\mathbf{Z}_{\delta_1 l}(\boldsymbol{\delta},\boldsymbol{\varrho}) & \mathbf{Z}_{\delta_1 s}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\delta_3 l}'(\boldsymbol{\delta})\\ \mathbf{Z}_{\delta_2 s}'(\boldsymbol{\delta}) & \mathbf{0}\end{bmatrix}$$

with $\mathbf{Z}_{\delta_1 l}(\boldsymbol{\delta},\boldsymbol{\varrho})=E[\mathbf{Z}_{\delta_1 lt}(\boldsymbol{\delta})|\boldsymbol{\delta},\boldsymbol{\varrho}]$ and $\mathbf{Z}_{\delta_1 s}(\boldsymbol{\delta},\boldsymbol{\varrho})=E[\mathbf{Z}_{\delta_1 st}(\boldsymbol{\delta})|\boldsymbol{\delta},\boldsymbol{\varrho}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$ and the constancy of $\mathbf{Z}_{\delta_2 st}(\boldsymbol{\delta})$ and $\mathbf{Z}_{\delta_3 lt}(\boldsymbol{\delta})$, together with the law of iterated expectations and the definition

$$\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}=V\left[\left.\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{array}\right|\boldsymbol{\delta},\boldsymbol{\varrho}\right].$$

Similarly, the unconditional covariance matrix of $\mathbf{s}_{\delta_2 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ and $\mathbf{s}_{\delta_3 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ will be

$$\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\delta_2 s}(\boldsymbol{\delta})\\ \mathbf{Z}_{\delta_3 l}(\boldsymbol{\delta}) & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\delta_3 l}'(\boldsymbol{\delta})\\ \mathbf{Z}_{\delta_2 s}'(\boldsymbol{\delta}) & \mathbf{0}\end{bmatrix}.$$

Hence, the residuals from the unconditional least squares projection of $\mathbf{s}_{\delta_1 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ on $\mathbf{s}_{\delta_2 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ and $\mathbf{s}_{\delta_3 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ will be:

$$\begin{aligned}\mathbf{s}_{\delta_1|\delta_2,\delta_3 t}(\boldsymbol{\delta},\boldsymbol{\varrho}) &= \mathbf{Z}_{\delta_1 lt}(\boldsymbol{\delta})\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})+\mathbf{Z}_{\delta_1 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ &\quad-\begin{bmatrix}\mathbf{Z}_{\delta_1 l}(\boldsymbol{\delta},\boldsymbol{\varrho}) & \mathbf{Z}_{\delta_1 s}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})\\ \mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})\end{bmatrix}\\ &= [\mathbf{Z}_{\delta_1 lt}(\boldsymbol{\delta})-\mathbf{Z}_{\delta_1 l}(\boldsymbol{\delta},\boldsymbol{\varrho})]\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})+[\mathbf{Z}_{\delta_1 st}(\boldsymbol{\delta})-\mathbf{Z}_{\delta_1 s}(\boldsymbol{\delta},\boldsymbol{\varrho})]\mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho}),\end{aligned}$$

because both $\mathbf{Z}_{\delta_2 s}(\boldsymbol{\delta})$ and $\mathbf{Z}_{\delta_3 l}(\boldsymbol{\delta})$ have full row rank when $\boldsymbol{\Delta}_2$ has full rank.

Although neither $\mathbf{e}_{lt}(\boldsymbol{\delta},\boldsymbol{\varrho})$ nor $\mathbf{e}_{st}(\boldsymbol{\delta},\boldsymbol{\varrho})$ will be conditionally orthogonal to arbitrary functions of $\boldsymbol{\varepsilon}_t^*$, their conditional covariance with any such function will be time-invariant. Hence, $\mathbf{s}_{\delta_1|\delta_2,\delta_3 t}(\boldsymbol{\delta},\boldsymbol{\varrho})$ will be unconditionally orthogonal to $\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}_i$ by virtue of the law of iterated expectations, which in turn implies that the unrestricted semiparametric estimator of $\boldsymbol{\delta}_1$ will be $(\boldsymbol{\delta}_2,\boldsymbol{\delta}_3)$-adaptive. □

Proposition 14

The proof of the first part trivially follows from Proposition 8 and the fact that the partitioned inverse formula implies that

$$\mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)=\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0)+\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\eta}'(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0).$$

To prove that $\mathcal{F}(\boldsymbol{\phi}_0)\leq\mathcal{H}(\boldsymbol{\phi}_0)$ it is convenient to note that both these matrices can also be decomposed into a component that reflects the asymptotic variance of these estimators if $\boldsymbol{\theta}_0$ were known, plus a second component that reflects the sample variability in the PML estimator $\tilde{\boldsymbol{\theta}}_T$. With respect to the first component, it is clear that $\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0)\leq\mathcal{L}(\boldsymbol{\phi}_0)/\mathcal{N}^2(\boldsymbol{\phi}_0)$. As for the second component, we must compare

$$\mathcal{I}_{\boldsymbol{\theta}\eta}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)/\mathcal{I}_{\eta\eta}^2(\boldsymbol{\phi}_0)=\left[\frac{2\,(N+2)\,\nu^2}{(\nu-2)\,(N+\nu)\,(N+\nu+2)\,\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)}\right]^2\mathbf{W}_s'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathbf{W}_s(\boldsymbol{\phi}_0)$$

with

$$\mathcal{Q}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{Q}(\boldsymbol{\phi}_0)/\mathcal{N}^2(\boldsymbol{\phi}_0)=\left[\frac{4(\nu-2)\,(\nu-4)}{N\nu^2\,(\nu-6)}\right]^2\mathbf{W}_s'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathbf{W}_s(\boldsymbol{\phi}_0).$$

The second expression will be larger than the first one if and only if

$$\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)-\frac{(N+2)\,N\nu^4\,(\nu-6)}{2\,(\nu-2)^2\,(\nu-4)\,(N+\nu)\,(N+\nu+2)}\geq 0.$$

We can then show that this inequality will be true for $N+2$ if it is true for N by using the recursion $\psi'(\nu/2)-\psi'(1+\nu/2)=-4\nu^2$ (see Abramowitz and Stegun (1964)), which reduces the problem to proving the inequality for $N=1$ and $N=2$. The proof for $N=2$ immediately follows from the same recursion. The proof for $N=1$ is more tedious, as it involves the asymptotic expressions for $\psi'(.)$ in Abramowitz and Stegun (1964).

To prove the last statement, it is also convenient to decompose the asymptotic variance of $\breve{\eta}_T$ into two components, namely:

$$\mathcal{G}(\boldsymbol{\phi}_0)=[\mathcal{E}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}(\boldsymbol{\phi}_0)]/\mathcal{N}^2(\boldsymbol{\phi}_0)$$
$$+\{[\mathcal{R}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]'\mathcal{C}(\boldsymbol{\phi}_0)[\mathcal{R}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]\}/\mathcal{N}^2(\boldsymbol{\phi}_0)$$

In this set up, it is straightforward to prove that

$$[\mathcal{R}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]=\mathcal{Q}(\boldsymbol{\phi}_0)$$

if condition (16) holds. As for the first component, since $\mathcal{L}(\boldsymbol{\phi}_0)$ is the residual variance in the regression of $m_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ on $\varsigma_t/N-1$, while $\mathcal{E}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}(\boldsymbol{\phi}_0)$ is the residual variance in the regression of $m_{\eta t}(\boldsymbol{\theta}_0,\eta_0)$ on $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)$, and the Gaussian pseudo-score can be written as $\mathbf{W}_s(\boldsymbol{\phi}_0)[\varsigma_t/N-1]$ plus an extra term that is orthogonal to ς_t, it is clear that

$$\mathcal{L}(\boldsymbol{\phi}_0)\leq\mathcal{E}(\boldsymbol{\phi}_0)-\mathcal{D}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}(\boldsymbol{\phi}_0),$$

with equality if and only if $[\varsigma_t/N-1]$ can be written as an exact linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)$, as in (17). $\square$

Proposition 15

The consistency of the Gaussian PML derives from the fact that $E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]=\mathbf{0}$. Thus, if the pseudo-true value of η, η_∞ say, is 0, then the student-t based pseudo-true values of the conditional mean and variance parameters, $\boldsymbol{\theta}_\infty$ say, will coincide with their true values $\boldsymbol{\theta}_0$ by the law of iterated expectations. But since η is estimated subject to the inequality constraint $\eta\geq 0$, the population KT conditions that define η_∞ will be

$$E[s_{\eta t}(\boldsymbol{\theta}_\infty,\eta_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]+\lambda_{\eta\infty}=0;\quad \eta_\infty\geq 0;\quad \lambda_{\eta\infty}\geq 0;\quad \eta_\infty\cdot\lambda_{\eta\infty}=0,$$

where $\lambda_{\eta\infty}$ is the pseudo-true value of the KT multiplier, and the expectation is taken with respect to the true unconditional distribution of the observations (see Calzolari, Fiorentini and Sentana (2004)). Hence, $\eta_\infty=0$ if and only if $E[s_{\eta t}(\boldsymbol{\theta}_0,0)|\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]\leq 0$.

Given that $\varsigma_t(\boldsymbol{\theta}_0)=\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$, we can write

$$\begin{aligned}s_{\eta t}(\boldsymbol{\theta}_0,0) &= \frac{N(N+2)}{4}-\frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0)+\frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0)\\ &= \frac{N(N+2)}{4}\left[\frac{(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^*)^2}{N(N+2)}-1\right]+\frac{N+2}{2}[(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^*)-N].\end{aligned}$$

But since we have normalised the innovations so that $E(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*}\boldsymbol{\varepsilon}_t^{*\prime}|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)=\mathbf{I}_N$, then

$$N=tr(\mathbf{I}_N)=tr[E(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*}\boldsymbol{\varepsilon}_t^{*\prime}|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)]=E[tr(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*}\boldsymbol{\varepsilon}_t^{*\prime})|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]=E(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^{*}|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$$

by the linearity of the expectation and trace operators. Therefore, it immediately follows that

$$\lambda_{\eta\infty}=\min\{0,-E[s_{\eta t}(\boldsymbol{\theta}_0,0)|\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]\}=\min\left\{0,-\frac{N(N+2)}{4}\kappa_0\right\}$$

in view of the definition of κ_0. Therefore, $\eta_\infty=0$ if and only if $\kappa_0\leq 0$.

To prove the second and third parts, we can use Propositions 1 and 2 in Calzolari, Fiorentini and Sentana (2004) if we regard the student t based estimator $\hat{\boldsymbol{\phi}}_T$ as the "inequality restricted" PML estimator of $\boldsymbol{\phi}$, and the Gaussian-based estimator $\tilde{\boldsymbol{\phi}}_T=(\tilde{\boldsymbol{\theta}}_T,0)$ as its "equality restricted" counterpart, both of which share not only the pseudo-true values $(\boldsymbol{\theta}_0,0,\lambda_{\eta\infty})$ when $\kappa_0\leq 0$, but also the modified pseudo-score $\mathbf{m}_t(\boldsymbol{\theta}_0,0,\lambda_{\eta\infty})=s_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_0,0)+\mathbf{e}_{p+1}\cdot\lambda_{\eta\infty}$, where $\mathbf{e}_{p+1}$ is the $(p+1)^{th}$ column of $\mathbf{I}_{p+1}$, as well as the expected value of the average Hessian $\mathcal{J}_0=E[\bar{\mathbf{h}}_T(\boldsymbol{\phi})|\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]$.

Specifically, Proposition 1 in Calzolari, Fiorentini and Sentana (2004) implies here that

$$\lambda_{\eta\infty}\cdot\sqrt{T}\hat{\eta}_T=o_p(1),$$

while their Proposition 2 implies that

$$\begin{array}{rcl}\left(\begin{array}{cc}\mathcal{J}_{\boldsymbol{\theta\theta}0} & \mathcal{J}_{\boldsymbol{\theta}\eta 0}\\ \mathcal{J}'_{\boldsymbol{\theta}\eta 0} & \mathcal{J}_{\eta\eta 0}\end{array}\right)\sqrt{T}\left(\begin{array}{c}\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \hat{\eta}_T\end{array}\right)+\mathbf{e}_{p+1}\sqrt{T}(\hat{\lambda}_{\eta T}-\lambda_{\eta\infty})-\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0,0,\lambda_{\eta\infty}) & = & o_p(1),\\ \left(\begin{array}{cc}\mathcal{J}_{\boldsymbol{\theta\theta}0} & \mathcal{J}_{\boldsymbol{\theta}\eta 0}\\ \mathcal{J}'_{\boldsymbol{\theta}\eta 0} & \mathcal{J}_{\eta\eta 0}\end{array}\right)\sqrt{T}\left(\begin{array}{c}\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ 0\end{array}\right)+\mathbf{e}_{p+1}\sqrt{T}(\tilde{\lambda}_{\eta T}-\lambda_{\eta\infty})-\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0,0,\lambda_{\eta\infty}) & = & o_p(1),\end{array}$$

where $\hat{\lambda}_{\eta T}$ and $\tilde{\lambda}_{\eta T}$ are the sample versions of the KT and Lagrange multipliers associated to the constraint $\eta = 0$. As a consequence,

$$\left(\begin{array}{cc}\mathcal{J}_{\boldsymbol{\theta\theta}0} & \mathcal{J}_{\boldsymbol{\theta}\eta 0}\\ \mathcal{J}'_{\boldsymbol{\theta}\eta 0} & \mathcal{J}_{\eta\eta 0}\end{array}\right)\sqrt{T}\left(\begin{array}{c}\hat{\boldsymbol{\theta}}_T-\tilde{\boldsymbol{\theta}}_T\\ \hat{\eta}_T\end{array}\right)+\mathbf{e}_{p+1}\sqrt{T}(\hat{\lambda}_{\eta T}-\tilde{\lambda}_{\eta T})=o_p(1).$$

Part 2 immediately follows from the fact that $\lambda_{\eta\infty} > 0$ when $\kappa_0 < 0$. Similarly, the first statement of Part 3 follows from the fact that $\lambda_{\eta\infty} = 0$ when $\kappa_0 = 0$. As for the condition (18), which derives directly from the expression for $\mathbf{h}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi})$ in Fiorentini, Sentana and Calzolari (2003) evaluated at $(\boldsymbol{\theta}_0, 0)$, its role is to guarantee that $\mathcal{J}_{\boldsymbol{\theta}\eta 0} = \mathbf{0}$. In this sense, it is worth mentioning that condition (18) will be satisfied if $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 = 0$ irrespective of whether or not it is Gaussian because in that case

$$E[(N+2-\varsigma_t)\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0] = E[(N+2-\varsigma_t)\sqrt{\varsigma_t}\mathbf{u}_t|\boldsymbol{\eta}_0] = \mathbf{0}$$

by the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t) = \mathbf{0}$, while

$$E[(N+2-\varsigma_t)\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0] = E[(N+2-\varsigma_t)\varsigma_t\mathbf{u}_t\mathbf{u}_t'|\boldsymbol{\eta}_0] = N^{-1}E[(N+2-\varsigma_t)\varsigma_t|\boldsymbol{\eta}_0]\mathbf{I}_N = \mathbf{0}$$

by the definition of κ_0 and the fact that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$. □

Proposition 16

As in the proof of Proposition 12, we can directly work in terms of the $\boldsymbol{\vartheta}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_s(.)$. For the sake of clarity, let us initially assume that we keep η fixed to some positive value. The student t score vector for the remaining parameters will then be given by (A8), where

$$\mathbf{e}_{lt}(\boldsymbol{\vartheta}, \eta) = \frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\vartheta})}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}), \tag{A12}$$

$$\mathbf{e}_{st}(\boldsymbol{\vartheta},\eta) = vec\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\vartheta})}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta}) - \mathbf{I}_N\right]. \tag{A13}$$

We can immediately see that

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}_{10}, \vartheta_2) = \frac{1}{\sqrt{\vartheta_2}}\boldsymbol{\Sigma}_t^{\circ-1/2}(\boldsymbol{\vartheta}_{10})[\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_{10})] = \sqrt{\frac{\vartheta_{20}}{\vartheta_2}}\boldsymbol{\varepsilon}_t^*,$$

so that

$$\varsigma_t(\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}) = \frac{\vartheta_{20}}{\vartheta_2}\varsigma_t.$$

Hence,

$$\begin{aligned}
\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{10}, \vartheta_2, \eta) &= \frac{N\eta+1}{1-2\eta+\eta(\vartheta_{20}/\vartheta_2)\varsigma_t}\sqrt{\frac{\vartheta_{20}}{\vartheta_2}}\boldsymbol{\varepsilon}_t^* \\
&= \frac{N\eta+1}{1-2\eta+\eta(\vartheta_{20}/\vartheta_2)\varsigma_t}\sqrt{\frac{\vartheta_{20}}{\vartheta_2}}\sqrt{\varsigma_t}\mathbf{u}_t, \\
\mathbf{e}_{st}(\boldsymbol{\vartheta}_{10}, \vartheta_2, \eta) &= vec\left[\frac{N\eta+1}{1-2\eta+\eta(\vartheta_{20}/\vartheta_2)\varsigma_t}\frac{\vartheta_{20}}{\vartheta_2}\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}-\mathbf{I}_N\right] \\
&= vec\left[\frac{N\eta+1}{1-2\eta+\eta(\vartheta_{20}/\vartheta_2)\varsigma_t}\frac{\vartheta_{20}}{\vartheta_2}\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right].
\end{aligned}$$

Then, it follows that $E[\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{10}, \vartheta_2, \eta)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0] = \mathbf{0}$ regardless of ϑ_2 and η because of the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t) = \mathbf{0}$. On the other hand,

$$E[\mathbf{e}_{st}(\boldsymbol{\vartheta}_{10}, \vartheta_2, \eta)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0] = E\left[\frac{N\eta+1}{1-2\eta+\eta(\vartheta_{20}/\vartheta_2)\varsigma_t}\frac{\vartheta_{20}}{\vartheta_2}\frac{\varsigma_t}{N}-1\middle|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\right]vec(\mathbf{I}_N)$$

because of the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$.

If we define $\vartheta_{2\infty}(\eta)$ as the value that solves the implicit equation

$$E\left[\frac{N\eta+1}{1-2\eta+\eta[\vartheta_{20}/\vartheta_{2\infty}(\eta)]\varsigma_t}\frac{\vartheta_{20}}{\vartheta_{2\infty}(\eta)}\frac{\varsigma_t}{N}-1\middle|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\right] = 0,$$

then it is straightforward to show that

$$E\{s_{\vartheta t}[\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\eta), \eta]|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\} = \mathbf{0}.$$

Finally, if we choose η_∞ as the solution to the implicit equation

$$E\{s_{\eta t}[\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\eta), \eta]|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\} = 0,$$

then it is clear that $\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\eta_\infty)$ and η_∞ will be the pseudo-true values of the parameters. □

Proposition 17

As in the proof of Proposition 13, we can directly work in terms of the $\boldsymbol{\delta}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. For the sake of clarity, let us initially assume that we keep η fixed to some positive value. The student t score vector for the remaining parameters will then be given by (A9), where the elements of $\mathbf{Z}_t$ are defined in (A10), $\mathbf{e}_{lt}(\boldsymbol{\delta}, \eta)$ and $\mathbf{e}_{st}(\boldsymbol{\delta}, \eta)$ are given by (A12) and (A13), respectively, and $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta})$ is defined in (A11).

We can immediately see from (A11) that

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) = \boldsymbol{\Delta}_2^{-1}(\boldsymbol{\delta}_{30} - \boldsymbol{\delta}_3) + \boldsymbol{\Delta}_2^{-1}\boldsymbol{\Delta}_{20}\boldsymbol{\varepsilon}_t^*,$$

so that both this variable and $\varsigma_t(\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3)$ will be *i.i.d.* conditional on $\mathbf{z}_t$ and I_{t-1}. Let $\boldsymbol{\delta}_{2\infty}(\eta)$ and $\boldsymbol{\delta}_{3\infty}(\eta)$ solve the implicit system of $N + N(N+1)/2$ equations

$$\begin{aligned} E\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]}\boldsymbol{\varepsilon}_t^*[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]\right|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\Bigg] &= \mathbf{0}, \\ vech\left\{E\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]}\boldsymbol{\varepsilon}_t^*[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]\right|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\Bigg]\right\} &= \mathbf{0}. \end{aligned}$$

Then, it is follows that

$$E\{s_{\boldsymbol{\delta}t}[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\} = \mathbf{0},$$

where we have exploited the symmetry of the matrix

$$\frac{N\eta+1}{1-2\eta+\eta\varsigma_t[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]}\boldsymbol{\varepsilon}_t^*[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta) \cdot \boldsymbol{\varepsilon}_t^{*\prime}[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]].$$

Finally, if we choose η_∞ as the solution to the implicit equation

$$E\{s_{\eta t}[\boldsymbol{\delta}_{10}, \boldsymbol{\delta}_{2\infty}(\eta), \boldsymbol{\delta}_{3\infty}(\eta)]|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0\} = 0,$$

then it is clear that $\boldsymbol{\delta}_{10}$, $\boldsymbol{\delta}_{2\infty}(\eta_\infty)$, $\boldsymbol{\delta}_{3\infty}(\eta_\infty)$ and η_∞ will be the pseudo-true values of the parameters. □

Proposition 18

Let $\boldsymbol{\phi}_\infty$ denote the pseudo-true values of $\boldsymbol{\phi}$ corresponding to the student t-based log-likelihood function, which can be implicitly characterised by the moment conditions

$$\begin{aligned} E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_\infty, \eta_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0] = \mathbf{0}, \\ E[s_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_\infty, \eta_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0] = \mathbf{0}. \end{aligned} \tag{A15}$$

The score version of the Hausman test can be regarded as an unconditional moment test of

$$E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_\infty, 0)|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0] = \mathbf{0}, \tag{A16}$$

which will hold if the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is *i.i.d.* $t(\mathbf{0}, \mathbf{I}, \eta_0)$ because $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ in that case. If we knew $\boldsymbol{\theta}_\infty$, it would be straightforward to test whether (A16) holds. But since we do not know $\boldsymbol{\theta}_\infty$, we replace it by its consistent estimator $\hat{\boldsymbol{\theta}}_T$, where $\hat{\boldsymbol{\theta}}_T$ and $\hat{\eta}_T$ satisfy the sample analogues of (A15). In order to account for the sampling variability that this introduces, we can compute the limiting unconditional least squares regression of $\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_\infty, 0)$ on $\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_\infty, \eta_\infty)$ and $\sqrt{T}\bar{s}_{\boldsymbol{\eta}T}(\boldsymbol{\theta}_\infty, \eta_\infty)$, and retain the residuals. But since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \eta_0)$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \eta_0)$ are martingale difference sequences under the null, we can simply regress the first on the last two. To do so, we need their joint asymptotic distribution, which in view of Propositions 1, 2 and 3 will be given by

$$\sqrt{T}\left[\begin{array}{c}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0,0)\\ \bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0,\eta_0)\\ \bar{s}_{\eta T}(\boldsymbol{\theta}_0,\eta_0)\end{array}\right]\xrightarrow{d}N\left\{\left(\begin{array}{c}\mathbf{0}\\ \mathbf{0}\\ 0\end{array}\right),\left[\begin{array}{ccc}\mathcal{B}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathbf{0}\\ \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\\ \mathbf{0}' & \mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0) & \mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)\end{array}\right]\right\}.$$

Hence, we can use standard arguments to show that

$$\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T,0)\xrightarrow{d}N[\mathbf{0},\mathcal{B}(\boldsymbol{\phi}_0)-\mathcal{A}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]$$

and

$$\sqrt{T}\left[\begin{array}{c}\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\end{array}\right]\xrightarrow{d}N\left\{\left(\begin{array}{c}\mathbf{0}\\ \mathbf{0}\end{array}\right),\left[\begin{array}{cc}\mathcal{C}(\boldsymbol{\phi}_0) & -\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\\ -\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\end{array}\right]\right\},$$

whence we can easily prove that

$$\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T,0)-\mathcal{A}(\boldsymbol{\phi}_0)\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T)=o_p(1),$$
$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T)\rightarrow N\left[\mathbf{0},\mathcal{C}(\boldsymbol{\phi}_0)-\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right],$$

as well as the asymptotic chi-square distribution of $H^W_{\boldsymbol{\theta}T}$. □

Proposition 19

The proof proceeds along the same lines of the previous one once we show that

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=-\frac{\partial E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]}{\partial\boldsymbol{\theta}}\tag{A17}$$

and

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]=-\frac{\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]}{\partial\boldsymbol{\theta}},\tag{A18}$$

where $\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ is shorthand for the symmetric semiparametric efficient score in (10). Condition (A17) is fairly straightforward because the regression residual

$$-\left\{\frac{2\partial g(\varsigma_t,\boldsymbol{\eta})}{\partial\varsigma}\frac{\varsigma_t}{N}+1\right\}-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)$$

is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$ by the law of iterated expectations, as shown in the proof of proposition 7. As a result,

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}]=\mathcal{A}(\boldsymbol{\theta}).$$

As for (A18), we can use the generalised information matrix inequality together with some of the arguments in the proof of proposition 7 to show that

$$-\frac{\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]}{\partial\boldsymbol{\theta}} = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}]$$

$$-E\left\{\mathbf{W}_s(\boldsymbol{\phi})\left[-\left\{\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\frac{\varsigma_t}{N}+1\right\}-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})\right|\boldsymbol{\phi}\right\}$$

$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathbf{W}_s(\boldsymbol{\phi})E\left\{\left[-\left\{\frac{2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\frac{\varsigma_t}{N}+1\right\}-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi})\right|\boldsymbol{\phi}\right\}\mathbf{Z}_d(\boldsymbol{\theta})$$

$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\text{-}\mathbf{W}_s(\boldsymbol{\phi})E\left\{\left[\left\{\frac{\text{-}2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\frac{\varsigma_t}{N}\text{-}1\right\}\text{-}\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}\text{-}1\right)\right]\left\{\frac{\text{-}2\partial g(\varsigma_t,\boldsymbol{\eta}_0)}{\partial\varsigma}\frac{\varsigma_t}{N}\text{-}1\right\}\right|\boldsymbol{\phi}\right\}\mathbf{W}'_s(\boldsymbol{\theta})$$

$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathbf{W}_s(\boldsymbol{\phi})\mathbf{W}'_s(\boldsymbol{\phi})\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]-\frac{4}{N\left[(N+2)\kappa+2\right]}\right\}=\mathring{\mathcal{S}}(\boldsymbol{\phi}).$$

□

References

Abramowitz, M. and Stegun, I.A. (1964): *Handbook of mathematic functions*, AMS 55, National Bureau of Standards.

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

Balestra, P., and Holly, A. (1990), "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.

Berkane, M. and Bentler, P.M. (1986): "Moments of elliptically distributed random variates", *Statistics and Probability Letters* 4, 333-335.

Bickel, P.J. (1982): "On adaptive estimation", *Annals of Statistics* 10, 647-671.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate Generalized Arch model", *Review of Economics and Statistics* 72, 498-505.

Bollerslev, T., and Wooldridge, J. M. (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.

Bontemps, C. and Meddahi, N. (2005): "Testing distributional assumptions: a GMM approach", mimeo, U. Montreal.

Calzolari, G., Fiorentini, G. and Sentana, E. (2004): "Constrained indirect estimation", *Review of Economic Studies* 71, 945-973.

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society* B, 38, 45-53.

Davidson, R. and MacKinnon, J.G. (1998): "Graphical methods for investigating the size and power of tests statistics", *The Manchester School* 66, 1-26.

Demos, A. and Sentana, E. (1998): "Testing for Garch effects: a one-sided approach", *Journal of Econometrics*, 86, 97-127.

Drost, F.C. and Klaassen, C.A.J. (1997): "Efficient estimation in semiparametric Garch models", *Journal of Econometrics* 80, 193-221.

Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997): "Adaptive estimation in time series models", *Annals of Statistics* 25, 786-817.

Engle, R.F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* 50, 987-1007.

Engle, R.F. and Gonzalez-Rivera, G. (1991): "Semiparametric Arch models", *Journal of Business and Economic Statistics* 9, 345-360.

Fang, K.-T., Kotz, S., and Ng, K.-W. (1990), *Symmetric multivariate and related distributions*, Chapman and Hall.

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.

Fiorentini, G., Sentana, E., and Calzolari, G. (2004): "On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models", *Economics Letters* 83, 307-312.

Gonzalez-Rivera, G. (1997): "A note on adaptation in Garch models", *Econometric Reviews* 16, 55-68.

Gonzalez-Rivera, G. and Drost, F.C. (1999): "Efficiency comparisons of maximum-likelihood-based estimators in Garch models", *Journal of Econometrics* 93, 93-111.

Hafner, C.M. and Rombouts, J.V.K. (2006): "Semiparametric multivariate volatility models", forthcoming in *Econometric Theory*.

Hall, P. and Yao, Q. (2003): "Inference in Arch and Garch models with heavy-tailed errors", *Econometrica* 71, 285-317.

Hodgson, D.J. (2000): "Unconditional pseudo-maximum likelihood and adaptive estimation in the presence of conditional heterogeneity of unknown form", *Econometric Reviews* 19, 175-206.

Hodgson, D.J. and Vorkink, K.P. (2003): "Efficient estimation of conditional asset pricing models", *Journal of Business and Economic Statistics* 21, 269-283.

Kotz, S. (1975): "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical Distributions in Scientific Work*, vol. I, 247-270, Reidel.

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989): "Robust statistical modeling using the t distribution", *Journal of the American Statistical Association* 84, 881-896.

Linton, O. (1993): "Adaptive estimation in Arch models", *Econometric Theory* 9, 539-569.

Linton, O. and Steigerwald, D. (2000): "Adaptive testing in Arch models", *Econometric Reviews* 19, 145-174.

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, Chichester.

Mardia, K.V. (1970): "Measures of multivariate skewness and kurtosis with applications", *Biometrika* 57, 519-530.

Maruyama, Y. and Seo, T. (2003): "Estimation of moment parameter in elliptical distributions", *Journal of the Japan Statistical Society* 33, 215-229.

Mencía, J. and Sentana, E. (2005): "Estimation and testing of dynamic models with generalised hyperbolic innovations", mimeo, CEMFI.

NAG (2001): *NAG Fortran 77 Library Mark 19 Reference Manual.*

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.

Newey, W.K. and Powell, J.L. (1998): "Two-step estimation, optimal moment conditions, and sample selection models", mimeo, MIT.

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-99.

Sentana, E. (2004): "Factor representing portfolios in large asset markets", *Journal of Econometrics* 119, 257-289.

Sentana, E. and Fiorentini, G. (2001): "Identification, estimation and testing of conditionally heteroskedastic factor models", *Journal of Econometrics* 102, 143-164.

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