On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models*

Gabriele Fiorentini
Università di Firenze and RCEA, 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>

Revised: October 2010

Abstract

We rank the efficiency of several likelihood-based parametric and semiparametric estimators of conditional mean and variance parameters in multivariate dynamic models with potentially asymmetric and leptokurtic strong white noise innovations. We detailedly study the elliptical case, and show that Gaussian pseudo maximum likelihood estimators are inefficient except under normality. We provide consistency conditions for distributionally misspecified maximum likelihood estimators, and show that they coincide with the partial adaptivity conditions of semiparametric procedures. We propose Hausman tests that compare Gaussian pseudo maximum likelihood estimators with more efficient but less robust competitors. Finally, we provide finite sample results through Monte Carlo simulations.

Keywords: Adaptivity, Elliptical Distributions, Hausman tests, Semiparametric Estimators.

JEL: C13, C14, C12, C51, C52

*We would like to thank Dante Amengual, Manuel Arellano, Nour Meddahi, Javier Mencía, Olivier Scaillet, 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), the Second Italian Congress on Econometrics and Empirical Economics (Rimini, 2007), as well as audiences at AUEB, Bocconi, Cass Business School, CEMFI, CREST, EUI, Florence, NYU, RCEA, Roma La Sapienza and Queen Mary for useful comments and suggestions. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 (Sentana) is gratefully acknowledged.

1 Introduction

As is well known, the Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) among many others remain root-T consistent for the conditional mean and variance parameters irrespective of the degree of asymmetry and kurtosis of the conditional distribution of the observed variables, so long as the first two moments are correctly specified and the fourth moments are bounded. Nevertheless, many empirical researchers prefer to specify a non-Gaussian parametric distribution for the standardised innovations, which they use to 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 is mostly interested in the first two conditional moments, the semiparametric (SP) estimators of Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999) 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 Sun and Stengos (2006). However, they suffer from the curse of dimensionality, which severely limits their use in multivariate models. To avoid this problem, Hodgson and Vorkink (2003) and Hafner and Rombouts (2007) have considered 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 less robust.

The main objective 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 one 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 misleadingly simple. The inclusion of means in multivariate models not only provides a unified perspective in an otherwise fragmented literature, but more importantly, it allows us to cover many empirically relevant applications beyond ARCH models, which have been the motivating example for most of the existing work. In particular, our results apply to conditionally homoskedastic, dynamic linear models such as VARS or multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts.

Another important differentiating feature of our analysis is that we explicitly look at the efficiency ranking of the feasible ML procedure that jointly estimates the shape parameters, as well as the Gaussian PML, SP, SSP and infeasible ML estimators considered in the existing literature.

In addition, we provide consistency conditions for distributionally misspecified maximum likelihood estimators, and show that they coincide with the partial adaptivity conditions of semiparametric procedures. Specifically, we find that the parameters that are efficiently estimated by the semiparametric procedures, and therefore by the feasible parametric estimators under correct specification, will continue to be consistently estimated by the latter under distributional misspecification. In contrast, all the other parameters, which the semiparametric procedures can only estimate with the efficiency of the Gaussian PML estimator, will be inconsistently estimated by distributionally misspecified parametric procedures. For that reason, we propose closed-form consistent estimators for those parameters.

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.

The rest of the paper is organised as follows. In section 2, we present closed-form expressions for the score vector, Hessian and conditional information matrices of log-likelihood functions with and without the assumption of elliptical symmetry, and derive the efficiency bounds of the Gaussian PML estimator and both SP estimators. Then, in section 3 we compare the efficiency of the different estimators of the conditional mean and variance parameters, and obtain some general results on partial adaptivity. In section 4, we first study the consistency of the conditional mean and variance 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 5. Finally, we present our conclusions in section 6. Proofs and auxiliary results are gathered in appendices.

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:

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

where $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are $N \times 1$ and $N(N+1)/2 \times 1$ vector 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 "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 martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\theta}_0) = \mathbf{I}_N$. Hence,

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

To complete the model, we need to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. Following most of the literature, we shall assume that, conditional on $\mathbf{z}_t$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed, or $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0 \sim i.i.d.\ D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ for short, where $\boldsymbol{\varrho}$ are some q additional parameters that determine the shape of the distribution. Importantly, this distribution could substantially depart from a multivariate normal both in terms of skewness and kurtosis.

2.2 Maximum likelihood estimators

Let $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ denote the conditional density of $\boldsymbol{\varepsilon}_t^*$ given $\mathbf{z}_t$, I_{t-1} and the shape parameters, which we assume is well defined. Let also $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho})'$ 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 for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \boldsymbol{\varrho}\right]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

The most common choices of square root matrices are the Cholesky decomposition, which leads to a lower triangular matrix for a given ordering of $\mathbf{y}_t$, or the spectral decomposition, which yields a symmetric matrix.[1] In what follows, we shall use the former because it is much faster to compute than the latter, especially when $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is time-varying. Nevertheless, we discuss some modifications required for the spectral decomposition in Appendix B.4.

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 $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$, respectively. If $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ and $\ln f(\boldsymbol{\varepsilon}^*, \boldsymbol{\varrho})$ are differentiable, then we show in Appendix B.1 that

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \end{aligned} \tag{2}$$

[1] The choice of square root matrix is non-trivial because $l_t(\boldsymbol{\phi})$ may depend on it. In fact, it might even be possible to identify $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ without imposing any restrictions such as lower triangularity or symmetry. One such instance would be a constant mean and covariance matrix model whose innovations follow a multivariate location-scale mixture of normals (see Mencía and Sentana (2009)), under the additional assumption that the vector of asymmetry parameters is known. Even if this vector is unknown, time-variation in $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ may suffice to identify this matrix unrestrictedly (see Mencía and Sentana (2010) for further details).

where

$$\left.\begin{array}{l}\mathbf{Z}_{lt}(\boldsymbol{\theta})=\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\\ \mathbf{Z}_{st}(\boldsymbol{\theta})=\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\end{array}\right\}, \tag{3}$$

and

$$\mathbf{e}_{dt}(\boldsymbol{\phi})=\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\left[\begin{array}{l}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*,\\ -vec\left\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right]. \tag{4}$$

Similarly, let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}'=\partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, we also show in Appendix B.1 that

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}+\mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}\\ &\quad+\left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}+\left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}, &(5)\\ \mathbf{h}_{\boldsymbol{\theta\varrho}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}'+\mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}', &(6)\\ \mathbf{h}_{\boldsymbol{\varrho\varrho}t}(\boldsymbol{\phi}) &= \partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}',\end{aligned}$$

where

$$\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}=\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \tag{7}$$

and

$$\begin{aligned}\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} &= \left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}+\left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}\\ &\quad\times\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. &(8)\end{aligned}$$

Importantly, while $\mathbf{Z}_{lt}(\boldsymbol{\theta})$, $\mathbf{Z}_{st}(\boldsymbol{\theta})$, $\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ and $\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, the first and second derivatives of $\ln f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ depend on the specific distribution assumed for estimation purposes.

Given correct specification, the results in Crowder (1976) imply that the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ at $\boldsymbol{\phi}_0$ follows a vector martingale difference. His results also imply that, under suitable regularity conditions,[2] 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 $\mathcal{I}(\boldsymbol{\phi}_0)=E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$, and

$$\mathcal{I}_t(\boldsymbol{\phi})=V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right]=-E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right].$$

In this context, we can prove the following result:

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

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

[2] In particular, Crowder (1976) requires: (i) $\boldsymbol{\phi}_0$ is locally identified and belongs to the interior of the admissible parameter space, which is a compact subset of $\mathbb{R}^{p+q}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\phi}_0$; (iii) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\boldsymbol{\phi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t\mathbf{h}_t(\boldsymbol{\phi})\right]T^{-1}\sum_t\mathbf{h}_t(\boldsymbol{\phi})\xrightarrow{p}\mathbf{I}_{p+q}$, where $E^{-1}\left[-T^{-1}\sum_t\mathbf{h}_t(\boldsymbol{\phi})\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$.

and

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

with

$$\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{st}(\boldsymbol{\phi})'|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{\prime *}\otimes\mathbf{I}_N)\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\cdot\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{*\prime}\otimes\mathbf{I}_N)|\boldsymbol{\varrho}\right]-\mathbf{K}_{NN},\\
\mathcal{M}_{lr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{sr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],
\end{aligned}$$

and

$$\mathcal{M}_{rr}(\boldsymbol{\varrho}) = V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\right],$$

where $\mathbf{K}_{mn}$ *is the commutation matrix of orders* m *and* n.

2.3 Elliptically symmetric maximum likelihood estimators

The multivariate Gaussian and Student t have been by far the two most popular choices made by empirical researchers to model the distribution of standardised innovations. For that reason, we specialise our previous results to those cases in which we make the additional assumption that $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho}_0)$ is some member of the spherical family with a well defined density (see Appendix A), 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. Elliptical distributions are attractive in our context because they remain tractable irrespective of the cross-sectional dimension N. In order to highlight the change in distributional assumption, we shall use $\boldsymbol{\eta}$ instead of $\boldsymbol{\varrho}$ to denote the parameters that determine the shape of the density of $\varsigma_t=\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. The most prominent elliptically symmetric example is the normal distribution, which we denote by $\boldsymbol{\eta}_0=\mathbf{0}$. As we mentioned before, another prominent example is 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 define η as $1/\nu$, which will always remain in the finite range $[0,1/2)$ under our assumptions.

Let $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$ denote the conditional density for $\boldsymbol{\varepsilon}_t^*$ given $\mathbf{z}_t$, I_{t-1} and the q shape parameters, where $c(\boldsymbol{\eta})$ corresponds to the constant of integration, and $g(\varsigma_t,\boldsymbol{\eta})$ to its kernel.[3] Let $\boldsymbol{\phi}=(\boldsymbol{\theta}',\boldsymbol{\eta})'$ denote the $p+q$ parameters of interest, which once again we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi})=\sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi})=d_t(\boldsymbol{\theta})+c(\boldsymbol{\eta})+g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ and $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$.

[3] Fiorentini, Sentana and Calzolari (2003) (FSC) provide expressions for $c(\eta)$ and $g(\varsigma_t,\eta)$ in the multivariate Student t case, which are obviously such that $c(0)=-\frac{1}{2}\pi$ and $g(\varsigma_t,0)=-\frac{1}{2}\varsigma_t$.

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 $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. We show in Appendix B.2 that if $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ are differentiable, then

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \tag{9}$$

while we can write $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ in (2) using[4]

$$\begin{aligned}
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})], \\
\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), & (10)\\
\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right\}, & (11)
\end{aligned}$$

where

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma \tag{12}$$

is a damping factor that reflects the tail-thickness of the distribution assumed for estimation purposes. Given that this factor is equal to 1 under Gaussianity, it is straightforward to check that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

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

Assuming twice differentiability of the different functions involved, we also show in Appendix B.2 that we can write

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} & (13)\\
\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma\partial\boldsymbol{\eta}', & (14)\\
\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}' + \partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}',
\end{aligned}$$

where $\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ and $\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, while $\partial^2 g(\varsigma,\eta)/(\partial\varsigma)^2$, $\partial^2 g(\varsigma,\eta)/\partial\varsigma\partial\boldsymbol{\eta}'$ and $\partial g(\varsigma,\eta)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ depend on the specific elliptical distribution assumed for estimation purposes (see FSC for the multivariate Student t).

The expressions in Proposition 1 simplify considerably in the elliptically symmetric case. The following result generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in FSC and 5.2 in Hafner and Rombouts (2007):

[4]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 $l_t(\boldsymbol{\phi})$. This result is not generally true for non-elliptical distributions.

Proposition 2 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with density* $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$, *then*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}, \tag{15}$$

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \tag{16}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = \mathrm{M}_{ss}(\boldsymbol{\eta})\,(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N), \tag{17}$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}), \tag{18}$$

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

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

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

FSC provide analytical expressions for M_{ll}, M_{ss} and M_{sr} in the multivariate Student t case, while Amengual and Sentana (2010a) do the same for the Kotz distribution (see Kotz (1975)) and discrete scale mixtures of normals. In this sense, an important point to note in relation to the Student t is that $\mathrm{M}_{ll}(\eta)$ increases without bound as $\nu \to 2^+$ while $\mathrm{M}_{ss}(\eta)$ remains bounded. This differential behaviour is also characteristic of other leptokurtic elliptical distributions, such as the normal-gamma mixture, the Kotz distribution, or the Pearson type II.

2.4 Gaussian pseudo maximum likelihood estimators

If the interest of the researcher lies exclusively in $\boldsymbol{\theta}$, which are the parameters characterising the conditional mean and variance functions, then one attractive possibility is to estimate an equality restricted version of the spherical 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 true conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is neither Gaussian nor spherical, 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 also 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:[5]

Proposition 3 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

[5] Throughout this paper, we use the high level regularity conditions in Bollerslev and Wooldridge (1992) because we want to leave unspecified the conditional mean vector and covariance matrix in order to maintain full generality. Primitive conditions for specific multivariate models can be found for example in Ling and McAleer (2003).

1. *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}$ *is i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ *with* $tr[\mathcal{K}(\boldsymbol{\varrho})] < \infty$*, then* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)]$*, with*

$$\begin{gathered}\mathcal{C}(\boldsymbol{\phi}) = \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}),\\ \mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\ \mathcal{A}_t(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),\\ \mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\ \mathcal{B}_t(\boldsymbol{\phi}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),\\ \mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \begin{bmatrix} \mathbf{I}_N & \boldsymbol{\Phi}(\boldsymbol{\varrho}) \\ \boldsymbol{\Phi}(\boldsymbol{\varrho}) & \boldsymbol{\Upsilon}(\boldsymbol{\varrho}) \end{bmatrix}, \qquad (19)\end{gathered}$$

where

$$\begin{aligned}\boldsymbol{\Phi}(\boldsymbol{\varrho}) &= E[\boldsymbol{\varepsilon}_t^* vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\, \boldsymbol{\phi}]\\ \boldsymbol{\Upsilon}(\boldsymbol{\varrho}) &= E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)|\, \boldsymbol{\phi}]\end{aligned}$$

depend on the multivariate third and fourth order cumulants of $\boldsymbol{\varepsilon}_t^*$*, so that* $\boldsymbol{\Phi}(\mathbf{0}) = \mathbf{0}$ *and* $\boldsymbol{\Upsilon}(\mathbf{0}) = (\mathbf{I}_{N^2} + \mathbf{K}_{NN})$ *if we use* $\boldsymbol{\varrho} = \mathbf{0}$ *to denote normality.*

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* $\kappa_0 < \infty$*, then (19) reduces to*

$$\mathcal{K}(\kappa) = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}, \qquad (20)$$

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

$$\kappa = E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)] - 1. \qquad (21)$$

But if $tr[\mathcal{K}(\boldsymbol{\varrho})]$ 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)).

2.5 Semiparametric estimators

As is well known, a single scoring iteration without line searches that started from $\tilde{\boldsymbol{\theta}}_T$ and some root-T consistent estimator of $\boldsymbol{\varrho}$, $\tilde{\boldsymbol{\varrho}}_T$ say, 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,

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

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

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

where

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

and

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)=\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\varrho}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\varrho}\boldsymbol{\varrho}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0) \tag{22}$$

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 $\boldsymbol{\varrho}$, $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_0)$. The residual score $\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$ is sometimes called the parametric efficient score of $\boldsymbol{\theta}$, and its variance,

$$\mathcal{P}(\boldsymbol{\phi}_0)=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\varrho}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\varrho}\boldsymbol{\varrho}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\varrho}}(\boldsymbol{\phi}_0), \tag{23}$$

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, $\hat{\boldsymbol{\theta}}_T$. In contrast, $\mathcal{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$ would give us the asymptotic variance of an infeasible restricted ML estimator, which we would obtain only if we could fix the shape parameters $\boldsymbol{\varrho}$ to their true values. For that reason, we shall refer to $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$ as the infeasible parametric efficiency bound.

In the elliptically symmetric case, we can easily prove that (22) and (23) reduce to

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

and

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

respectively, where

$$\mathbf{W}_s(\boldsymbol{\phi}_0)=\mathbf{Z}_d(\boldsymbol{\phi}_0)[\mathbf{0}',vec'(\mathbf{I}_N)]'=E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}',vec'(\mathbf{I}_N)]'$$
$$=E\left\{\frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]/\partial\boldsymbol{\theta}\cdot vec[\boldsymbol{\Sigma}^{-1}_t(\boldsymbol{\theta}_0)]\right|\boldsymbol{\phi}_0\left.\right\}=E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]=-E\left\{\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|\,\boldsymbol{\phi}_0\right\}. \tag{24}$$

It is worth noting that the last summand of (22) 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 Lemma 2. Such an interpretation immediately suggests alternative estimators of $\boldsymbol{\theta}$ that replace a 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 3 and 4 in Hafner and Rombouts (2007) 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 4 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0$ *is i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ *with density function* $f(\varepsilon_t^*; \boldsymbol{\varrho})$, *where* $\boldsymbol{\varrho}$ *contains some shape parameters and* $\boldsymbol{\varrho} = \mathbf{0}$ *denotes normality, such that both its Fisher information matrix for location and scale,* $\mathcal{M}_{dd}(\boldsymbol{\varrho})$, *and the matrix of third and fourth order central moments* $\mathcal{K}(\boldsymbol{\varrho})$ *are bounded, then the semiparametric efficient score will be given by:*

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})\right], \tag{25}$$

while the semiparametric efficiency bound is

$$\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho}_0)\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0), \tag{26}$$

where $+$ *denotes Moore-Penrose inverses.*

In practice, however, $f(\varepsilon_t^*; \boldsymbol{\varrho})$ has to be replaced by a nonparametric estimator, which suffers from the curse of dimensionality. For this reason, Hodgson and Vorkink (2001), Hafner and Rombouts (2007) 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 amends and extends Proposition 9 in Hafner and Rombouts (2007), provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 5 *When* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ *with* $-2/(N+2) < \kappa_0 < \infty$, *the elliptically symmetric semiparametric efficient score is given by:*

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2}{(N+2)\kappa_0 + 2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\right\}, \tag{27}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]}\right\}. \tag{28}$$

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.[6]

3 The relative efficiency of the different estimators

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, restricted ML estimator, whose computation requires knowledge of $\boldsymbol{\varrho}_0$; (2) the feasible, unrestricted ML estimator, which simultaneously estimates $\boldsymbol{\varrho}$; (3) the elliptically symmetric

[6] Hodgson, Linton and Vorkink (2002) also consider alternative estimators that iterate the semiparametric adjustment until it becomes negligible. However, since they have the same first-order asymptotic distribution, we shall not discuss them separately.

semiparametric estimator, which restricts $\boldsymbol{\varepsilon}_t^*$ 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 $\boldsymbol{\varepsilon}_t^*$ is *i.i.d.*$(\mathbf{0}, \mathbf{I}_N)$; and (5) the Gaussian PML estimator, which imposes $\boldsymbol{\eta} = \mathbf{0}$ even though the true conditional distribution of $\boldsymbol{\varepsilon}_t^*$ could be neither normal nor spherical. The following proposition ranks (in the usual positive semidefinite sense) the "information matrices" of those five estimators:

Proposition 6 *1. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ with $tr[\mathcal{K}(\boldsymbol{\varrho})] < \infty$, then*

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

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 $\kappa_0 < \infty$, 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 such instance within the spherical family:

Proposition 7 *1. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $N(\mathbf{0}, \mathbf{I}_N)$, then*

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

where

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

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 $-2/(N+2) < \kappa_0 < \infty$, 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 $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and $\mathbf{Z}_l(\boldsymbol{\phi}_0) \neq \mathbf{0}$, 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 FSC, implies that as far as $\boldsymbol{\theta}$ is concerned, there is no asymptotic efficiency loss in estimating $\boldsymbol{\varrho}$ when $\boldsymbol{\varrho}_0 = \mathbf{0}$. The second part, which generalises the results in Gonzalez-Rivera (1997), implies that the SSP estimator can be fully efficient only if $\boldsymbol{\varepsilon}_t^*$ has a conditional Kotz distribution, 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 7 generalises Result 2 in Drost and Gonzalez-Rivera (1999) and Proposition 6 in Hafner and Rombouts (2007).

While it is relatively straightforward to obtain closed-form expressions for the different efficiency bounds in conditionally homoskedastic, dynamic linear models such as multivariate regressions or VARs (see e.g. Amengual and Sentana (2010a)), it is virtually impossible to do so in dynamic conditionally heteroskedastic models, as one has to resort to numerical or Monte Carlo integration methods to compute the expected values of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ or $\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta})$ (see e.g. Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999)).[7]

3.2 General results on partial adaptivity

There are situations in which some, but not all elements of $\boldsymbol{\theta}$ can be estimated as efficiently as if $\boldsymbol{\varrho}_0$ 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 partialling 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 reparametrisation provides a general sufficient condition in multivariate dynamic models under ellipticity:

Reparametrisation 1 *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$[\partial \mathbf{r}'_s(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}] = p$, such that*

$$\left.\begin{array}{r} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1) \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_2 \boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_1) \end{array}\right\} \quad \forall t. \tag{29}$$

Such a reparametrisation is not unique, since we can always multiply the overall scale parameter ϑ_2 by some scalar positive smooth function of $\boldsymbol{\vartheta}_1$, $k(\boldsymbol{\vartheta}_1)$ say, and divide $\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_1)$ by the same function without violating (29) or redefining $\boldsymbol{\vartheta}_1$. As we shall see, a particularly convenient function would be such that after re-scaling[8]

$$E[\partial \ln |\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_1)| / \partial \boldsymbol{\vartheta}_1 | \boldsymbol{\phi}] = \mathbf{0}. \tag{30}$$

The following proposition generalises and extends earlier results by Bickel (1982), Linton (1993), Drost, Klaassen and Werker (1997) and Hodgson and Vorkink (2003):

Proposition 8 *1. If $\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ and (29) holds, then:*

(a) the elliptically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_1$ is ϑ_2-adaptive,

[7] But see Fiorentini and Sentana (2009, 2010) for closed-form expressions in the context of tests for univariate or multivariate conditional homoskedasticity.

[8] Amengual and Sentana (2010a) provide an example of a reparametrisation that achieves (30) in an unrestricted conditionally homoskedastic context. Specifically, they model $\boldsymbol{\Sigma}$ as $\vartheta_2 \boldsymbol{\Sigma}^{\circ}(\boldsymbol{\vartheta}_1)$, where $\boldsymbol{\vartheta}_1$ are $N(N+1)/2 - 1$ parameters that ensure that $|\boldsymbol{\Sigma}^{\circ}(\boldsymbol{\vartheta}_1)| = 1 \; \forall \boldsymbol{\vartheta}_1$. In other words, their reparametrisation is such that $\vartheta_2 = |\boldsymbol{\Sigma}|^{1/N}$ and $\boldsymbol{\Sigma}^{\circ}(\boldsymbol{\vartheta}_1) = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/N}$.

(b) *If* $\mathring{\boldsymbol{\vartheta}}_T$ *denotes the iterated elliptically symmetric semiparametric estimator of* $\boldsymbol{\vartheta}$, *then* $\mathring{\vartheta}_{2T} = \vartheta_{2T}(\mathring{\boldsymbol{\vartheta}}_{1T})$, *where*

$$\begin{aligned} \vartheta_{2T}(\boldsymbol{\vartheta}_1) &= \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\varsigma_t^{\circ}(\boldsymbol{\vartheta}_1), & (31)\\ \varsigma_t^{\circ}(\boldsymbol{\vartheta}_1) &= [\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1)]'\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_1)[\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1)], & (32)\end{aligned}$$

(c) *rank* $\left[\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) - \mathcal{C}^{-1}(\boldsymbol{\phi}_0)\right] \leq \dim(\boldsymbol{\vartheta}_1) = p-1$.

2. *If in addition condition (30) holds at* $\boldsymbol{\vartheta}_{10}$, *then:*

(a) $\mathcal{I}_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}}(\boldsymbol{\phi}_0), \mathcal{P}(\boldsymbol{\phi}_0), \mathring{\mathcal{S}}(\boldsymbol{\phi}_0), \mathcal{S}(\boldsymbol{\phi}_0)$ *and* $\mathcal{C}(\boldsymbol{\phi}_0)$ *are block-diagonal between* $\boldsymbol{\vartheta}_1$ *and* ϑ_2.

(b) $\sqrt{T}(\mathring{\vartheta}_{2T} - \tilde{\vartheta}_{2T}) = o_p(1)$, *where* $\tilde{\boldsymbol{\vartheta}}_T' = (\tilde{\boldsymbol{\vartheta}}_{1T}', \tilde{\vartheta}_{2T})$ *is the Gaussian PMLE of* $\boldsymbol{\vartheta}$, *with* $\tilde{\vartheta}_{2T} = \vartheta_{2T}(\tilde{\boldsymbol{\vartheta}}_{1T})$.

This proposition provides a saddle point characterisation of the asymptotic efficiency of the elliptically symmetric semiparametric estimator of $\boldsymbol{\theta}$, in the sense that in principle it can estimate $p-1$ "parameters" as efficiently as if we fully knew the true conditional distribution of the data, while for the remaining scalar "parameter" it only achieves the efficiency of the Gaussian PMLE. Obviously, the feasible, unrestricted ML estimator of $\boldsymbol{\vartheta}_1$ will also be ϑ_2-adaptive when the assumed parametric conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is correct in view of Proposition 6.

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:

Reparametrisation 2 *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 parameter set* $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', \boldsymbol{\psi}_3')'$, *where* $\boldsymbol{\psi}_2 = vech(\boldsymbol{\Psi}_2)$, $\boldsymbol{\Psi}_2$ *is an unrestricted positive (semi)definite matrix of order* N, $\boldsymbol{\psi}_3$ *is* $N \times 1$, *and* $\mathbf{r}_g(\boldsymbol{\theta})$ *is twice continuously differentiable in a neighbourhood of* $\boldsymbol{\theta}_0$ *with rank*$\left[\partial\mathbf{r}_g'(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\right] = p$, *such that*

$$\left.\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_1) + \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)\boldsymbol{\psi}_3 \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)\boldsymbol{\Psi}_2\boldsymbol{\Sigma}_t^{\diamond 1/2\prime}(\boldsymbol{\psi}_1) \end{aligned}\right\} \quad \forall t. \tag{33}$$

This parametrisations simply requires the pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_1) = \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_1)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_1)] \tag{34}$$

to be *i.i.d.* $(\boldsymbol{\psi}_3, \boldsymbol{\Psi}_2)$. Again, (33) is not unique, since it continues to hold with the same $\boldsymbol{\psi}_1$ if we replace $\boldsymbol{\Psi}_2$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_1)\boldsymbol{\Psi}_2\mathbf{K}^{-1/2\prime}(\boldsymbol{\psi}_1)$ and $\boldsymbol{\psi}_3$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_1)\boldsymbol{\psi}_3 - \mathbf{l}(\boldsymbol{\psi}_1)$, and adjust $\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_1)$ and $\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)$ accordingly, where $\mathbf{l}(\boldsymbol{\psi}_1)$ and $\mathbf{K}(\boldsymbol{\psi}_1)$ are a $N \times 1$ vector and a $N \times N$ positive definite matrix of smooth functions of $\boldsymbol{\psi}_1$, respectively. As we shall see, particularly convenient forms for these functions would be those which achieve that after re-centring and re-scaling

$$\left.\begin{aligned} E\left[\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_1)/\partial\boldsymbol{\psi}_1 \cdot \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_1)\right|\boldsymbol{\phi}\Big] = \mathbf{0} \\ E\left\{\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)]/\partial\boldsymbol{\psi}_1 \cdot \left[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_1)\right]\right|\boldsymbol{\phi}\Big\} = \mathbf{0} \end{aligned}\right\}. \tag{35}$$

The following proposition, which does not require sphericity, generalises and extends Theorems 3.1 in Drost and Klaassen (1997) and 3.2 in Sun and Stengos (2006):

Proposition 9 *1. If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$, and (33) holds, then*

(a) the semiparametric estimator of $\boldsymbol{\psi}_1$, $\breve{\boldsymbol{\psi}}_{1T}$, is $(\boldsymbol{\psi}_2, \boldsymbol{\psi}_3)$-adaptive,

(b) If $\breve{\boldsymbol{\psi}}_T$ denotes the iterated semiparametric estimator of $\boldsymbol{\psi}$, then $\breve{\boldsymbol{\psi}}_{2T} = \boldsymbol{\psi}_{2T}(\breve{\boldsymbol{\psi}}_{1T})$ and $\breve{\boldsymbol{\psi}}_{3T} = \boldsymbol{\psi}_{3T}(\breve{\boldsymbol{\psi}}_{1T})$, where

$$\boldsymbol{\psi}_{2T}(\boldsymbol{\psi}_1) = vech\left\{\frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\psi}_1) - \boldsymbol{\psi}_{3T}(\boldsymbol{\psi}_1)][\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\psi}_1) - \boldsymbol{\psi}_{3T}(\boldsymbol{\psi}_1)]'\right\}, \quad (36)$$

$$\boldsymbol{\psi}_{3T}(\boldsymbol{\psi}_1) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\psi}_1) \quad (37)$$

(c) rank $\left[\mathcal{S}(\boldsymbol{\phi}_0) - \mathcal{C}^{-1}(\boldsymbol{\phi}_0)\right] \leq \dim(\boldsymbol{\psi}_1) = p - N - N(N+1)/2$.

2. If in addition condition (35) holds at $\boldsymbol{\psi}_{10}$, then

(a) $\mathcal{I}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\phi}_0), \mathcal{P}(\boldsymbol{\phi}_0), \mathcal{S}(\boldsymbol{\phi}_0)$ and $\mathcal{C}(\boldsymbol{\phi}_0)$ are block diagonal between $\boldsymbol{\psi}_1$ and $(\boldsymbol{\psi}_2, \boldsymbol{\psi}_3)$.

(b) $\sqrt{T}[(\breve{\boldsymbol{\psi}}'_{2T} - \tilde{\boldsymbol{\psi}}'_{2T}), (\breve{\boldsymbol{\psi}}'_{3T} - \tilde{\boldsymbol{\psi}}'_{3T})]' = o_p(1)$, where $\tilde{\boldsymbol{\psi}}'_T = (\tilde{\boldsymbol{\psi}}'_{1T}, \tilde{\boldsymbol{\psi}}'_{2T}, \tilde{\boldsymbol{\psi}}'_{3T})$ is the Gaussian PMLE of $\boldsymbol{\psi}$, with $\tilde{\boldsymbol{\psi}}_{2T} = \boldsymbol{\psi}_{2T}(\tilde{\boldsymbol{\psi}}'_{1T})$ and $\tilde{\boldsymbol{\psi}}_{3T} = \boldsymbol{\psi}_{3T}(\tilde{\boldsymbol{\psi}}'_{1T})$.

This proposition provides a saddle point characterisation of the asymptotic efficiency of the semiparametric estimator of $\boldsymbol{\theta}$, in the sense that in principle it can estimate $p - N(N+3)/2$ "parameters" as efficiently as if we fully knew the true conditional distribution of the data, while for the remaining "parameters" it only achieves the efficiency of the Gaussian PMLE.

Many conditionally homokedastic multivariate regression models, including VARs, can be written as in (33) by identifying $\boldsymbol{\psi}_1$ with the slope coefficients after suitably redefining the intercepts. In contrast, 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 as (33) *if* we additionally assume that $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mathbf{0}\ \forall t$, in which case $\boldsymbol{\psi}_3$ is unnecessary. And even in that case, we could only adaptively estimate the parameters of $\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1) = \mathbf{D}_t(\boldsymbol{\theta}_1)\{E[\mathbf{D}_t(\boldsymbol{\theta}_1)]|\boldsymbol{\phi}_0\}^{-1}$, which will typically correspond to the relative scale parameters of the N univariate ARCH models for the elements of $\mathbf{y}_t$, although Ling and McAleer (2003) consider a more general specification. In most other models, we may need to artificially augment the original parametrisation with $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$ even though we know that $\boldsymbol{\psi}_{20} = vech(\mathbf{I}_N)$ and $\boldsymbol{\psi}_{30} = \mathbf{0}$, which could be associated with a substantial efficiency cost. Furthermore, in doing so, we must guarantee that the parameters $\boldsymbol{\psi}_1$ 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 8 and 9 is that in the elliptically symmetric case we can restrict $\boldsymbol{\Psi}_2$ to be a scalar matrix, and $\boldsymbol{\psi}_3$ to $\mathbf{0}$ regardless of the mean specification, which reduces the number of parameters involved by a factor of $N(N+3)/2$.

4 Distributional misspecification and parameter consistency

4.1 Parameter estimation

So far, we have maintained the assumption that the true conditional distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ is correctly specified. 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, some of the elements of an efficient 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, as forcefully argued by Newey and Steigerwald (1997) in the univariate case. To focus our discussion on the effects of distributional misspecification, in the remaining of this section we shall assume that (1) is true.

Let us first consider situations in which the true distribution is *i.i.d.* elliptical but different from the parametric one assumed for estimation purposes, which will often be chosen for convenience or familiarity. Note that this covers situations in which the conditionally elliptical distribution is correctly specified, but we fix $\boldsymbol{\eta}$ to some a priori chosen value $\bar{\boldsymbol{\eta}}$ which does not coincide with the true value $\boldsymbol{\eta}_0$.

For simplicity, we shall define the pseudo-true values of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ as consistent roots of the expected elliptical pseudo log-likelihood score, which under appropriate regularity conditions will maximise the expected value of the pseudo log-likelihood function. The first part of the following proposition extends the first part of Theorem 1 in Newey and Steigerwald (1997) to multivariate dynamic models, while the rest does the same thing for Proposition 5 in Amengual and Sentana (2010a).

Proposition 10 *If (29) holds, and $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\varphi}_0$, is i.i.d. $s(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then:*

1. *The pseudo-true value of a feasible spherically-based ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\vartheta}_1', \vartheta_2, \boldsymbol{\eta})'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\vartheta}_{1\infty}$ is equal to the true value $\boldsymbol{\vartheta}_{10}$.*

2.
$$E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{0},$$
$$\mathcal{O}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\vartheta}_\infty)\mathcal{M}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\vartheta}_\infty),$$
$$\mathcal{H}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = -E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\vartheta}_\infty)\mathcal{M}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\vartheta}_\infty),$$

where both $\mathcal{M}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{M}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ share the structure of (15), (16), (17) and (18), with

$$\begin{gathered}
\mathrm{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{\delta^2[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]\middle|\,\boldsymbol{\varphi}\right\} \\
\mathrm{M}^O_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) = N(N+2)^{-1}\left[1+V\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]|\,\boldsymbol{\varphi}\right\}\right], \\
\mathrm{M}^O_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left[\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]-1\right\}\mathbf{e}'_{rt}(\boldsymbol{\phi})\middle|\,\boldsymbol{\varphi}\right], \\
\mathcal{M}^O_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = V[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\varphi}], \\
\mathrm{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]/\partial\varsigma\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]+\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\,\boldsymbol{\varphi}\right\}, \\
\mathrm{M}^H_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]/\partial\varsigma\cdot\varsigma^2_t(\boldsymbol{\vartheta})/[N(N+2)]\middle|\,\boldsymbol{\varphi}\right\}+1, \\
\mathrm{M}^H_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = -E\left\{[\varsigma_t(\boldsymbol{\vartheta})/N]\cdot\partial\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]/\partial\boldsymbol{\eta}|\boldsymbol{\varphi}\right\}, \\
\mathcal{M}^H_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = -E[\,\partial\mathbf{e}_{rt}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\,\boldsymbol{\varphi}].
\end{gathered}$$

3. *If in addition (30) holds at $\boldsymbol{\vartheta}_{10}$, then $E[\mathcal{O}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ and $E[\mathcal{H}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ will be block diagonal between $\boldsymbol{\vartheta}_1$ and $(\vartheta_2,\boldsymbol{\eta})$.*

Part 1 says that a spherically-based, unrestricted PMLE can consistently estimate all the parameters except the expected value of $\varsigma^\circ_t(\boldsymbol{\vartheta}_{10})$ in (32), while Part 2 allows us to obtain the asymptotic variance of the spherically-based PML estimators with the usual sandwich formula. Importantly, the above results also apply suitably modified to restricted spherically-based ML estimators of $\boldsymbol{\vartheta}$ that fix $\boldsymbol{\eta}$ to some a priori chosen value $\bar{\boldsymbol{\eta}}$.

Remarkably, note that the transformed parameters that we can estimate in a partially adaptive manner by means of the SSP estimator, and therefore by the feasible parametric procedures under correct specification, coincide with the parameters that we continue to estimate consistently with a misspecified, spherically-based, pseudo-ML estimator. In contrast, the remaining parameter, which the SSP procedure can only estimate with the efficiency of the Gaussian PML estimator, will be inconsistently estimated by distributionally misspecified parametric procedures. Nevertheless, it should be straightforward to consistently estimate the overall scale parameter ϑ_2 by combining $\hat{\boldsymbol{\vartheta}}_{1T}$ with the expression for the concentrated Gaussian PML and iterated SSP estimators in (31).

If $\boldsymbol{\varepsilon}^*_t|\mathbf{z}_t,I_{t-1},\boldsymbol{\varphi}_0$ is not *spherical*, then in general some elements of the feasible elliptically-based PML estimator will be inconsistent, and the same applies to the SSP estimator. Indeed, such inconsistencies will also affect a parametric non-elliptical estimator if the distribution used for computing the log-likelihood function does not nest the true distribution, or even if it is correctly specified but we fix $\boldsymbol{\varrho}$ to some a priori chosen value $\bar{\boldsymbol{\varrho}}$ which differs from the true value $\boldsymbol{\varrho}_0$. Once again, though, it may still be possible to estimate consistently some parameters:

Proposition 11 *If (33) holds, and $\boldsymbol{\varepsilon}^*_t|\mathbf{z}_t,I_{t-1};\boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then:*

1. *The pseudo-true value of the feasible parametric ML estimator of* $\boldsymbol{\phi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', \boldsymbol{\psi}_3', \boldsymbol{\varrho})'$, $\boldsymbol{\phi}_\infty$, *is such that* $\boldsymbol{\psi}_{1\infty}$ *is equal to the true value* $\boldsymbol{\psi}_{10}$.

2.

$$E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0] = \mathbf{0},$$
$$\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\psi}_\infty)\mathcal{M}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\psi}_\infty),$$
$$\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\psi}_\infty)\mathcal{M}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\psi}_\infty),$$

where $\mathcal{M}^O(\boldsymbol{\phi}; \boldsymbol{\varphi}) = V[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\varphi}]$, *while*

$$\begin{aligned}
\mathcal{M}_{ll}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right| \boldsymbol{\varphi}\right\},\\
\mathcal{M}_{ls}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime} \cdot [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) \otimes \mathbf{I}_N])\right| \boldsymbol{\varphi}\right\},\\
\mathcal{M}_{ss}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) \otimes \mathbf{I}_N] \cdot \partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime} \cdot [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) \otimes \mathbf{I}_N]|\boldsymbol{\varphi}\right\} - \mathbf{K}_{NN}\\
\mathcal{M}_{lr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= -E\left[\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right],\\
\mathcal{M}_{sr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= -E\left[[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right\},
\end{aligned}$$

and

$$\mathcal{M}_{rr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right\}.$$

3. *If in addition (35) holds at* $\boldsymbol{\psi}_{10}$, *then* $E[\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ *and* $E[\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ *will be block diagonal between* $\boldsymbol{\psi}_1$ *and* $(\boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \boldsymbol{\varrho})$.

The first part of this proposition is the multivariate generalisation of Theorem 2 in Newey and Steigerwald (1997).[9] Obviously, it also applies when the density assumed for estimation purposes is elliptical, whether or not it is parametrically specified, although in that case the expressions for $\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ and $\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ will simplify considerably along the lines of Proposition 10. And as in the case of Proposition 10, the above results also apply to restricted ML estimators of $\boldsymbol{\psi}$ that fix $\boldsymbol{\varrho}$ to some a priori chosen value $\bar{\boldsymbol{\varrho}}$. In simple terms, Proposition 11 says that in general, a misspecified parametric ML estimator cannot consistently estimate either the mean or the covariance matrix of the *i.i.d.* pseudo-standardised residuals $\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\psi}_{10})$ in (34), which is the precise multivariate analogue to the Newey and Steigerwald (1997) univariate result, who only needed ψ_2 and ψ_3 scalar.

Once again, note that the transformed parameters that we can estimate in a partially adaptive manner by means of the unrestricted semiparametric estimator, and therefore by the feasible parametric procedures under correct specification, coincide with the parameters that we continue to estimate consistently with a misspecified parametric ML estimator. In contrast, all the other parameters, which the semiparametric procedures can only estimate with the efficiency

[9] 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 $\boldsymbol{\psi}_3$ is unnecessary. Since a zero conditional mean assumption is in principle as contentious as any other parametric specification for the first moment, we shall not separately discuss this case any further.

of the Gaussian PML estimator, will be inconsistently estimated by distributionally misspecified parametric procedures. However, it should be straightforward to consistently estimate $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$ by combining $\breve{\boldsymbol{\psi}}_{1T}$ with the expressions for the concentrated Gaussian PML and SP estimators in (36) and (37).[10]

Propositions 10 and 11 will trivially yield the expressions in Proposition 3 when the distribution used for estimation purposes is the multivariate normal. It turns out that there are other cases in which the whole of $\boldsymbol{\theta}$ will be consistently estimated despite distributional misspecification. In particular, imagine that we decide to use a Student t (pseudo) log-likelihood function, which requires us to impose the inequality constraint $\eta \geq 0$:

Proposition 12 *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 (21) 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*

$$\mathcal{H}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[[N + 2 - \varsigma_t(\boldsymbol{\theta}_0)]\{\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)|vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\}\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\varphi}_0] = \mathbf{0}, \quad (38)$$

where $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)$, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$ the rest of the time.

In fact, as far as $\hat{\boldsymbol{\theta}}_T$ is concerned, this result is valid not only for the Student t, but also for any pseudo ML estimator based on a symmetric generalised hyperbolic distribution (see Mencía and Sentana (2010) for details). In addition, it is also true for ML estimators based on fourth order elliptically symmetric expansions of the multivariate normal density, as well as on discrete scale mixtures of normals in which the odds ratio of the components is given (see Amengual and Sentana (2010b)). More generally, it will be true for any leptokurtic spherical distribution that nests the normal as a limiting case, and which is such that the scores with respect to the shape parameters evaluated under Gaussianity are proportional to the second generalised Laguerre polynomial

$$\varsigma_t^2(\boldsymbol{\theta})/4 - (N+2)\varsigma_t(\boldsymbol{\theta})/2 + N(N+2)/4. \quad (39)$$

In all those cases $\hat{\boldsymbol{\theta}}_T = \tilde{\boldsymbol{\theta}}_T$ whenever $\hat{\boldsymbol{\eta}}_T = \mathbf{0}$, which will occur when the sample coefficient of excess kurtosis is non-positive.

[10]See also Fan, Qi and Xiu (2010) for consistent estimators of univariate GARCH models with zero conditional mean.

Finally, it is worth pointing out that the semiparametric estimator may also become inconsistent if the *i.i.d.* assumption does not hold.[11] In this sense, one should 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.

4.2 Hausman tests

There are several ways in which we can test the validity of the parametric assumption made for estimation purposes. 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 (2010), who use the generalised hyperbolic family as nesting distribution for the multivariate normal and Student t. An alternative procedure would be an information matrix test that compares some or all the elements of $\mathcal{M}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{M}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ in Propositions 10 or 11 by means of an unconditional moment test. But we can also consider a Hausman specification test. The rationale is that the feasible parametric 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 $\boldsymbol{\varepsilon}_t^*$ is misspecified, then $\tilde{\boldsymbol{\theta}}_T$ will remain root-T consistent as long as the fourth order moments are bounded, while $\hat{\boldsymbol{\theta}}_T$ will probably not, as Propositions 10 and 11 illustrate. More formally

Proposition 13 *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 parametric ML estimator $\hat{\boldsymbol{\theta}}_T$. If the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied and $tr[\mathcal{K}(\mathbf{0})] < \infty$, 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 $s = rank\left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right]$.

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 those matrices as sample averages of the corresponding conditional expressions in Propositions 1 or 2 and Proposition 3 evaluated at a common estimator of $\boldsymbol{\phi}$, such as the joint MLE $\hat{\boldsymbol{\phi}}_T$, or the Gaussian PML $\tilde{\boldsymbol{\theta}}_T$ coupled with sequential ML or method

[11] Hodgson (2000) shows that the consistency of the conditional mean parameters is preserved in non-linear univariate regression models when the innovations are conditionally symmetric but not *i.i.d.* if certain conditions are satisfied. See also Proposition 7 in Amengual and Sentana (2010a) for a multivariate example.

of moments estimators of $\boldsymbol{\varrho}$ (see Amengual, Fiorentini and Sentana (2010)), the latter often being such that $\mathcal{B}(\boldsymbol{\theta},\boldsymbol{\varrho})$ remains bounded.

Unfortunately, these feasible versions of the Hausman tests will not work properly when $tr[\mathcal{K}(\mathbf{0})]$ becomes unbounded, which violates one of the assumptions of Proposition 3. Similarly, in view of Propositions 7 and 12, a feasible Hausman test for the Student t and related distributions will become numerically unstable when the true distribution is Gaussian but the estimator of η is strictly positive 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 that may be more affected by such distributional misspecification. For instance, in the situation discussed in Proposition 10 power would be maximised if we based our Hausman test on the overall scale parameter ϑ_2 exclusively, and the same will be true in the context of Proposition 11 if we look at $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$, which contain the variance and mean parameters of the pseudo standardised residuals $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_1)$ in (34), respectively.

Given that the SSP estimator is also efficient relative to the PML estimator under sphericity, but it may lose its consistency otherwise, we can assess the elliptical assumption with the following alternative specification tests:

Proposition 14 *Let*

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

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* $H^W_{\mathring{\boldsymbol{\theta}}_T}\overset{d}{\to}\chi^2_s$ *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* $s=rank[\mathcal{C}(\boldsymbol{\phi}_0)-\mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)]\leq p-1$.

Once again, it may be convenient to concentrate on the parameters that are more likely to reflect the distributional misspecification, such as $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$.

5 Monte Carlo Evidence

5.1 Design and estimation details

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 the single factor version of the conditionally heteroskedastic factor model in Sentana and Fiorentini (2001) with covariance stationary diagonal VAR(1) dynamics for the mean, and GARCH dynamics for the variance of the common factor. Thus:

$$\left.\begin{array}{c}\mathbf{y}_t=\boldsymbol{\mu}_t(\boldsymbol{\pi}_0,\boldsymbol{\rho}_0)+\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}_0)\varepsilon_t^*,\\ \boldsymbol{\mu}_t(\boldsymbol{\pi},\boldsymbol{\rho})=[\mathbf{I}_N-diag(\boldsymbol{\rho})]\boldsymbol{\pi}+diag(\boldsymbol{\rho})\mathbf{y}_{t-1},\\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta})=\mathbf{cc}'\lambda_t(\boldsymbol{\theta})+\boldsymbol{\Gamma},\\ \lambda_t(\boldsymbol{\theta})-\lambda=\alpha[f_{kt-1}^2(\boldsymbol{\theta})+\omega_{t-1}(\boldsymbol{\theta})-\lambda]+\beta[\lambda_{t-1}(\boldsymbol{\theta})-\lambda],\\ \varepsilon_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\eta}_0\sim i.i.d.\ s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0),\end{array}\right\}\quad(40)$$

where $f_{kt}(\boldsymbol{\theta})$ is the conditionally linear Kalman filter estimator of the underlying common factor, $\omega_t(\boldsymbol{\theta})$ the corresponding conditional mean square error (see Sentana (2004) for details), $\boldsymbol{\theta}=(\boldsymbol{\pi}',\boldsymbol{\rho}',\mathbf{c}',\boldsymbol{\gamma}',\alpha,\beta)'$, $\boldsymbol{\pi}=(\pi_1,\ldots,\pi_N)'$, $\boldsymbol{\rho}=(\rho_1,\ldots,\rho_N)'$, and $\boldsymbol{\gamma}=vecd(\boldsymbol{\Gamma})$. Specifically, we simulate and estimate a model in which $N=6$, $\boldsymbol{\pi}_0=.1\cdot\boldsymbol{\iota}_6$, $\boldsymbol{\rho}_0=.1\cdot\boldsymbol{\iota}_6$, $\mathbf{c}_0=\boldsymbol{\iota}_6$, $\boldsymbol{\gamma}_0=2\cdot\boldsymbol{\iota}_6$, $\boldsymbol{\iota}_6=(1,1,1,1,1,1)'$, $\lambda_0=1$, $\alpha_0=.1$ and $\beta_0=.85$. As for $\boldsymbol{\varepsilon}_t^*$, we consider a Gaussian distribution, and two multivariate Student t's with 8 and 4 degrees of freedom respectively. In order to assess the effects of distributional misspecification, we also consider an *i.i.d.* normal-gamma mixture with the same coefficient of multivariate excess kurtosis as the t_8, an *i.i.d.* asymmetric Student t such that the marginal distribution of an equally-weighted average of the six series has the maximum negative skewness possible for the kurtosis of the t_8, and a symmetric Student t distribution with time-varying kurtosis, in which the degrees of freedom parameter evolves according to the following stochastic difference equation

$$\nu_t=.8+.8(f_{kt-1}^2+\omega_{t-1})\lambda_{t-1}^{-1}+.8\nu_{t-1},$$

which can be regarded as a multivariate version of expression (7) in Demos and Sentana (1998).[12] We exploit the results in Mencía and Sentana (2010) to simulate standardised versions of all these distributions by appropriately mixing a 6-dimensional spherical normal vector with a univariate gamma random variable, which we obtain from the NAG Fortran 77 Mark 19 library routines G05DDF and G05FFF, respectively (see Numerical Algorithm Group (2001) for details). As we mentioned in section 2.2, we systematically resort to Cholesky decompositions to factorise $\boldsymbol{\Sigma}_t$ with the objective of speeding up the computations. This choice is inconsequential for all simulated distributions except the asymmetric t, and all estimators except the SP one. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T=1,000$ observations (plus another 100 for initialisation) based on 10,000 Monte Carlo replications. This sample size corresponds roughly to 20 years of weekly data, or 4 years of daily data.

Our ML estimation procedure employs the following numerical strategy. First, we estimate the conditional mean and variance parameters $\boldsymbol{\theta}$ under normality with a scoring algorithm that

[12] A direct application of the formulas in Demos and Sentana (1998, sect.3.1) yields $\inf_t \nu_t=4$ and $E(\nu_t)=8$.

combines the E04LBF routine with the analytical expressions for the score in Appendix B.3 and the $\mathcal{A}(\boldsymbol{\phi}_0)$ matrix in Proposition 3. Then, we compute the sequential MM estimator $\breve{\eta}_T$ proposed by FSC, which is given by

$$\breve{\eta}_T = \frac{\max[0, \bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)]}{4\max[0, \bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)] + 2}, \tag{41}$$

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. Then we use $\breve{\eta}_T$ as initial value for a univariate optimisation procedure that obtains the sequential ML estimator of η that maximises the Student t log-likelihood function with the E04ABF routine keeping $\boldsymbol{\theta}$ fixed at its Gaussian PMLE, $\tilde{\boldsymbol{\theta}}_T$. This estimator, together with the PML of $\boldsymbol{\theta}$, become the initial values for the t-based ML estimators, which are obtained with the same scoring algorithm as the PML estimator, but this time using the analytical expressions for the information matrix $\mathcal{I}(\boldsymbol{\phi}_0)$ in Proposition 2. We rule out numerically problematic solutions by imposing the inequality constraints $|\rho_i| \leq .999$ and $\gamma_i \geq 10^{-10}$ for $i = 1, \ldots, N$, $\alpha \geq 10^{-4}$, $\beta \geq 0$, $\alpha + \beta \leq .999$ and $0 \leq \eta \leq .499$.[13] Given that the scale of the common factor is free, we set $\lambda = 1$ in estimation for computational convenience but report results for the alternative normalisation $c_1 = 1$.

Computational details for the two semiparametric procedures can be found in Appendices B.3 and B.4. Given that a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise with $N = 6$, we have done some experimentation to choose "optimal" bandwidths by scaling up and down the automatic choices given in Silverman (1986).[14]

5.2 Sampling distributions of estimators

Figures 1A-1F display box-plots with the sampling distributions of the Gaussian- and t-based ML estimators, and the two semiparametric ones. In the case of vector parameters, we report the values corresponding to the third series. As usual, the central boxes describe the first and third quartiles of the sampling distributions, as well as their median. The maximum length of the whiskers is one interquartile range. Finally, we also report the fraction of estimates outside those whiskers to complement the information on the tails of the distributions.

[13] We implicitly impose the restrictions on α and β by numerically maximising the Gaussian and t log-likelihood functions with respect to θ_I^* and θ_{II}^* subject to the restrictions $10^{-4} \leq \theta_I^* \leq .999$ and $0 \leq \theta_{II}^* \leq .999$, where $\beta = \theta_I^*\theta_{II}^*$ and $\alpha = \theta_I^*(1 - \theta_{II}^*)$. Nevertheless, we always compute scores and information bounds in terms of α and β, using the chain rule for derivatives whenever necessary.

[14] We considered .3, .5, .8, 1, 1.25, 1.5, 2, 2.5, 3 and 4 times the bandwidth $[4/(N+2)]^{1/(N+4)} \cdot s \cdot T^{-1/(N+4)}$ recommended by Silverman (1986) for multivariate density estimation under normality, where s^2 is the second sample moment of $\varepsilon_{it}^*(\tilde{\boldsymbol{\theta}}_T)$ averaged across t and i in the case of the SP estimator, and the sample variance of $\sqrt[3]{\varsigma_t(\tilde{\boldsymbol{\theta}}_T)}$ in the case of the SSP estimator. The reported results use scaling factors of 1.25 (SSP) and 2.5 (SP).

As expected from Proposition 7.1, the distribution of the four estimators is essentially identical under normality across all the parameters, with the only exception of the SP estimator of γ_3, which is not very surprising given that the ML and PML are numerically identical over half the time. However, they progressively differ under correct Student t specification as the degrees of freedom decrease.

Another thing to note is that the sampling distributions of the Gaussian PML estimators of π_3 and ρ_3 do not seem to be affected much by the true conditional distribution of the data, which suggests that the different information bounds of the simulated model are almost block diagonal between the conditional mean parameters $(\boldsymbol{\pi}, \boldsymbol{\rho})$ and the rest. The same seems to be true for the SP estimator of π_3, which essentially reflects the fact that there is no SP adjustment for unconditional means. In contrast, the behaviour of the SP estimator of the autoregressive coefficient ρ_3 described in Figure 1B is very much at odds with the theoretical predictions, probably as a result of the fact that the adjustment of this parameter described in (25) becomes very noisy once we replace the unknown score by the one obtained with the multivariate kernel estimator.

On the other hand, the sampling distributions of the SSP and t-based ML estimators of π_3 and ρ_3 are quite sensitive to the nature of the underlying distribution. In particular, when the true distribution is elliptical, the sampling distributions of those estimators are narrower than the distributions of the PML and SP estimators. This is particularly noticeable in the t_4 case, but also in the normal-gamma case, for which the ML estimator should lose its asymptotic efficiency but not its consistency according to Proposition 10. At the same time, an asymmetric distribution introduces substantial positive biases in the ML and SSP estimators of π_3. Intuitively, since the true distribution of the standardised innovations is negatively skewed, those estimators are re-centring their estimated distributions so as to make them more symmetric. Somewhat surprisingly, though, the biases in the unconditional mean seem to go a long way in mopping up the biases in the autocorrelation coefficients. As for time-varying kurtosis, it seems to have little effect on the estimators of the two conditional mean parameters that we analyse, with results that broadly resemble the ones obtained for the t_8.

Unlike what happens with the conditional mean parameters, the sampling distributions of the PML estimators of both the static variance parameters c_3 and γ_3, and the dynamic variance parameters α and β are quite sensitive to the distribution of the innovations. In this sense, the first thing to note is that those sampling distributions deteriorate 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 of the ARCH and GARCH parameters is clearly non-standard, as discussed after Proposition 3. On the other hand, the PML estimators of α and β are the least affected by the existence of time-varying higher order moments. The SP estimators of the conditional variance parameters also suffer when κ_0 increases, becoming substantially downward biased in the case of γ_3, as well as in the case of α when the innovations are t_4.

In contrast, the ML estimators of the conditional variance parameters behave very much as expected: there are substantial efficiency gains when the distribution of the innovations coincides with the assumed one, and some noticeable biases when it does not. However, it is interesting to note that those biases only affect γ_3 and α in the normal-gamma case, and α and β in the time-varying leptokurtic case. The unbiasedness results that we obtain with the asymmetric t are somewhat remarkable, and suggest once again that the biases in the unconditional mean that we observe in Figure 1A adequately re-centre the estimated distribution of the innovations.

The behaviour of the SSP estimators of the conditional variance parameters is mixed. When the distribution is elliptical, this estimator does a reasonably good job, although by no means does it achieve the efficiency of the ML estimator. This is especially true in the case of t_4 innovations, when it also shares a downward bias for α with the SP estimator. Like the ML estimators, though, the SSP estimators also seem somewhat resilient to misspecification, since the only noticeable biases correspond to γ_3 for the asymmetric Student t, and α and β for the t distribution with time-varying degrees of freedom.

Model (40) can be easily reparametrised as in (29) if we ignore the small adjustment term $\omega_{t-j}(\boldsymbol{\theta})$. For instance, we can choose ϑ_2 to be the cross-sectional average of the idiosyncratic variances ($= \boldsymbol{\gamma}'\boldsymbol{\iota}_N/N$), and then re-scale λ, α and the elements of $\boldsymbol{\gamma}$ accordingly. Figures 1G and 1H display box-plots of γ_3/ϑ_2 and α/ϑ_2. As can be seen, the t-based ML estimators of these two derived parameters become consistent when the true distribution is normal-gamma, which confirms Proposition 10.a (see also Thm.1 in Newey and Steigerwald (1997)). But contrary to the asymptotic results in Proposition 8.a, they seem to be at least as efficient as the SSP estimator in that case. Similarly, the SSP estimators also seem to be consistent in the case of the asymmetric Student t, but the downward bias that affects α when the distribution is t_4 continues to contaminate α/ϑ_2.

5.3 Finite sample performance of Hausman tests

Following our discussion on power in section 4.2, we focus our attention on two parameters only: the cross-sectional mean of the unconditional mean parameters $\pi's$ and the cross-sectional mean of the idiosyncratic variances $\gamma's$. In the remaining of this section, we shall refer to those two average parameters as $\bar{\pi}$ and $\bar{\gamma}$. The Wald version of single coefficient tests is straightforward. The LM version is also easy to obtain if we use the results in the proofs of Propositions 13 and

14 to show that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) - \mathcal{A}^{-1}(\boldsymbol{\phi}_0)\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, 0) = o_p(1),$$
$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \mathring{\boldsymbol{\theta}}_T) - \mathcal{A}^{-1}(\boldsymbol{\phi}_0)\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\mathring{\boldsymbol{\theta}}_T, 0) = o_p(1).$$

To simplify the comparisons between parametric and semiparametric testing procedures, we systematically use the PML estimator of $\boldsymbol{\theta}$ in computing the different information bounds. We also use the sequential MM estimator of η in (41), which amounts to replacing κ_0 by its sample analogue when it is positive.[15] We provide further details on how we compute the SSP bound $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ in Appendix B.3.

The first two panels of Table 1 report the fraction of simulations in which the parametric and SSP Hausman tests in Propositions 13 and 14, respectively, exceed the 1, 5 and 10% critical values of a χ_1^2 when the true distribution is a Student t_8, while the last panel reports the corresponding fractions for the SSP test in the normal-gamma case. All tests tend to overreject, but the size distortions of the parametric tests are typically small, especially if compared to the huge distortions shown by the SSP Hausman procedures based on $\bar{\gamma}$. Although the estimators of $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ are noisier than the estimators of $\mathcal{I}(\boldsymbol{\phi}_0)$ or $\mathcal{C}(\boldsymbol{\phi}_0)$, the main problem with the SSP tests is that the difference between the Monte Carlo variances of the PML estimators of $\bar{\pi}$ and $\bar{\gamma}$ and its asymptotically efficient SSP counterparts is smaller than the Monte Carlo variance of the difference between those two estimators, which violates the principle underlying Hausman tests. In fact, the Monte Carlo variance of the SSP estimator of $\bar{\gamma}$ turns out to be higher than that of the PML estimator both in the case of the Student t_8 and the normal-gamma mixture, despite the fact that the Monte Carlo variances of the estimators of the individual $\gamma_i's$ are in the correct order, which suggests that the SSP estimators of the $\gamma_i's$ have a more positive cross-sectional correlation. Monte Carlo experiments with $T = 10,000$ indicate, though, that those problems are mitigated as the first-order asymptotic results become more representative.

Table 2 contains the fraction of simulations in which the parametric (upper panels) and SSP (lower panels) Hausman tests exceed the 1, 5 and 10% empirical critical values obtained by simulation when the true distribution is a Student t_8 (see Table 1).

As expected, the parametric test based on $\bar{\pi}$ has little power when the true distribution is normal-gamma, which is not surprising given that in that case the ML estimators of the conditional mean parameters are consistent, albeit no longer efficient. In contrast, the power is essentially 1 if we base the test on the idiosyncratic variance parameter $\bar{\gamma}$. In the case of the

[15] As we mentioned before, the feasible versions of the Hausman tests will not work properly when $\eta \geq 1/4$ because in that case κ becomes unbounded in the population but not in the sample. Moreover, it may also have poor finite sample properties for $\eta_0 \geq 1/8$ because the asymptotic distribution of $\check{\eta}_T$ will not be root-T consistent in that case (see Amengual, Fiorentini and Sentana (2010) for further details).

asymmetric t, though, the parametric Hausman tests based on the unconditional means have substantially more power than the tests based on the unconditional idiosyncratic variances, which is also in line with the Monte Carlo distributions presented in the previous section. Finally, neither of those parameters is useful to detect a t distribution with time-varying degrees of freedom.

In turn, the SSP Hausman test based on $\bar{\pi}$ and $\bar{\gamma}$ have a lot of power to detect departures in the asymmetric direction, but again no power against time-varying kurtosis. The odd size-adjusted power results observed at the 1% level simply reflect the imprecision of the estimated Monte Carlo critical values.

6 Conclusions

In the context of general multivariate dynamic models with non-zero conditional means and possibly time-varying variances and covariances, we compare the efficiency of the feasible ML procedure that jointly estimates the shape parameters with the efficiency of the Gaussian PML, SP, SSP and infeasible ML estimators of the conditional mean and variance parameters considered in the existing literature. As one would expect, we show that if the standardised innovations are strong white noise with a possibly asymmetric and leptokurtic distribution the ranking is infeasible ML, feasible ML, SP and Gaussian PML. We then particularise our results to elliptical distributions, and show that the efficiency bound of the SSP estimator lies between those of the feasible ML and SP estimators, the second of which in turn is more efficient than the Gaussian PMLE, 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) and Hafner and Rombouts (2007), who look at univariate models and multivariate models with zero means, respectively. By explicitly considering a multivariate framework with non-zero conditional means we are able to cover many empirically relevant applications beyond Arch models, which have been the motivating example for most of the existing work. In particular, our results apply to conditionally homoskedastic, dynamic linear models such as Vars or multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts.

More generally, we show that in the elliptical case the SSP estimator is adaptive for all but one global scale parameter in an appropriate reparametrisation of the model. This result directly generalises the one obtained for univariate Garch models by Linton (1993), as well as the results in Hodgson and Vorkink (2003) for a specific multivariate Garch-M model. We also show that when the conditional distribution is not only leptokurtic or platykurtic but also potentially asymmetric the general SP estimator is adaptive for a much more restricted set of parameters in an alternative reparametrisation that in a conditionally heteroskedastic context only seems to fit the constant conditional correlation model of Bollerslev (1987) when the conditional mean is 0, but which covers the slope coefficients of many conditionally homokedastic multivariate regression models, including VARs. This second result generalises the ones obtained for specific univariate GARCH models by Drost and Klaassen (1997) and Sun and Stengos (2006), which seem overly simple from a multivariate perspective. Importantly, we prove that both semiparametric estimators share a saddle point efficiency property, in that they are as inefficient as the Gaussian PMLE for the parameters that they cannot estimate adaptively.

We also thoroughly analyse the effects of distributional misspecification on the consistency of the conditional mean and variance parameter estimators. In particular, we show that when the true conditional distribution is elliptical but different from the parametric one assumed for estimation purposes, the feasible spherically-based ML estimator is consistent for exactly the same parameters for which the SSP estimator is adaptive, and the same is true when we fix the shape parameters to some a priori chosen value which does not coincide with the true one. This result generalises Theorem 1 in Newey and Steigerwald (1997), which applies to univariate models.

Furthermore, we show that when the conditional distribution is not necessarily spherical, the feasible ML estimator based on a misspecified parametric distribution will be 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. This second result also directly generalises Theorem 2 in Newey and Steigerwald (1997), which again looks misleadingly simple from a multivariate perspective.

In both cases, we also show that the remaining parameters, which the semiparametric procedures can only estimate with the efficiency of the Gaussian PML estimator, will be inconsistently estimated by distributionally misspecified parametric procedures. For that reason, we provide closed-form expressions for consistent estimators of those parameters.

Intuitively, the reparametrisations that we consider are such that both the covariance of the (pseudo) score and the expected Hessian are block diagonal between a subset of the conditional mean and variance parameters and the rest, including those that characterise the shape of the distribution. In turn, this block diagonality leads to full efficiency under correct specification, while under misspecification it protects the estimators of those parameters from inconsistencies in the remaining ones.

In addition, we show that when the conditional distribution is either platykurtic or mesokurtic, so that the coefficient of multivariate excess kurtosis is non-negative, the feasible ML estimators based on certain leptokurtic spherical distributions, including the multivariate Student t and indeed any symmetric generalised hyperbolic distribution, as well some discrete scale mixtures and Laplace expansions of the multivariate normal, provide consistent estimators of all the parameters irrespective of the ellipticity of the true distribution because they converge to the Gaussian PML estimators.

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

In a detailed Monte Carlo experiment we find that there is a substantial difference between the estimation of the following four groups of parameters: (a) the unconditional mean parameters, (b) the unconditional variance parameters, (c) the dynamic mean parameters, and (d) the dynamic variance parameters. We also find that the finite sample performance of the semiparametric procedures is not well approximated by the first-order asymptotic theory that justifies them. This is particularly true of the SP estimators of the dynamic mean and variance parameters, but also affects the SSP estimators of the latter. As for the feasible ML estimators based on a Student t log-likelihood function, we find that they offer substantial efficiency gains relative to the Gaussian PML estimators when the true distribution coincides with the one assumed for estimation purposes, but they may be biased otherwise. Nevertheless, we find that the biases seem to be limited to the unconditional mean parameters when the true distribution is asymmetric, and the variance parameters when it is elliptical but not t. In this second case, our simulation results also confirm that we can obtain consistent estimators of all parameters but one by using one of the reparametrisations previously discussed.

As for the Hausman tests, we find that the one based on the feasible ML estimator works quite well, both in terms of size and power, while the one based on the SSP estimator suffers from substantial size distortions when we base it on the unconditional variance parameters. In this sense, it would be useful to explore bootstrap procedures that exploit the fact that elliptical distributions are parametric in $N-1$ dimensions, and non-parametric in only one.

Further work is required in at least four other 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. For that reason, Mencía and Sentana (2009, 2010) model the asymmetry parameters as a function of the information set in such a way that this dependence disappears. However, this implies that there is no longer a clear separation between the parameters that enter in the first two moments, and those that determine higher order ones. The beta t ARCH model of Harvey and Chakravarty (2008), or the conditionally heteroskedastic factor model of Harvey, Ruiz and Sentana (1992) with elliptically distributed factors are other examples in which this separation also breaks down.

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, empirical researchers are often interested in features of the distribution beyond the first two conditional moments, which implies that one cannot simply treat the shape parameters as if they were nuisance parameters. For that reason, Amengual, Fiorentini and Sentana (2010) consider sequential estimators of the shape parameters, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators. In particular, they consider sequential ML estimators, as well as sequential GMM estimators. 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.

More generally, 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 its quantiles or the probability of the joint occurrence of several events. 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

A 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

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

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

Lemmata

Lemma 1 *Let ς denote a scalar random variable with continuously differentiable density function $h(\varsigma;\boldsymbol{\eta})$ over the possibly infinite domain $[a,b]$, and let $m(\varsigma)$ denote a continuously differentiable function over the same domain such that $E[m(\varsigma)|\boldsymbol{\eta}] = k(\boldsymbol{\eta}) < \infty$. Then*

$$E[\partial m(\varsigma)/\partial\varsigma|\,\boldsymbol{\eta}] = -E[m(\varsigma)\partial \ln h(\varsigma;\boldsymbol{\eta})/\partial\varsigma|\,\boldsymbol{\eta}],$$

as long as the required expectations are defined and bounded.

Proof. If we differentiate

$$k(\boldsymbol{\eta}) = E[m(\varsigma)|\,\boldsymbol{\eta}] = \int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})d\varsigma$$

with respect to ς, we get

$$0=\int_a^b \frac{\partial m(\varsigma)}{\partial \varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma+\int_a^b m(\varsigma)\frac{\partial h(\varsigma;\boldsymbol{\eta})}{\partial \varsigma}d\varsigma=\int_a^b \frac{\partial m(\varsigma)}{\partial \varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma+\int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})\frac{\partial \ln h(\varsigma;\boldsymbol{\eta})}{\partial \varsigma}d\varsigma,$$

as required. □

Lemma 2 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density function $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})$, where $\boldsymbol{\varrho}=\mathbf{0}$ denotes normality, then*

$$E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{e}'_{rt}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\middle|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\}=[\mathcal{K}(\mathbf{0})\,|\mathbf{0}]. \tag{A3}$$

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

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

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 $\boldsymbol{\varrho}$ when regarded as the influence 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 $\mathbf{e}_t(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$. □

Proposition 1

Since the distribution of $\boldsymbol{\varepsilon}_t^*$ given $\mathbf{z}_t$, I_{t-1} is assumed to be *i.i.d.*, then it is easy to see from (2) that $\mathbf{e}_t(\boldsymbol{\phi})=[\mathbf{e}'_{dt}(\boldsymbol{\phi}),\mathbf{e}'_{rt}(\boldsymbol{\phi})]'$ will inherit the martingale difference property of the score $\mathbf{s}_t(\boldsymbol{\phi}_0)$. As a result, the conditional information matrix will be given by

$$\begin{aligned}&\left[\begin{array}{ccc}\mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{array}\right]\left[\begin{array}{ccc}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho})\\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho})\\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho})\end{array}\right]\left[\begin{array}{cc}\mathbf{Z}'_{lt}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{array}\right]\\ &=\left[\begin{array}{cc}\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ll}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}'_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta})+\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{ss}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{lr}(\boldsymbol{\varrho})+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{sr}(\boldsymbol{\varrho})\\ \mathcal{M}'_{lr}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta})+\mathcal{M}'_{sr}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathcal{M}_{rr}(\boldsymbol{\varrho})\end{array}\right],\end{aligned}$$

where

$$\left[\begin{array}{ccc}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho})\\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho})\\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho})\end{array}\right]=V\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})\\ \mathbf{e}_{rt}(\boldsymbol{\theta},\boldsymbol{\varrho})\end{array}\middle|\,\boldsymbol{\theta},\boldsymbol{\varrho}\right],$$

which confirms the variance of the score part of the proposition.

As for the expected value of the Hessian expressions, it is easy to see that

$$E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi})|z_t,I_{t-1};\boldsymbol{\phi}]=\mathbf{Z}_{lt}(\boldsymbol{\theta})E\left[\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}\middle|\,z_t,I_{t-1};\boldsymbol{\phi}\right]+\mathbf{Z}_{st}(\boldsymbol{\theta})E\left[\frac{\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}\middle|\,z_t,I_{t-1};\boldsymbol{\phi}\right]$$

because

$$E\left[\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})|z_t,I_{t-1};\boldsymbol{\phi}\right]=-E\left[\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*|z_t,I_{t-1};\boldsymbol{\phi}\right]=\mathbf{0} \tag{A4}$$

and

$$E\left[\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})|z_t,I_{t-1};\boldsymbol{\phi}\right]=-E\left[vec\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\}|z_t,I_{t-1};\boldsymbol{\phi}\right]=\mathbf{0}. \qquad \text{(A5)}$$

Expression (7) then leads to

$$E\left[\left.\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}\right|z_t,I_{t-1};\boldsymbol{\phi}\right]=E\left[\left.\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}\right|z_t,I_{t-1};\boldsymbol{\phi}\right]$$
$$=E\left[\left.\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\right|\boldsymbol{\phi}\right]\mathbf{Z}_{lt}'(\boldsymbol{\theta})+E\left[\left.\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|\boldsymbol{\phi}\right]\mathbf{Z}_{st}'(\boldsymbol{\theta}).$$

Likewise, equation (8) leads to

$$E\left[\left.\frac{\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}\right|z_t,I_{t-1};\boldsymbol{\phi}\right]=E\left[\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}+\left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}\right.$$
$$\left.\times\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}\right|z_t,I_{t-1};\boldsymbol{\phi}\Big]=E\left[\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\right|\boldsymbol{\phi}\right]\mathbf{Z}_{lt}'(\boldsymbol{\theta})$$
$$+E\left[\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|z_t,I_{t-1};\boldsymbol{\phi}\right]\mathbf{Z}_{st}'(\boldsymbol{\theta})-\mathbf{K}_{NN}\mathbf{Z}_{st}'(\boldsymbol{\theta})$$

because of (A4) and (A5), which in turn implies

$$\begin{aligned}&E\left\{\left.\left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|z_t,I_{t-1};\boldsymbol{\phi}\right\}\\
=\ &\mathbf{K}_{NN}E\left\{\left.\mathbf{K}_{NN}\left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|z_t,I_{t-1};\boldsymbol{\phi}\right\}\\
=\ &\mathbf{K}_{NN}E\left\{\left.\left[\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\otimes\mathbf{I}_N\right][\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|z_t,I_{t-1};\boldsymbol{\phi}\right\}\\
=\ &\mathbf{K}_{NN}E\left\{\left.\left[\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N\right]\right|z_t,I_{t-1};\boldsymbol{\phi}\right\}=-\mathbf{K}_{NN}\end{aligned}$$

in view of Theorem 3.1 in Magnus (1988).

As a result, the information matrix equality implies that

$$\begin{aligned}\mathcal{M}_{ll}(\boldsymbol{\varrho})&=E\left\{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\phi}\right\}\\
\mathcal{M}_{ls}(\boldsymbol{\varrho})&=E\left\{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|\boldsymbol{\phi}\right\}\\
\mathcal{M}_{ss}(\boldsymbol{\varrho})&=E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\right|\boldsymbol{\phi}\right\}-\mathbf{K}_{NN}\end{aligned}$$

Similarly, equation (6) implies that

$$E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi})|z_t,I_{t-1};\boldsymbol{\phi}]=E[\mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'+\mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t,I_{t-1};\boldsymbol{\phi}].$$

But then the information matrix equality together with equations (C37) and (C38) imply that

$$\begin{aligned}E[\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t,I_{t-1};\boldsymbol{\phi}]&=-E\{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\}=\mathcal{M}_{lr}(\boldsymbol{\varrho}),\\
E[\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t,I_{t-1};\boldsymbol{\phi}]&=-E\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\}=\mathcal{M}_{sr}(\boldsymbol{\varrho}).\end{aligned}$$

Finally, the information matrix equality also implies that

$$\mathcal{M}_{rr}(\boldsymbol{\varrho})=-E\{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\},$$

as required. □

Proposition 2

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

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

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

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\delta(\varsigma_t,\boldsymbol{\eta})\sqrt{\varsigma_t}|\boldsymbol{\eta}]\cdot E(\mathbf{u}_t)=\mathbf{0},$$

and that

$$E[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = vec\left\{E\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t|\boldsymbol{\eta}\right]\cdot E(\mathbf{u}_t\mathbf{u}_t')-\mathbf{I}_N\right\} = vec(\mathbf{I}_N)\left\{E\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)(\varsigma_t/N)|\boldsymbol{\eta}\right]-1\right\}.$$

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

$$h(\varsigma_t;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})], \tag{A6}$$

whence

$$[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1] = -\frac{2}{N}\left[1+\varsigma_t\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\right]. \tag{A7}$$

On this basis, we can use Lemma 1 to show that $E(\varsigma_t)=N<\infty$ implies

$$E\left[\varsigma_t\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma|\boldsymbol{\eta}\right] = -E\left[1\right]=-1,$$

which in turn implies that

$$E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1|\boldsymbol{\eta}\right]=0 \tag{A8}$$

in view of (A7). Consequently, $E[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}]=\mathbf{0}$, as required.

Similarly, we can also show that

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

by virtue of (A1), and

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

by virtue of (A2), (A7) and (A8).

Finally, it is clear from (9) that $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ will be a function of ς_t but not of $\mathbf{u}_t$, which immediately implies that $E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}]=\mathbf{0}$, and that

$$
\begin{aligned}
E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{vec\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t\cdot\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N\right]\mathbf{e}'_{rt}(\boldsymbol{\phi})\right\}\\
&= vec(\mathbf{I}_N)E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1\right]\mathbf{e}'_{rt}(\boldsymbol{\phi})\right\}.
\end{aligned}
$$

To obtain the expected value of the Hessian, it is also convenient to write $\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}_0)$ in (13) as

$$
\begin{aligned}
&-4\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{I}_N\otimes\{\delta[\varsigma_t(\boldsymbol{\theta}_0),\eta_0]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)-\mathbf{I}_N\}]\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)\\
&+[\mathbf{e}'_{lt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_p]\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left[\frac{\partial\boldsymbol{\mu}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\\
&+\frac{1}{2}\{\mathbf{e}'_{st}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)]\otimes\mathbf{I}_p\}\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\right\}\\
&-2\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)[\mathbf{e}'_{lt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\otimes\mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{e}_{lt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\otimes\mathbf{I}_N]\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)\\
&-\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)-\frac{2\partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0\right]}{\partial\varsigma}\{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)vec'[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)]\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)]vec'[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)]\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)\}.
\end{aligned}
$$

Clearly, the first four lines have zero conditional expectation, and the same is true of the sixth line by virtue of (A1). As for the remaining terms, we can write them as

$$
\begin{aligned}
&-\delta(\varsigma_t,\boldsymbol{\eta}_0)\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)-2\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma\cdot\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)\\
&-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)-2\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma\cdot\varsigma_t^2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{u}_t\mathbf{u}'_t)vec'(\mathbf{u}_t\mathbf{u}'_t)\mathbf{Z}'_{st}(\boldsymbol{\theta}_0),
\end{aligned}
$$

whose conditional expectation will be

$$
\begin{aligned}
&-\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}_0)E[\delta(\varsigma_t;\boldsymbol{\eta}_0)+2(\varsigma_t/N)\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0]-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}_0)\\
&-\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\frac{2E[\varsigma_t^2\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0]}{N(N+2)}[(\mathbf{I}_{N^2}\otimes\mathbf{K}_{NN})+vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]\mathbf{Z}'_{st}(\boldsymbol{\theta}_0).
\end{aligned}
$$

As for $\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}_0)$, it follows from (14) and (B33) that we can write it as

$$\begin{aligned} & \{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) + \mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right]\} \cdot \partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\boldsymbol{\eta}' \\ = \; & [\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{u}_t\sqrt{\varsigma_t} + \mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{u}_t\mathbf{u}_t')\varsigma_t] \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta})/\partial\boldsymbol{\eta}', \end{aligned}$$

whose conditional expected value will be $\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{I}_N)E[(\varsigma_t/N) \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}'|\boldsymbol{\eta}]$. □

Proposition 3

The proof of the first part is based on a straightforward application of Proposition 1 in Bollerslev and Wooldridge (1992) to the *i.i.d.* 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 $V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0]$, which justifies (19). Further, we will have that

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

for any spherical distribution, with ς_t and $\mathbf{u}_t$ both mutually and serially independent. Then (20) 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 the expression for $\mathcal{B}_t(\boldsymbol{\phi}_0)$ under multivariate normality ($\boldsymbol{\varrho}_0 = \mathbf{0}$) by the (conditional) information matrix equality. □

Proposition 4

It trivially follows from (19) and (A3) that

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

for any distribution. In addition, we also know that

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

Hence, the second summand of (25), 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 (25) as

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

then we can use the law of iterated expectations to show that the semiparametric efficient score (25) 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 $E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta}, \boldsymbol{\varrho})|\boldsymbol{\theta}, \boldsymbol{\varrho}\right] = \mathbf{0}$.

Finally, the expression for the semiparametric efficiency bound will be

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

by virtue of (19), (A3) and the law of iterated expectations. □

Proposition 5

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

$$\begin{aligned}\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,\mathbf{0}) &= E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta}_0,\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta}_0,\mathbf{0})\end{array}\middle|\varsigma_t;\boldsymbol{\phi}_0\right]=E\left\{\begin{array}{c}\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\\ vec\left[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)-\mathbf{I}_N\right]\end{array}\middle|\varsigma_t;\boldsymbol{\phi}_0\right\}\\ &= E\left[\begin{array}{c}\sqrt{\varsigma_t}\mathbf{u}_t\\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N)\end{array}\middle|\varsigma_t\right]=\left(\frac{\varsigma_t}{N}-1\right)\left[\begin{array}{c}\mathbf{0}\\ vec(\mathbf{I}_N)\end{array}\right],\end{aligned}\tag{A9}$$

and

$$\begin{aligned}\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)=E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi}_0)\\ \mathbf{e}_{st}(\boldsymbol{\phi}_0)\end{array}\middle|\varsigma_t;\boldsymbol{\phi}_0\right]\\ =E\left\{\begin{array}{c}\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\cdot\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\\ vec\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\cdot\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)-\mathbf{I}_N\right]\end{array}\middle|\varsigma_t;\boldsymbol{\phi}_0\right\}\\ =E\left\{\begin{array}{c}\delta(\varsigma_t,\boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t\\ vec[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N]\end{array}\middle|\varsigma_t\right\}=\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\left[\begin{array}{c}\mathbf{0}\\ vec(\mathbf{I}_N)\end{array}\right],\end{aligned}\tag{A10}$$

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 ς_t 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

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]\middle|\boldsymbol{\eta}\right\}$$

and

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

In this respect, we can use (21) 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 Lemma 1 to show that $E(\varsigma_t^2) = N(N+2)(\kappa+1) < \infty$ implies

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

If we then combine this result with (A7) and (A8), we will have that for any spherically symmetric distribution

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]\middle|\boldsymbol{\eta}\right\} = \frac{2}{N},$$

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

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

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

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

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 (27) 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 (27) 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

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\begin{array}{c}\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\mathring{\mathcal{K}}^+(\kappa)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\right\} \\ \times\left\{\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^+(\kappa)\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_d(\boldsymbol{\phi})\right\}\end{array}\middle|\boldsymbol{\phi}\right]$$

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

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

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

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

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

by virtue of the law of iterated expectations. □

Proposition 6

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 $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_0)$, 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 $\mathcal{S}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\theta})$ follows from the fact that it coincides with the residual covariance matrix in the theoretical regression of the semiparametric efficient score on the Gaussian pseudo-score since

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

because $\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})$ is conditionally orthogonal to $[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^+(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})]$ by construction. □

Proposition 7

The proof of the first part is trivial, except perhaps for the fact that $\mathcal{M}_{sr}(\mathbf{0}) = \mathbf{0}$, which follows from Lemma 2 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 $\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t/N-1$ 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

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

which in view of (A6) implies that

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

i.e. the density of Gamma random variable with mean N and variance $N[(N+2)\kappa_0+2]$. In this sense, it is worth recalling that $\kappa \geq -2/(N+2)$ for all elliptical distributions, with the lower limit corresponding to the uniform.

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

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

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 $\delta(\varsigma_t,\boldsymbol{\eta}_0)\boldsymbol{\varepsilon}_t^*$ on $\boldsymbol{\varepsilon}_t^*$ is zero for all t, or equivalently, that $\delta(\varsigma_t,\boldsymbol{\eta}_0) = 1$. But since the solution to this differential equation is $g(\varsigma_t,\boldsymbol{\eta}) = -.5\varsigma_t + C$, then the result follows from (A6). □

Proposition 8

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

$$\mathbf{Z}_{dt}(\boldsymbol{\vartheta}) = \left\{\begin{array}{cc} \vartheta_2^{-1/2}\left[\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_1)/\partial\boldsymbol{\vartheta}_1\right]\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1) & \frac{1}{2}\{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_1)]/\partial\boldsymbol{\vartheta}_1\}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)] \\ 0 & \frac{1}{2}\vartheta_2^{-1}vec'(\mathbf{I}_N)\end{array}\right\} = \left[\begin{array}{cc} \mathbf{Z}_{\boldsymbol{\vartheta}_1lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\vartheta_2st}(\boldsymbol{\vartheta})\end{array}\right]. \quad \text{(A11)}$$

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

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\vartheta}_1t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ s_{\vartheta_2t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right] = \left[\begin{array}{c} \mathbf{Z}_{\boldsymbol{\vartheta}_1lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + \mathbf{Z}_{\boldsymbol{\vartheta}_1st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ \mathbf{Z}_{\vartheta_2st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right], \quad \text{(A12)}$$

where $\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ are given in (10) and (11), respectively.

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\vartheta_1t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $s_{\vartheta_2t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is

$$E\left\{\left[\begin{array}{cc}\mathbf{Z}_{\boldsymbol{\vartheta}_1lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1st}(\boldsymbol{\vartheta})\end{array}\right]\left[\begin{array}{cc}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{array}\right]\left[\begin{array}{c} 0 \\ \mathbf{Z}'_{\vartheta_2st}(\boldsymbol{\vartheta})\end{array}\right]\middle|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\}$$

$$= \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta}) + N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_2}E\left\{\frac{1}{2}\frac{\partial vec'[\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)]}{\partial\boldsymbol{\vartheta}_1}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)]\middle|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\}vec(\mathbf{I}_N)$$

$$= \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta}) + N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_2}\mathbf{Z}_{\boldsymbol{\vartheta}_1s}(\boldsymbol{\vartheta},\boldsymbol{\eta})vec(\mathbf{I}_N),$$

with $\mathbf{Z}_{\boldsymbol{\vartheta}_1s}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = E[\mathbf{Z}_{\boldsymbol{\vartheta}_1st}(\boldsymbol{\vartheta})|\boldsymbol{\vartheta},\boldsymbol{\eta}]$, 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 2.

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

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

Hence, the residuals from the unconditional regression of $\mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ 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 $\mathbf{s}_{\vartheta_1|\vartheta_2 t}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ is clearly conditionally orthogonal to any function of $\varsigma_t(\boldsymbol{\vartheta}_0)$. In contrast, the second term is not conditionally orthogonal to functions of $\varsigma_t(\boldsymbol{\vartheta}_0)$, but since the conditional covariance between any such function and $\mathbf{e}_{st}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ will be time-invariant, it will be unconditionally orthogonal by the law of iterated expectations. As a result, $\mathbf{s}_{\boldsymbol{\vartheta}_1|\vartheta_2 t}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ 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.

To prove Part 1b, note that Proposition 5 and (A11) imply that the elliptically symmetric semiparametric efficient score corresponding to ϑ_2 will be given by

$$
\begin{aligned}
\mathring{s}_{\vartheta_2 t}(\boldsymbol{\vartheta}) &= -\frac{1}{2\vartheta_2}vec'(\mathbf{I}_N)vec\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta})-\mathbf{I}_N\right\} \\
&-\frac{N}{2\vartheta_2}\left\{\left[\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]-\frac{2}{(N+2)\kappa+2}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]\right\} \\
= \frac{1}{2\vartheta_2}\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\vartheta})-N\} &- \frac{N}{2\vartheta_2}\left\{\left[\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]-\frac{2}{(N+2)\kappa+2}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]\right\} \\
&= \frac{N}{\vartheta_2[(N+2)\kappa+2]}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right].
\end{aligned}
$$

But since the iterated elliptically symmetric semiparametric estimator of $\boldsymbol{\vartheta}$ must set to 0 the sample average of this modified score, it must be the case that $\sum_{t=1}^T \varsigma_t(\mathring{\boldsymbol{\vartheta}}_T)=\sum_{t=1}^T \varsigma_t^\circ(\mathring{\boldsymbol{\vartheta}}_{1T})/\mathring{\vartheta}_{2T}=NT$, which is equivalent to (31).

To prove Part 1c note that

$$
\mathbf{s}_{\vartheta_2 t}(\boldsymbol{\vartheta},\mathbf{0})=\frac{1}{2\vartheta_2}\left[\varsigma_t(\boldsymbol{\vartheta})-N\right] \tag{A13}
$$

is proportional to the elliptically symmetric semiparametric efficient score $\mathring{s}_{\vartheta_2 t}(\boldsymbol{\vartheta})$, which means that the residual covariance matrix in the theoretical regression of this efficient score on the

Gaussian score will have rank $p-1$ at most. But this residual covariance matrix coincides with $\mathring{\mathcal{S}}(\boldsymbol{\phi})-\mathcal{A}(\boldsymbol{\phi})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\phi})$ since

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}]=\mathcal{A}(\boldsymbol{\theta}) \tag{A14}$$

because the regression residual

$$\left[\delta(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)$$

is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ by the law of iterated expectations, as shown in the proof of proposition 5.

Tedious algebraic manipulations that exploit the block-triangularity of (A11) and the constancy of $\mathbf{Z}_{\vartheta_2 st}(\boldsymbol{\vartheta})$ show that the different information matrices will be block diagonal when $\mathbf{W}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\phi}_0)$ is 0. Then, part 2a follows from the fact that $\mathbf{W}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\phi}_0)=-E\left\{\partial d_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\vartheta}_1|\,\boldsymbol{\phi}_0\right\}$ will trivially be 0 if (30) holds.

Finally, to prove Part 2b note that (A13) implies that the Gaussian PMLE will also satisfy (31). But since the asymptotic covariance matrices in both cases will be block-diagonal between $\boldsymbol{\vartheta}_1$ and ϑ_2 when (30) holds, the effect of estimating $\boldsymbol{\vartheta}_1$ becomes irrelevant. $\square$

Proposition 9

We can directly work in terms of the $\boldsymbol{\psi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Given the specification for the conditional mean and variance in (33), 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

$$\begin{bmatrix}\mathbf{s}_{\boldsymbol{\psi}_1 t}(\boldsymbol{\psi},\boldsymbol{\varrho})\\ \mathbf{s}_{\boldsymbol{\psi}_2 t}(\boldsymbol{\psi},\boldsymbol{\varrho})\\ \mathbf{s}_{\boldsymbol{\psi}_3 t}(\boldsymbol{\psi},\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi})\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho})+\mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi})\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\boldsymbol{\psi}_2 st}(\boldsymbol{\psi})\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\boldsymbol{\psi}_3 lt}(\boldsymbol{\psi})\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho})\end{bmatrix}, \tag{A15}$$

where

$$\left.\begin{aligned}
\mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi})&=\left\{\partial\boldsymbol{\mu}^{\diamond\prime}_t(\boldsymbol{\psi}_1)/\partial\boldsymbol{\psi}_1+\partial vec'[\boldsymbol{\Sigma}^{\diamond 1/2}_t(\boldsymbol{\psi}_1)]/\partial\boldsymbol{\psi}_1\cdot(\boldsymbol{\psi}_3\otimes\mathbf{I}_N)\right\}\boldsymbol{\Sigma}^{\diamond-1/2\prime}_t(\boldsymbol{\psi}_1)\boldsymbol{\Psi}^{-1/2\prime}_2,\\
\mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi})&=\partial vec'[\boldsymbol{\Sigma}^{\diamond 1/2}_t(\boldsymbol{\psi}_1)]/\partial\boldsymbol{\psi}_1\cdot[\boldsymbol{\Psi}^{1/2}_2\otimes\boldsymbol{\Sigma}^{\diamond-1/2\prime}_t(\boldsymbol{\psi}_1)\boldsymbol{\Psi}^{-1/2\prime}_2],\\
\mathbf{Z}_{\boldsymbol{\psi}_2 st}(\boldsymbol{\psi})&=\partial vec'(\boldsymbol{\Psi}^{1/2})/\partial\boldsymbol{\psi}_2\cdot(\mathbf{I}_N\otimes\boldsymbol{\Psi}^{-1/2\prime}_2)=\mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}),\\
\mathbf{Z}_{\boldsymbol{\psi}_3 lt}(\boldsymbol{\psi})&=\boldsymbol{\Psi}^{-1/2\prime}_2=\mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi}),
\end{aligned}\right\} \tag{A16}$$

$\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho})$ are given in (4), with

$$\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi})=\boldsymbol{\Psi}^{-1/2}_2\boldsymbol{\Sigma}^{\diamond-1/2}_t(\boldsymbol{\psi}_1)[\mathbf{y}_t-\boldsymbol{\mu}^{\diamond}_t(\boldsymbol{\psi}_1)-\boldsymbol{\Sigma}^{\diamond 1/2}_t(\boldsymbol{\psi}_1)\boldsymbol{\psi}_3]. \tag{A17}$$

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\boldsymbol{\psi}_1 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and the remaining elements of the score will be given by

$$\begin{bmatrix}\mathbf{Z}_{\boldsymbol{\psi}_1 l}(\boldsymbol{\psi},\boldsymbol{\varrho}) & \mathbf{Z}_{\boldsymbol{\psi}_1 s}(\boldsymbol{\psi},\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{Z}'_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi})\\ \mathbf{Z}'_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}) & \mathbf{0}\end{bmatrix}$$

with $\mathbf{Z}_{\boldsymbol{\psi}_1 l}(\boldsymbol{\psi},\boldsymbol{\varrho}) = E[\mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi})|\boldsymbol{\psi},\boldsymbol{\varrho}]$ and $\mathbf{Z}_{\boldsymbol{\psi}_1 s}(\boldsymbol{\psi},\boldsymbol{\varrho}) = E[\mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi})|\boldsymbol{\psi},\boldsymbol{\varrho}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$ and the constancy of $\mathbf{Z}_{\boldsymbol{\psi}_2 st}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_3 lt}(\boldsymbol{\psi})$, together with the law of iterated expectations and the definition

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

Similarly, the unconditional covariance matrix of $\mathbf{s}_{\boldsymbol{\psi}_2 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{s}_{\boldsymbol{\psi}_3 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ will be

$$\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}) \\ \mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi}) & \mathbf{0} \end{array}\right]\left[\begin{array}{cc} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}'_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi}) \\ \mathbf{Z}'_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}) & \mathbf{0} \end{array}\right].$$

Hence, the residuals from the unconditional least squares projection of $\mathbf{s}_{\boldsymbol{\psi}_1 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ on $\mathbf{s}_{\boldsymbol{\psi}_2 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{s}_{\boldsymbol{\psi}_3 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ will be:

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\psi}_1|\boldsymbol{\psi}_2,\boldsymbol{\psi}_3 t}(\boldsymbol{\psi},\boldsymbol{\varrho}) &= \mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi})\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) + \mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi})\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ &\quad - \left[\begin{array}{cc} \mathbf{Z}_{\boldsymbol{\psi}_1 l}(\boldsymbol{\psi},\boldsymbol{\varrho}) & \mathbf{Z}_{\boldsymbol{\psi}_1 s}(\boldsymbol{\psi},\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}) \end{array}\right] \\ &= [\mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi}) - \mathbf{Z}_{\boldsymbol{\psi}_1 l}(\boldsymbol{\psi},\boldsymbol{\varrho})]\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) + [\mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi}) - \mathbf{Z}_{\boldsymbol{\psi}_1 s}(\boldsymbol{\psi},\boldsymbol{\varrho})]\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}), \end{aligned}$$

because both $\mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi})$ have full row rank when $\boldsymbol{\Psi}_2$ has full rank in view of the discussion that follows expression (B36).

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

To prove Part 1b note that the semiparametric efficient scores corresponding to $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$ will be given by

$$\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}) \\ \mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi}) & \mathbf{0} \end{array}\right]\mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho}_0)\left\{\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) \\ vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) - \mathbf{I}_N] \end{array}\right\}$$

because $\mathbf{Z}_{\boldsymbol{\psi}_2 st}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\vartheta})$ and $\mathbf{Z}_{\boldsymbol{\psi}_3 lt}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\vartheta})\ \forall t$. But if (36) and (37) hold, then the sample averages of $\mathbf{e}_{lt}[\boldsymbol{\psi}_1,\boldsymbol{\psi}_{2T}(\boldsymbol{\psi}_1),\boldsymbol{\psi}_{3T}(\boldsymbol{\psi}_1);\mathbf{0}]$ and $\mathbf{e}_{st}[\boldsymbol{\psi}_1,\boldsymbol{\psi}_{2T}(\boldsymbol{\psi}_1),\boldsymbol{\psi}_{3T}(\boldsymbol{\psi}_1);\mathbf{0}]$ will be 0, and the same is true of the semiparametric efficient score.

To prove Part 1c note that

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\psi}_2 t}(\boldsymbol{\psi},\mathbf{0}) \\ \mathbf{s}_{\boldsymbol{\psi}_3 t}(\boldsymbol{\psi},\mathbf{0}) \end{array}\right] = \left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\psi}_2 s}(\boldsymbol{\psi}) \\ \mathbf{Z}_{\boldsymbol{\psi}_3 l}(\boldsymbol{\psi}) & \mathbf{0} \end{array}\right]\left[\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) \\ vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) - \mathbf{I}_N] \end{array}\right], \tag{A18}$$

which implies that the residual covariance matrix in the theoretical regression of the semiparametric efficient score on the Gaussian score will have rank $p - N(N+3)/2$ at most because both

$\mathbf{Z}_{\psi_2 s}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\psi_3 l}(\boldsymbol{\psi})$ have full row rank when $\boldsymbol{\Psi}_2$ has full rank. But as we saw in the proof of Proposition 6, that residual covariance matrix coincides with $\mathcal{S}(\boldsymbol{\phi}_0)-\mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\theta})$.

Tedious algebraic manipulations that exploit the block structure of (A16) and the constancy of $\mathbf{Z}_{\psi_2 st}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\psi_3 lt}(\boldsymbol{\psi})$ show that the different information matrices will be block diagonal when $\mathbf{Z}_{\psi_1 l}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{Z}_{\psi_1 s}(\boldsymbol{\psi},\boldsymbol{\varrho})$ are both 0. But those are precisely the necessary and sufficient conditions for $\mathbf{s}_{\psi_1 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ to be equal to $\mathbf{s}_{\psi_1|\psi_2,\psi_3 t}(\boldsymbol{\psi},\boldsymbol{\varrho})$, which is also guaranteed by (35). In this sense, please note that the reparametrisation of $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$ associated with (35) will be such that the Jacobian matrix of $vech[\mathbf{K}^{-1/2}(\boldsymbol{\psi}_1)\boldsymbol{\Psi}_2\mathbf{K}^{-1/2\prime}(\boldsymbol{\psi}_1)]$ and $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_1)\boldsymbol{\psi}_3-\mathbf{l}(\boldsymbol{\psi}_1)$ with respect to $\boldsymbol{\psi}$ evaluated at the true values is equal to

$$\left\{-V^{-1}\left[\begin{array}{c}\mathbf{s}_{\psi_{2t}}(\boldsymbol{\psi}_0)\\ \mathbf{s}_{\psi_{3t}}(\boldsymbol{\psi}_0)\end{array}\middle|\boldsymbol{\phi}_0\right]E\left[\begin{array}{c}\mathbf{s}_{\psi_{2t}}(\boldsymbol{\psi}_0)\mathbf{s}'_{\psi_{1t}}(\boldsymbol{\psi}_0)\\ \mathbf{s}_{\psi_{3t}}(\boldsymbol{\psi}_0)\mathbf{s}'_{\psi_{1t}}(\boldsymbol{\psi}_0)\end{array}\middle|\boldsymbol{\phi}_0\right]\middle|\begin{array}{c|c}\mathbf{I}_{N(N+1)/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_N\end{array}\right\}.$$

Finally, to prove Part 2b simply note that (A18) implies that the Gaussian PMLE will also satisfy (36) and (37). But since the asymptotic covariance matrices in both cases will be block-diagonal between $\boldsymbol{\psi}_1$ and $(\boldsymbol{\psi}_2,\boldsymbol{\psi}_3)$ when (35) holds, the effect of estimating $\boldsymbol{\psi}_1$ becomes irrelevant. □

Proposition 10

As in the proof of Proposition 8, we can directly work in terms of the $\boldsymbol{\vartheta}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_s(.)$. Let us initially keep $\boldsymbol{\eta}$ fixed to some admissible value. The elliptically symmetric score vector for the remaining parameters will then be given by (A12). But since

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

so that

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

we will have that

$$\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{10},\vartheta_2,\boldsymbol{\eta})=\delta[(\vartheta_{20}/\vartheta_2)\varsigma_t,\eta]\sqrt{\vartheta_{20}/\vartheta_2}\boldsymbol{\varepsilon}_t^*=\delta[(\vartheta_{20}/\vartheta_2)\varsigma_t,\boldsymbol{\eta}]\sqrt{\vartheta_{20}/\vartheta_2}\sqrt{\varsigma_t}\mathbf{u}_t,$$
$$\mathbf{e}_{st}(\boldsymbol{\vartheta}_{10},\vartheta_2,\boldsymbol{\eta})=vec\left[\delta[(\vartheta_{20}/\vartheta_2)\varsigma_t,\boldsymbol{\eta}](\vartheta_{20}/\vartheta_2)\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}-\mathbf{I}_N\right]=vec\left[\delta[(\vartheta_{20}/\vartheta_2)\varsigma_t,\boldsymbol{\eta}](\vartheta_{20}/\vartheta_2)\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right].$$

Then, it follows that $E[\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{10},\vartheta_2,\boldsymbol{\eta})|\mathbf{z}_t,I_{t-1};\boldsymbol{\varphi}_0]=\mathbf{0}$ regardless of ϑ_2 and $\boldsymbol{\eta}$ 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,

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

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

$$E\left[\delta\{[\vartheta_{20}/\vartheta_2(\boldsymbol{\eta})]\varsigma_t, \boldsymbol{\eta}\}[\vartheta_{20}/\vartheta_2(\boldsymbol{\eta})](\varsigma_t/N) - 1|\,\boldsymbol{\varphi}_0\right] = 0, \quad \text{(A19)}$$

which we assume is positive, then it is straightforward to show that

$$E\{\mathbf{s}_{\boldsymbol{\vartheta}t}[\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\boldsymbol{\eta}), \boldsymbol{\eta}]|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0\} = \mathbf{0}, \quad \text{(A20)}$$

which means that $\boldsymbol{\vartheta}_{10}$ and $\vartheta_{2\infty}(\boldsymbol{\eta})$ will be the pseudo-true values of the parameters corresponding to a restricted PML estimator that keeps $\boldsymbol{\eta}$ fixed.

If instead we choose $\boldsymbol{\eta}_\infty$ as the solution to the implicit equation

$$E\{\mathbf{s}_{\boldsymbol{\eta}t}[\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\boldsymbol{\eta}), \boldsymbol{\eta}]|\boldsymbol{\varphi}_0\} = \mathbf{0},$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\boldsymbol{\eta}_\infty)$ and $\boldsymbol{\eta}_\infty$ will be the pseudo-true values of the parameters corresponding to an unrestricted PMLE that also estimates $\boldsymbol{\eta}$. In addition, since $\mathbf{s}_{\boldsymbol{\eta}t}[\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}(\boldsymbol{\eta}), \boldsymbol{\eta}]$ only depends on $\varsigma_t(\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty})$, which is *i.i.d.* over time, we will have that

$$E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty}, \boldsymbol{\eta}_\infty)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0] = \mathbf{0}, \quad \text{(A21)}$$

which confirms the martingale difference nature of the elliptical score evaluated at the pseudo-true values.

To obtain the variance of the elliptically symmetric score under misspecification, we can follow exactly the same steps as in the proof of Proposition 2 by exploiting the fact that (A20) and (A21) hold at the pseudo-true parameter values $\boldsymbol{\phi}_\infty$.

These conditions also allow us to obtain the expected value of the Hessian along the lines of Proposition 2.

As we mentioned in the proof of Proposition (8), we can tediously show that the condition for block-diagonality of the expected value of the Hessian and the covariance matrix of the score is $E[\mathbf{W}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty})|\boldsymbol{\varphi}_0] = \mathbf{0}$. But this condition will be satisfied if (30) holds because $\mathbf{W}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}_{10}, \vartheta_{2\infty})$ coincides with $\mathbf{W}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}_{10}, \vartheta_{20})$ in view of (A11). □

Proposition 11

As in the proof of Proposition 9, we can directly work in terms of the $\boldsymbol{\psi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Let us initially keep $\boldsymbol{\varrho}$ fixed to some admissible value. The parametric score vector for the remaining parameters will then be given by (2), with $\mathbf{Z}_{\boldsymbol{\psi}_2 lt}(\boldsymbol{\psi}) = \mathbf{0}$, $\mathbf{Z}_{\boldsymbol{\psi}_3 st}(\boldsymbol{\psi}) = \mathbf{0}$ and the remaining elements in (A16).

Since we are systematically working with lower triangular square root decompositions, we can write

$$\begin{aligned} \mathbf{Z}_{\psi_1 st}(\psi) &= \partial vech'[\Sigma_t^{\diamond 1/2}(\psi_1)]/\partial\psi_1 \cdot \mathbf{L}_N[\Psi_2^{1/2} \otimes \Sigma_t^{\diamond -1/2\prime}(\psi_1)\Psi_2^{-1/2\prime}], \\ \mathbf{Z}_{\psi_2 s}(\psi) &= \partial vech'(\Psi_2^{1/2})/\partial\psi_2 \cdot \mathbf{L}_N[\mathbf{I}_N \otimes \Psi_2^{-1/2\prime}], \end{aligned}$$

where $\mathbf{L}_N$ is the elimination matrix of order N (see Magnus (1988)), which is such that $vec(\mathbf{A}) = \mathbf{L}_N' vech(\mathbf{A})$ for any $N \times N$ lower triangular matrix $\mathbf{A}$.

Given that $\Psi_2^{1/2\prime}$ is upper triangular, $\Psi_2^{-1/2}\Sigma_t^{\diamond -1/2}(\psi_1)$ is lower triangular and $\mathbf{I}_N$ is diagonal, Theorem 5.7.i in Magnus (1988) implies that

$$\begin{aligned} [\Psi_2^{1/2\prime} \otimes \Psi_2^{-1/2}\Sigma_t^{\diamond -1/2}(\psi_1)]\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N[\Psi_2^{1/2\prime} \otimes \Psi_2^{-1/2}\Sigma_t^{\diamond -1/2}(\psi_1)]\mathbf{L}_N', \\ [\mathbf{I}_N \otimes \Psi_2^{-1/2}]\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N[\mathbf{I}_N \otimes \Psi_2^{-1/2}]\mathbf{L}_N', \end{aligned}$$

whence

$$\begin{aligned} \mathbf{Z}_{\psi_1 st}(\psi) &= \partial vech'[\Sigma_t^{\diamond 1/2}(\psi_1)]/\partial\psi_1 \cdot \mathbf{L}_N[\Psi_2^{1/2} \otimes \Sigma_t^{\diamond -1/2\prime}(\psi_1)\Psi_2^{-1/2\prime}]\mathbf{L}_N'\mathbf{L}_N, \\ \mathbf{Z}_{\psi_2 s}(\psi) &= \partial vech'(\Psi_2^{1/2})/\partial\psi_2 \cdot \mathbf{L}_N[\mathbf{I}_N \otimes \Psi_2^{-1/2\prime}]\mathbf{L}_N'\mathbf{L}_N. \end{aligned}$$

As a result,

$$\begin{aligned} \mathbf{s}_{\psi_2 t}(\psi, \varrho) &= -\partial vech'(\Psi_2^{1/2})/\partial\psi_2 \cdot \mathbf{L}_N[\mathbf{I}_N \otimes \Psi_2^{-1/2\prime}]\mathbf{L}_N' vech\left\{\mathbf{I}_N + \frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*}\varepsilon_t^{*\prime}(\psi)\right\} \\ \mathbf{s}_{\psi_3 t}(\psi, \varrho) &= -\Psi_2^{-1/2\prime}\frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*} \end{aligned}$$

and

$$\begin{aligned} \mathbf{s}_{\psi_1 t}(\psi, \varrho) = &\left\{\partial\mu_t^{\diamond\prime}(\psi_1)/\partial\psi_1 + \partial vec'[\Sigma_t^{\diamond 1/2}(\psi_1)]/\partial\psi_1 \cdot (\psi_3 \otimes \mathbf{I}_N)\right\}\Sigma_t^{\diamond -1/2\prime}(\psi_1)\mathbf{s}_{\psi_3 t}(\psi, \varrho) \\ &-\partial vech'[\Sigma_t^{\diamond 1/2}(\psi_1)]/\partial\psi_1 \cdot \mathbf{L}_N[\Psi_2^{1/2} \otimes \Sigma_t^{\diamond -1/2\prime}(\psi_1)\Psi_2^{-1/2\prime}]\mathbf{L}_N' vech\left\{\mathbf{I}_N + \frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*}\varepsilon_t^{*\prime}(\psi)\right\} \end{aligned}$$

since $vech(\mathbf{A}) = \mathbf{L}_N vec(\mathbf{A})$ for any $N \times N$ square matrix $\mathbf{A}$ regardless of its structure.

Let $\psi_{2\infty}(\varrho)$ and $\psi_{3\infty}(\varrho)$ denote the solution to the implicit system of $N + N(N+1)/2$ equations

$$\left.\begin{aligned} E\{\mathbf{s}_{\psi_2 t}[\psi_{10}, \psi_{2\infty}(\varrho), \psi_{3\infty}(\varrho), \varrho]|\varphi_0\} = \mathbf{0} \\ E\{\mathbf{s}_{\psi_3 t}[\psi_{10}, \psi_{2\infty}(\varrho), \psi_{3\infty}(\varrho), \varrho]|\varphi_0\} = \mathbf{0} \end{aligned}\right\}, \tag{A22}$$

which we assume is such that $\Psi_{2\infty}(\varrho)$ is p.d. Given that

$$\varepsilon_t^*(\psi) = \Psi_2^{-1/2}\Sigma_t^{\diamond -1/2}(\psi_1)[\mathbf{y}_t - \mu_t^{\diamond}(\psi_1) - \Sigma_t^{\diamond 1/2}\psi_3],$$

so that

$$\varepsilon_t^*(\psi_{10}, \psi_2, \psi_3) = \Psi_2^{-1/2}(\psi_{30} - \psi_3) + \Psi_2^{-1/2}\Psi_{20}^{1/2}\varepsilon_t^*,$$

we can immediately see that this variable will be $i.i.d.[\mathbf{\Psi}_2^{-1/2}(\boldsymbol{\psi}_{30}-\boldsymbol{\psi}_3), \mathbf{\Psi}_2^{-1/2}\mathbf{\Psi}_{20}\mathbf{\Psi}_2^{-1/2\prime}]$ conditional on $\mathbf{z}_t$ and I_{t-1}. This, together with the full rank of $\mathbf{\Psi}_2^{-1/2\prime}$ implies that

$$E\left[\left.\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho})]; \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right| z_t, I_{t-1}, \boldsymbol{\varphi}_0\right] = \mathbf{0}.$$

In addition, we know from Theorem 5.6 in Magnus (1988) that the matrix

$$\mathbf{L}_N[\mathbf{I}_N \otimes \mathbf{\Psi}_2^{-1/2\prime}]\mathbf{L}_N'$$

will be upper triangular of full rank. Similarly, given that we have defined $\boldsymbol{\psi}_2 = vech(\mathbf{\Psi}_2)$, the matrix $\partial vech'(\mathbf{\Psi}_2^{1/2})/\partial\boldsymbol{\psi}_2$ would also be of full rank in view of the discussion that follows expression (B36).

As a result, we will also have that

$$vech\left\{E\left[\mathbf{I}_N + \left.\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho})]; \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho})]\right| z_t, I_{t-1}, \boldsymbol{\varphi}_0\right]\right\} = \mathbf{0}.$$

Consequently, we will have that

$$E\{\mathbf{s}_{\boldsymbol{\psi}t}[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0\} = \mathbf{0}, \qquad \text{(A23)}$$

which confirms that $\boldsymbol{\psi}_{10}$, $\boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho})$ and $\boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho})$ will be the pseudo-true values corresponding to a restricted PML estimator that keeps $\boldsymbol{\varrho}$ fixed.

If instead we choose $\boldsymbol{\varrho}_\infty$ as the solution to the q equations

$$E\{\mathbf{s}_{\boldsymbol{\varrho}t}[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|\boldsymbol{\varphi}_0\} = \mathbf{0}$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\psi}_{10}$, $\boldsymbol{\psi}_{2\infty} = \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}_\infty)$, $\boldsymbol{\psi}_{3\infty} = \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho}_\infty)$ and $\boldsymbol{\varrho}_\infty$ will be the pseudo-true values of the parameters corresponding to an unrestricted PMLE that also estimates $\boldsymbol{\varrho}$. In addition, since $\mathbf{s}_{\boldsymbol{\varrho}t}[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]$ only depends on $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}, \boldsymbol{\psi}_{3\infty})$, which is $i.i.d.$ over time, we will have that

$$E\{\mathbf{s}_{\boldsymbol{\varrho}t}[\boldsymbol{\psi}_{10}, \boldsymbol{\psi}_{2\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{3\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|\mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0\} = \mathbf{0}, \qquad \text{(A24)}$$

which confirms that the score evaluated at the pseudo-true values will remain a martingale difference sequence.

Therefore, in order to compute the variance of the average score we can follow exactly the same steps as in the proof of Proposition 1 by exploiting the fact that (A23) and (A24) hold at the pseudo-true parameter values $\boldsymbol{\phi}_\infty$. The martingale difference nature of the score also allows us to obtain the expected value of the Hessian along the lines of Proposition 2.

As we mentioned in the proof of Proposition (9), we can tediously show that the conditions for block-diagonality of the expected value of the Hessian and the covariance matrix of the score are that $E[\mathbf{Z}_{\psi_1 lt}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ and $E[\mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ are both 0. But given that

$$\begin{aligned}\mathbf{Z}_{\psi_1 lt}(\boldsymbol{\psi}_{10},\boldsymbol{\psi}_2,\boldsymbol{\psi}_3) &= \left[\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_{10})/\partial\boldsymbol{\psi}_1\cdot\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{10})\right]\boldsymbol{\Psi}_2^{-1/2\prime}\\ &\quad+\left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_{10})]/\partial\boldsymbol{\psi}_1\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{10})]\right\}(\boldsymbol{\psi}_3\otimes\boldsymbol{\Psi}_2^{-1/2\prime}),\\ \mathbf{Z}_{\psi_1 st}(\boldsymbol{\psi}_{10},\boldsymbol{\psi}_2,\boldsymbol{\psi}_3) &= \left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_{10})]/\partial\boldsymbol{\psi}_1\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{10})]\right\}(\boldsymbol{\Psi}_2^{1/2}\otimes\boldsymbol{\Psi}_2^{-1/2\prime}),\end{aligned}$$

those condition will be satisfied if (35) holds in view of the full rank of $\boldsymbol{\Psi}_2$. .□

Proposition 12

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 $E[s_{\eta t}(\boldsymbol{\theta}_0,0)|\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0]\leq 0$.

FSC show that in the multivariate Student t case $s_{\eta t}(\boldsymbol{\theta}_0,0)$ it is proportional to the second generalised Laguerre polynomial (39). Given that $\varsigma_t(\boldsymbol{\theta}_0)=\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$, we can write

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

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

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

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 $\mathbf{m}_t(\boldsymbol{\theta}_0, 0, \lambda_{\eta\infty}) = s_{\phi t}(\boldsymbol{\theta}_0, 0) + \mathbf{e}_{p+1} \cdot \lambda_{\eta\infty}$, 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 $\mathcal{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[\bar{\mathbf{h}}_T(\boldsymbol{\phi}_0)|\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0]$.

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

$$\begin{bmatrix} \mathcal{H}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{H}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\eta\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{bmatrix} \sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \hat{\eta}_T \end{pmatrix} + \mathbf{e}_{p+1}\sqrt{T}(\hat{\lambda}_{\eta T} - \lambda_{\eta\infty}) - \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0, 0, \lambda_{\eta\infty}) = o_p(1),$$

$$\begin{bmatrix} \mathcal{H}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{H}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\eta\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{bmatrix} \sqrt{T} \begin{pmatrix} \tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ 0 \end{pmatrix} + \mathbf{e}_{p+1}\sqrt{T}(\tilde{\lambda}_{\eta T} - \lambda_{\eta\infty}) - \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0, 0, \lambda_{\eta\infty}) = o_p(1),$$

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,

$$\begin{bmatrix} \mathcal{H}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{H}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{H}_{\eta\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{bmatrix} \sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T \\ \hat{\eta}_T \end{pmatrix} + \mathbf{e}_{p+1}\sqrt{T}(\hat{\lambda}_{\eta T} - \tilde{\lambda}_{\eta T}) = o_p(1).$$

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 (38), which derives directly from the expression for $\mathbf{h}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi})$ in FSC evaluated at $(\boldsymbol{\theta}_0, 0)$, its role is to guarantee that $\mathcal{H}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = \mathbf{0}$. In this sense, it is worth mentioning that condition (38) will be satisfied for instance 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

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

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

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

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

Proposition 13

Let $\boldsymbol{\phi}_\infty$ denote the pseudo-true values of $\boldsymbol{\phi}$ corresponding to the assumed log-likelihood function. If we assume that they belong to the interior of the admissible parameter space, we can be implicitly characterise $\boldsymbol{\phi}_\infty$ by the moment conditions

$$\begin{array}{c} E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_\infty, \boldsymbol{\varrho}_\infty)|\boldsymbol{\varphi}_0] = \mathbf{0}, \\ E[s_{\boldsymbol{\varrho}t}(\boldsymbol{\theta}_\infty, \boldsymbol{\varrho}_\infty)|\boldsymbol{\varphi}_0] = \mathbf{0}. \end{array} \tag{A25}$$

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

$$E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_\infty, \mathbf{0})|\boldsymbol{\varphi}_0] = \mathbf{0}, \tag{A26}$$

which will hold if the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is *i.i.d.* $D(\mathbf{0}, \mathbf{I}, \boldsymbol{\varrho}_0)$ because $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ in that case. If we knew $\boldsymbol{\theta}_\infty$, it would be straightforward to test whether (A26) 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 $\hat{\boldsymbol{\varrho}}_T$ satisfy the sample analogues of (A25). In order to account for the sampling variability that this introduces, we can compute the limiting unconditional least squares regression of $\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_\infty, \mathbf{0})$ on $\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_\infty, \boldsymbol{\varrho}_\infty)$ and $\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\varrho}_0 T}(\boldsymbol{\theta}_\infty, \boldsymbol{\varrho}_\infty)$, and retain the residuals. But since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \mathbf{0})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)$ and $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)$ 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, 3 and Lemma 2 will be given by

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

Hence, we can use standard arguments to show that

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

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

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

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

Proposition 14

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

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = -\partial E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta} \tag{A27}$$

and

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}. \tag{A28}$$

Condition (A27) follows immediately from (A14) and the generalised information matrix equality. As for (A28), we can use the same equality together with some of the arguments in the proof of Proposition 5 to show that

$$-\frac{\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]}{\partial\boldsymbol{\theta}} = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]$$
$$-E\left\{\mathbf{W}_s(\boldsymbol{\phi}_0)\left[\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)\right|\boldsymbol{\phi}_0\right\}$$
$$=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)E\left\{\left[\left\{\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right\}-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\right|\boldsymbol{\phi}_0\right\}\mathbf{Z}_d(\boldsymbol{\theta}_0)$$
$$=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)\text{-}\mathbf{W}_s(\boldsymbol{\phi}_0)E\left[\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\text{-}\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}\text{-}1\right)\right\}\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\right|\boldsymbol{\phi}_0\right]\mathbf{W}'_s(\boldsymbol{\phi}_0)$$
$$=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\}=\mathring{\mathcal{S}}(\boldsymbol{\phi}_0).$$

□

B Computational issues

B.1 Score and Hessian for non-elliptical distributions

Since $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \boldsymbol{\varrho}\right]$, it trivially follows that

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\varrho}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}.$$

To prove (2), we can then use the fact that

$$\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N)$$

and

$$\begin{aligned}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}\\ &= -\{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\},\end{aligned} \tag{B29}$$

where $\mathbf{Z}_{dt}(\boldsymbol{\theta}) = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]$ are defined in (3).

As for the Hessian, given that

$$d\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho}) = -d\{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\}, \tag{B30}$$

expression (B29) implies that

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}.$$

In turn,

$$d\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho}) = -dvec\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]$$

$$= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] d\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \qquad \text{(B31)}$$

implies that

$$\frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

$$= \left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}.$$

Finally, (B30) and (B31) trivially imply that

$$\begin{aligned}
\frac{\partial^2 \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\varrho}'} &= -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}, \\
\frac{\partial^2 \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\varrho}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}.
\end{aligned}$$

B.2 Score and Hessian for elliptically symmetric distributions

Since in this case $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$, it trivially follows that

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial \varsigma} \frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})] \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \qquad \text{(B32)} \\
\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial c(\boldsymbol{\eta})/\partial \boldsymbol{\eta} + \partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial \boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}),
\end{aligned}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta}) vec(\mathbf{I}_N) \\
\partial \varsigma_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta}) vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, \qquad \text{(B33)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}), \\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2} \partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial \boldsymbol{\theta} \cdot [\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})], \\
\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \\
\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right\}, \\
\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] &= -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial \varsigma.
\end{aligned}$$

As for the Hessian function $\mathbf{h}_t(\boldsymbol{\phi}) = \partial \mathbf{s}_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'$, we will have

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{\partial^2 g[\varsigma_t(\boldsymbol{\theta}),\eta]}{(\partial \varsigma)^2} \frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\eta]}{\partial \varsigma} \frac{\partial^2 \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \\
\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \partial \varsigma_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \partial^2 g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial \varsigma \partial \boldsymbol{\eta}', \\
\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 c(\boldsymbol{\eta})/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}' + \partial^2 g[\varsigma_t(\boldsymbol{\theta}),\eta]/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}',
\end{aligned}$$

where

$$\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}', \quad \text{(B34)}$$

$$\begin{aligned}\partial^2 \varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' &= 2\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + 8\mathbf{Z}_{st}(\boldsymbol{\theta})[\mathbf{I}_N\otimes\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})]\mathbf{Z}_{st}'(\boldsymbol{\theta}) + 4\mathbf{Z}_{lt}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\\ &+4\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}) - 2[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_p]\partial vec[\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]\partial\boldsymbol{\theta}'\\ &-\{vec'[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\otimes\mathbf{I}_p\}\partial vec\{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}/\partial\boldsymbol{\theta}'.\end{aligned}$$

B.3 Elliptically symmetric efficient score and semiparametric efficiency bound for model (40)

The vector of conditional mean and variance parameters corresponding to model (40) is given by $\boldsymbol{\theta} = (\boldsymbol{\pi}', \boldsymbol{\rho}', \mathbf{c}', \boldsymbol{\gamma}', \alpha, \beta)'$ after normalising the unconditional variance parameter λ to 1.

The Jacobian matrices of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are:

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = [\mathbf{I}_N - diag(\boldsymbol{\rho})]\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'} + diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\theta}'}$$

and

$$\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'} = (\mathbf{I}_{N^2} + \mathbf{K}_{NN})[\lambda_t(\boldsymbol{\theta})\mathbf{c}\otimes\mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'} + \mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'} + (\mathbf{c}\otimes\mathbf{c})\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

respectively, where $\mathbf{E}_N' = (\mathbf{e}_1\mathbf{e}_1'|\ldots|\mathbf{e}_N\mathbf{e}_N')$, with $(\mathbf{e}_1|\ldots|\mathbf{e}_N) = \mathbf{I}_N$, is the unique $N^2\times N$ "diagonalisation" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}_N' vec(\mathbf{A})$ (see Magnus (1988)).

After some straightforward algebraic manipulations, expressions (B32) and (11) lead to:

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) &= \begin{pmatrix} [\mathbf{I}_N - diag(\boldsymbol{\rho})]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\\ diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\\ \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\lambda_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\lambda_t(\boldsymbol{\theta})\\ \frac{1}{2}vecd\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\\ 0\\ 0\end{pmatrix}\\ &+\frac{1}{2}\frac{\partial\boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\right],\\ \frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \alpha\left[2f_{kt-1}(\boldsymbol{\theta})\frac{\partial f_{kt-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\omega_{t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right] + \beta\frac{\partial\lambda_{t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\\ &+[f_{kt-1}^2(\boldsymbol{\theta}) + \omega_{t-1}(\boldsymbol{\theta}) - 1]\frac{\partial\alpha}{\partial\boldsymbol{\theta}} + [\lambda_{t-1}(\boldsymbol{\theta}) - 1]\frac{\partial\beta}{\partial\boldsymbol{\theta}}.\end{aligned}$$

Finally, if we take as initial conditions $\boldsymbol{\mu}_1(\boldsymbol{\theta}) = \boldsymbol{\pi}$ and $\lambda_1(\boldsymbol{\theta}) = 1$, then $\partial\boldsymbol{\mu}_1(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = \partial\boldsymbol{\pi}/\partial\boldsymbol{\theta}'$ and $\partial\lambda_1(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = \mathbf{0}$.

If $\boldsymbol{\gamma} > \mathbf{0}$, we can use the Woodbury formula to prove that

$$\begin{aligned}
f_{kt}(\boldsymbol{\theta}) &= \omega_t(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \\
\omega_t(\boldsymbol{\theta}) &= [\lambda_t^{-1}(\boldsymbol{\theta}) + \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}]^{-1}, \\
\varsigma_t(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) - f_{kt}^2(\boldsymbol{\theta})/\omega_t(\boldsymbol{\theta}), \\
\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) &= \boldsymbol{\Gamma}^{-1} - \omega_t(\boldsymbol{\theta})\boldsymbol{\Gamma}^{-1}\mathbf{c}\mathbf{c}'\boldsymbol{\Gamma}^{-1},
\end{aligned}$$

$$\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} = \boldsymbol{\Gamma}^{-1}\mathbf{c}\omega_t(\boldsymbol{\theta})/\lambda_t(\boldsymbol{\theta}),$$

$$\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} = \mathbf{c}\boldsymbol{\Gamma}^{-1}\mathbf{c}\omega_t(\boldsymbol{\theta})/\lambda_t(\boldsymbol{\theta})$$

$$\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\lambda_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\lambda_t(\boldsymbol{\theta}) = \boldsymbol{\Gamma}^{-1}[\mathbf{v}_t(\boldsymbol{\theta})f_{kt}(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathbf{c}\omega_t(\boldsymbol{\theta})],$$

$$\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}^{-1}[\mathbf{v}_t(\boldsymbol{\theta})\mathbf{v}_t'(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] + \omega_{kt}(\boldsymbol{\theta})\mathbf{c}\mathbf{c}' - \boldsymbol{\Gamma}]\boldsymbol{\Gamma}^{-1}$$

and

$$\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} = \frac{f_{kt}^2(\boldsymbol{\theta})}{\lambda_t^2(\boldsymbol{\theta})}\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \frac{\omega_t(\boldsymbol{\theta})}{\lambda_t(\boldsymbol{\theta})}\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c},$$

where $\mathbf{v}_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) - \mathbf{c}f_{kt}(\boldsymbol{\theta})$, which greatly simplifies the computations (see Sentana (2000)). Specifically,

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) &= \begin{pmatrix}
[\mathbf{I}_N - diag(\boldsymbol{\rho})]\boldsymbol{\Gamma}^{-1}\mathbf{v}_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] \\
diag(\mathbf{y}_{t-1} - \boldsymbol{\pi})\boldsymbol{\Gamma}^{-1}\mathbf{v}_t(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] \\
\boldsymbol{\Gamma}^{-1}[\mathbf{v}_t(\boldsymbol{\theta})f_{kt}(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathbf{c}\omega_t(\boldsymbol{\theta})] \\
\frac{1}{2}vecd\left\{\boldsymbol{\Gamma}^{-1}[\mathbf{v}_t(\boldsymbol{\theta})\mathbf{v}_t'(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] + \omega_t(\boldsymbol{\theta})\mathbf{c}\mathbf{c}' - \boldsymbol{\Gamma}]\boldsymbol{\Gamma}^{-1}\right\} \\
0 \\
0
\end{pmatrix} \\
&\quad + \frac{1}{2}\frac{\partial\lambda_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left[\frac{f_{kt}^2(\boldsymbol{\theta})\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\lambda_t^2(\boldsymbol{\theta})} - \frac{\omega_t(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}}{\lambda_t(\boldsymbol{\theta})}\right].
\end{aligned}$$

The last two items that we require for the score are

$$\begin{aligned}
\frac{\partial f_{kt}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \mathbf{c}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\frac{\partial\omega_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}\boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\omega_t(\boldsymbol{\theta}) \\
&\quad - \frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'[\boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) \otimes \omega_t(\boldsymbol{\theta})\boldsymbol{\Gamma}^{-1}\mathbf{c}] - \frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{c}'\boldsymbol{\Gamma}^{-1}\omega_t(\boldsymbol{\theta})
\end{aligned}$$

and

$$\frac{\partial\omega_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -2\omega_t^2(\boldsymbol{\theta})\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}\boldsymbol{\Gamma}^{-1}\mathbf{c} + \omega_t(\boldsymbol{\theta})\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'(\boldsymbol{\Gamma}^{-1}\mathbf{c} \otimes \boldsymbol{\Gamma}^{-1}\mathbf{c}) + \frac{\omega_t^2(\boldsymbol{\theta})}{\boldsymbol{\lambda}_t^2(\boldsymbol{\theta})}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}.$$

To compute the elliptically symmetric semiparametric bound we need expressions for

$$\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

$$\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'},$$

and

$$\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]vec'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'}.$$

The first term will be given by

$$\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}=\frac{\partial\boldsymbol{\pi}'}{\partial\boldsymbol{\theta}}[\mathbf{I}_N-diag(\boldsymbol{\rho})]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})[\mathbf{I}_N-diag(\boldsymbol{\rho})]\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\rho}'}{\partial\boldsymbol{\theta}}diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\rho}'}{\partial\boldsymbol{\theta}}diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})[\mathbf{I}_N-diag(\boldsymbol{\rho})]\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'}+\frac{\partial\boldsymbol{\pi}'}{\partial\boldsymbol{\theta}}[\mathbf{I}_N-diag(\boldsymbol{\rho})]\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})diag(\mathbf{y}_{t-1}-\boldsymbol{\pi})\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\theta}'},$$

which effectively has four non-zero blocks only, two of which are equal by symmetry.

The second term is also straightforward. Specifically:

$$\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'}$$
$$=\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}[\lambda_t(\boldsymbol{\theta})\mathbf{c}'\otimes\mathbf{I}_N](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\lambda_t(\boldsymbol{\theta})\mathbf{c}\otimes\mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}+\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\mathbf{c}'\otimes\mathbf{c}')[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{c}\otimes\mathbf{c})\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}[\lambda_t(\boldsymbol{\theta})\mathbf{c}'\otimes\mathbf{I}_N](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\lambda_t(\boldsymbol{\theta})\mathbf{c}\otimes\mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}[\lambda_t(\boldsymbol{\theta})\mathbf{c}'\otimes\mathbf{I}_N](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{c}\otimes\mathbf{c})\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\mathbf{c}'\otimes\mathbf{c}')[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\lambda_t(\boldsymbol{\theta})\mathbf{c}\otimes\mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})](\mathbf{c}\otimes\mathbf{c})\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\mathbf{c}'\otimes\mathbf{c}')[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}$$
$$=2\boldsymbol{\lambda}_t^2(\boldsymbol{\theta})\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}\{[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\cdot\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})+\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\odot\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}+[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}]^2\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$$
$$+2\lambda_t(\boldsymbol{\theta})\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}+2\lambda_t(\boldsymbol{\theta})\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}_N'[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+2\lambda_t(\boldsymbol{\theta})[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}]\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+2\lambda_t(\boldsymbol{\theta})[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}]\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{c}'\Sigma_t^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}$$
$$+\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}\odot\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}]\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\odot\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'},$$

where $\odot$ denotes Hadamard products.

But if we assume that $\boldsymbol{\gamma}>\mathbf{0}$, we can use again the Woodbury formula to considerably simplify the previous expressions. The only slightly complex term left is

$$[\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\mathbf{E}_N$$

But if we exploit the explicit shape of $\mathbf{E}_N$, then we can show that the (i,j)th element of this matrix takes the following form

$$\frac{\omega_t(\boldsymbol{\theta})}{\lambda_t(\boldsymbol{\theta})}\frac{b_j}{\gamma_j}\left[\frac{I(i=j)}{\gamma_i}-\frac{b_i b_j}{\gamma_i\gamma_j}\omega_t(\boldsymbol{\theta})\right],$$

where $I(.)$ is the usual indicator function.

Finally,

$$\begin{aligned}\mathbf{W}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] = \frac{1}{2}\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}[\lambda_t(\boldsymbol{\theta})\mathbf{c}'\otimes\mathbf{I}_N](\mathbf{I}_{N^2}+\mathbf{K}_{NN})vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] \\ &+\frac{1}{2}\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}\mathbf{E}'_N vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]+\frac{1}{2}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\mathbf{c}'\otimes\mathbf{c}')vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] \\ &= \lambda_t(\boldsymbol{\theta})\frac{\partial\mathbf{c}'}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c}+\frac{1}{2}\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}}vecd[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]+\frac{1}{2}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c},\end{aligned}$$

whose computation can also be greatly simplified by using the Woodbury formula.

To estimate $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ non-parametrically, we can exploit expression (A7) to write

$$-\frac{2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}=-\frac{2\partial\ln h[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}+\frac{N-2}{2}\frac{1}{\varsigma_t(\boldsymbol{\theta})}.$$

Then, we can compute $h[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]$ either directly by using a kernel for positive random variables (see Chen (2000)), or indirectly by using a faster standard Gaussian kernel after exploiting the Box-Cox-type transformation $\upsilon=\varsigma^k$ (see Hodgson, Linton and Vorkink (2002)). In the second case, the usual change of variable formula yields

$$p(\upsilon;\boldsymbol{\eta})=\frac{\pi^{N/2}}{k\Gamma(N/2)}\upsilon^{-1+N/2k}\exp[c(\boldsymbol{\eta})+g(\upsilon^{1/k};\boldsymbol{\eta})],$$

whence

$$g(\upsilon^{1/k};\boldsymbol{\eta})=\ln p(\upsilon;\boldsymbol{\eta})+\left(1-\frac{N}{2k}\right)\ln\upsilon-\frac{N}{2}\ln 2\pi+\ln k-\ln\Gamma(N/2)-c(\boldsymbol{\eta})$$

and

$$\frac{\partial g(\upsilon^{1/k};\boldsymbol{\eta})}{\partial\upsilon^{1/k}}=k\frac{\partial\ln f(\upsilon;\boldsymbol{\eta})}{\partial\upsilon}\upsilon^{1-1/k}+\frac{k-N/2}{\upsilon^{1/k}}.$$

We use the second procedure in our Monte Carlo simulations because the distribution of $\varsigma_t(\boldsymbol{\theta})$ becomes more normal-like as N increases, which reduces the advantages of using kernels for positive variables. Still, we use a cubic root transformation to improve the approximation, with a common bandwidth parameter for both the density and its first derivative.

The last thing we need is to estimate $\mathrm{M}_{ll}(\boldsymbol{\eta})$ and $\mathrm{M}_{ss}(\boldsymbol{\eta})$. In our experience, the sample analogue of the OOS expression for $\mathrm{M}_{ll}(\boldsymbol{\eta})$ in Proposition 10 based on the nonparametric estimators of $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ tends to overestimate $\mathrm{M}_{ll}(\boldsymbol{\eta})$ even in fairly large samples because $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$

is imprecisely estimated when ς_t is either very small or very large. For that reason, we have considered an alternative estimator based on the following equivalent expression:

$$\mathrm{M}_{ll}(\boldsymbol{\eta}) = cov\left\{ \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\frac{\varsigma_t}{N} \middle| \boldsymbol{\eta} \right\} + (N-2)E[\varsigma^{-1}(\boldsymbol{\theta})|\boldsymbol{\eta}],$$

where we have exploited (A8), as well as Lemma 1 applied to $m(1) = 1$, which yields

$$E[\delta(\varsigma_t, \boldsymbol{\eta})] = -(N-2)E[\varsigma^{-1}|\boldsymbol{\eta}], \tag{B35}$$

as long as $E[\varsigma^{-1}|\boldsymbol{\eta}]$ is bounded, which in the Gaussian case, for instance, requires $N \geq 3$. Importantly, note that (B35) does not depend at all on the semiparametric estimator. Still, its sample analogue typically underestimates $\mathrm{M}_{ll}(\boldsymbol{\eta})$, for which reason in the end we average the two estimators.

As for $\mathrm{M}_{ss}(\boldsymbol{\eta})$, our experience is that the sample analogue of the OOS expression for $\mathrm{M}_{ss}(\boldsymbol{\eta})$ in Proposition 10 tends to underestimate it. For that reason, we divide it by the square of the sample mean of $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\varsigma_t/N$, which converges in probability to 1 asymptotically in view of (A8).

In order to make sure that $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) - \mathcal{S}(\boldsymbol{\phi}_0)$ is positive semidefinite, we also impose the theoretical restrictions $\mathrm{M}_{ll}(\boldsymbol{\eta}_0) \geq 1$ and

$$V\left[\left\{\frac{\delta(\varsigma_t, \boldsymbol{\eta})\varsigma_t}{N} - 1\right\} - \frac{2}{(N+2)\kappa_0 + 2}\left(\frac{\varsigma_t}{N} - 1\right)\right] = \left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]} \geq 0,$$

after replacing κ_0 by its sample analogue. These restrictions also guarantee that our estimates of $\mathcal{C}(\boldsymbol{\phi}_0) - \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)$ will be positive semidefinite too as long as we evaluate these matrices at the same parameter values using the analytical expressions in Propositions 3 and 5. Finally, we deal with the fact that $rank[\mathcal{C}(\boldsymbol{\phi}_0) - \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)] \leq p - 1$ in view of Proposition 8.1.c by setting to 0 those eigenvalues that are smaller than $10^{-7}/T$ in computing the Moore-Penrose inverse of the difference between those matrices.

B.4 The semiparametric efficient score of model (40)

As we mentioned in footnote 1, the first thing to note regarding a non-elliptical distribution function for the innovations is that the choice of $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ affects the value of the log-likelihood function and its score. For the standard (i.e. lower triangular) Cholesky decomposition of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, we will have that

$$dvec(\boldsymbol{\Sigma}_t) = [(\boldsymbol{\Sigma}_t^{1/2} \otimes \mathbf{I}_N) + (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}]dvec(\boldsymbol{\Sigma}_t^{1/2}).$$

Unfortunately, this transformation is singular, which means that we must find an analogous transformation between the corresponding $dvech's$. In this sense, we can write the previous expression as

$$dvech(\mathbf{\Sigma}_t) = [\mathbf{L}_N(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\mathbf{\Sigma}_t^{1/2}), \tag{B36}$$

where $\mathbf{L}_N$ is the elimination matrix (see Magnus, 1988). We can then use the results in chapter 5 of Magnus (1988) to show that the above mapping will be lower triangular of full rank as long as $\mathbf{\Sigma}_t^{1/2}$ has full rank, which means that we can readily obtain the Jacobian matrix of $vech(\mathbf{\Sigma}_t^{1/2})$ from the Jacobian matrix of $vech(\mathbf{\Sigma}_t)$.

In the case of the symmetric square root matrix, the analogous transformation would be

$$dvech(\mathbf{\Sigma}_t) = [\mathbf{D}_N^+(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{D}_N + \mathbf{D}_N^+(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2})\mathbf{D}_N]dvech(\mathbf{\Sigma}_t^{1/2}),$$

where $\mathbf{D}_N$ is the duplication matrix and $\mathbf{D}_N^+ = (\mathbf{D}_N'\mathbf{D}_N)^{-1}\mathbf{D}_N'$ its Moore-Penrose inverse (see Magnus and Neudecker, 1988).

From a numerical point of view, the calculation of both $\mathbf{L}_N(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N'$ and $\mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$ is straightforward. Specifically, given that $\mathbf{L}_N vec(\mathbf{A}) = vech(\mathbf{A})$ for any square matrix $\mathbf{A}$, the effect of premultiplying by the $\frac{1}{2}N(N+1) \times N^2$ matrix $\mathbf{L}_N$ is to eliminate rows N+1, 2N+1 and 2N+2, 3N+1, 3N+2 and 3N+3, etc. Similarly, given that $\mathbf{L}_N\mathbf{K}_{NN}vec(\mathbf{A}) = vech(\mathbf{A}')$, the effect of postmultiplying by $\mathbf{K}_{NN}\mathbf{L}_N'$ is to delete all columns but those in positions 1, N+1, 2N+1,...,N+2, 2N+2,..., N+3, 2N+3,..., N^2.

Let $\mathbf{F}_t$ denote the transpose of the inverse of $\mathbf{L}_N(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$, which will be upper triangular. The fastest way to compute

$$\frac{\partial vec'[\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{-1/2}(\boldsymbol{\theta})] = \frac{1}{2}\frac{\partial vech'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}\mathbf{F}_t\mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{-1/2})$$

is as follows:

1. From the expression for $\partial vec'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$ we can readily obtain $\partial vech'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$ by simply avoiding the computation of the duplicated columns

2. Then we postmultiply the resulting matrix by $\mathbf{F}_t$

3. Next, we construct the matrix

$$\mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2}) = \mathbf{L}_N \begin{pmatrix} \mathbf{\Sigma}_t^{-1/2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t^{-1/2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_t^{-1/2} \end{pmatrix}$$

   by eliminating the first row from the second block, the first two rows from the third block, ..., and all the rows but the last one from the last block

4. Finally, we premultiply the resulting matrix by $\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot\mathbf{F}_t$.

The last task that we must perform is the computation of $\mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$. The two main problems here are the singular nature of $\mathcal{K}(\boldsymbol{\varrho})$, and its positive semidefiniteness. The first problem is easy to solve because

$$\mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) = \underset{\cdot}{\mathrm{K}}(0)\underset{\cdot}{\mathrm{K}}^{-1}(\boldsymbol{\varrho})\underset{\cdot}{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

where

$$\underset{\cdot}{\mathrm{K}}(0)=\begin{pmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & 2\mathbf{D}_N^{+\prime}\end{pmatrix},\ \underset{\cdot}{\mathrm{K}}(\boldsymbol{\rho})=\begin{pmatrix}\mathbf{I}_N & \boldsymbol{\Phi}'\\ \boldsymbol{\Phi} & \boldsymbol{\Upsilon}\end{pmatrix},\underset{\cdot}{\mathrm{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})=\begin{Bmatrix}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\\ vech\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right]\end{Bmatrix},$$

$$\boldsymbol{\Phi}=E\{vech[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N]\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\}$$

and

$$\boldsymbol{\Upsilon}=E\{vech[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N]\cdot vech'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N]|\boldsymbol{\theta},\boldsymbol{\varrho}\}.$$

As for the second problem, there are two alternative solutions:

1. Re-centre and orthogonalise $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ as $\boldsymbol{\varepsilon}_t^{**}(\boldsymbol{\theta})=\bar{\mathbf{P}}_T^{-1/2}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})-\bar{\mathbf{p}}_T]$, where $\bar{\mathbf{p}}_T$ is the sample mean of $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ and $\bar{\mathbf{P}}_T$ its sample covariance. In this way, the sample covariance matrix of the vector $\{\boldsymbol{\varepsilon}_t^{**\prime}(\boldsymbol{\theta}),vech'[\boldsymbol{\varepsilon}_t^{**}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{**\prime}(\boldsymbol{\theta})]\}$ will have exactly the same structure as $\underset{\cdot}{\mathrm{K}}(\boldsymbol{\varrho})$.

2. Replace $\underset{\cdot}{\mathrm{K}}(\boldsymbol{\varrho})$ by either the sample covariance matrix or the second moment matrix of the vector $\underset{\cdot}{\mathrm{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$.

The advantage of the first procedure is that we can exploit the fact that the sample covariance matrix of $\boldsymbol{\varepsilon}_t^{**}(\boldsymbol{\theta})$ will be the identity matrix in using the partitioned inverse formula for $\underset{\cdot}{\mathrm{K}}(\boldsymbol{\varrho})$. On the other hand, the advantage of the second procedure is that there is no need to standardise again the standardised innovations $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, which in our experience makes it more attractive.

It is also worth mentioning that the most convenient way to compute $\underset{\cdot}{\mathrm{K}}(0)\underset{\cdot}{\mathrm{K}}^{-1}(\boldsymbol{\varrho})\underset{\cdot}{\mathrm{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$ is by first computing $\underset{\cdot}{\mathrm{K}}^{-1}(\boldsymbol{\varrho})\underset{\cdot}{\mathrm{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$, and then exploiting the shape of $\underset{\cdot}{\mathrm{K}}(0)$ as follows: (a) copy the first N elements of $\underset{\cdot}{\mathrm{K}}^{-1}(\boldsymbol{\varrho})\underset{\cdot}{\mathrm{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$; and (b) duplicate the remaining $\frac{1}{2}$N(N+1) elements, but doubling the ones in the following positions: N+1, 2N+1, 3N, 4N-1, 5N-2,...,N+N^2. Intuitively, in doing so we are simply using the fact that $2\mathbf{D}_N^{+\prime}vech(\mathbf{A}_L)=vec(\mathbf{A}_L+\mathbf{A}_L')$ for any lower triangular matrix $\mathbf{A}_L$.

Finally, we use a multivariate spherical Gaussian kernel to compute the density of $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ and its derivatives with a common bandwidth parameter.

References

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

Amengual, D. and Sentana (2010a): "A comparison of mean-variance efficiency tests", *Journal of Econometrics* 154, 16-34.

Amengual, D. and Sentana (2010b): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Amengual, D., Fiorentini, G., and Sentana (2010): "Sequential estimators of shape parameters in multivariate dynamic models", work in progress, CEMFI.

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

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.

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

Chen, S.X. (2000): "Probability density function estimation using gamma kernels", *Annals of the Institute of Statistical Mathematics* 52, 471-480.

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

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

Fan, J., Lei, Q., and Xiu, D. (2010): "Non-Gaussian quasi maximum likelihood estimation of GARCH models", mimeo, Princeton University.

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

Fiorentini, G., and Sentana, E. (2009): "Dynamic specification tests for static factor models", CEMFI Working Paper 0912.

Fiorentini, G., and Sentana, E. (2010): "New testing approaches for mean-variance predictability", mimeo, CEMFI.

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. (2007): "Semiparametric multivariate volatility models", *Econometric Theory* 23, 251-280.

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

Harvey, A., and Chakravarty, T. (2008): "Beta-t-(E)GARCH", Cambridge Working Papers in Economics 08340.

Harvey, A., Ruiz, E. and Sentana, E. (1992): "Unobservable component time series models with ARCH disturbances", *Journal of Econometrics* 52, 129-158.

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., Linton, O. and Vorkink, K.P. (2002): "Testing the Capital Asset Pricing Model efficiently under elliptical symmetry: a semiparametric approach", *Journal of Applied Econometrics* 17, 617-639.

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.

Ling, S. and McAleer, M. (2003): "Asymptotic theory for a vector Arma-Garch model", *Econometric Theory* 19, 280-310.

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. (1988): *Linear structures*, Oxford University Press, New York.

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.

Mencía, J. and Sentana, E. (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105-121.

Mencía, J. and Sentana, E. (2010): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", forthcoming in the *Review of Economics and Statistics*.

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

Sentana, E. (2000): "The likelihood function of conditionally heteroskedastic factor models", *Annales d'Economie et de Statistique* 58, 1-19.

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.

Sun, Y. and Stengos, T. (2006): "Semiparametric efficient adaptive estimation of asymmetric Garch models", *Journal of Econometrics* 127, 373-386.

Table 1

Size properties of Hausman tests in finite samples

Parametric

Nominal size (%)	Student t_8			
	$\bar{\pi}$		$\bar{\gamma}$	
	Wald	LM	Wald	LM
1	1.68	1.77	2.35	1.33
5	6.28	6.67	6.69	5.23
10	11.2	11.7	11.1	10.2

Semiparametric

Nominal size (%)	Student t_8			
	$\bar{\pi}$		$\bar{\gamma}$	
	Wald	LM	Wald	LM
1	2.68	4.75	36.1	23.1
5	8.95	11.4	52.5	36.9
10	15.2	17.5	61.9	45.7

Nominal size (%)	normal-gamma			
	$\bar{\pi}$		$\bar{\gamma}$	
	Wald	LM	Wald	LM
1	1.13	2.53	66.0	48.4
5	5.40	7.03	80.9	66.1
10	10.5	12.2	87.0	74.5

Table 2

Size-adjusted power properties of Hausman tests in finite samples

Parametric

Actual size (%)	normal-gamma $\bar{\pi}$ Wald	$\bar{\pi}$ LM	$\bar{\gamma}$ Wald	$\bar{\gamma}$ LM
1	3.40	3.04	99.9	99.9
5	11.1	10.1	100.	100.
10	18.5	16.8	100.	100.

Actual size (%)	asymmetric t $\bar{\pi}$ Wald	$\bar{\pi}$ LM	$\bar{\gamma}$ Wald	$\bar{\gamma}$ LM
1	100.	100.	52.5	55.0
5	100.	100.	78.7	76.5
10	100.	100.	87.9	84.6

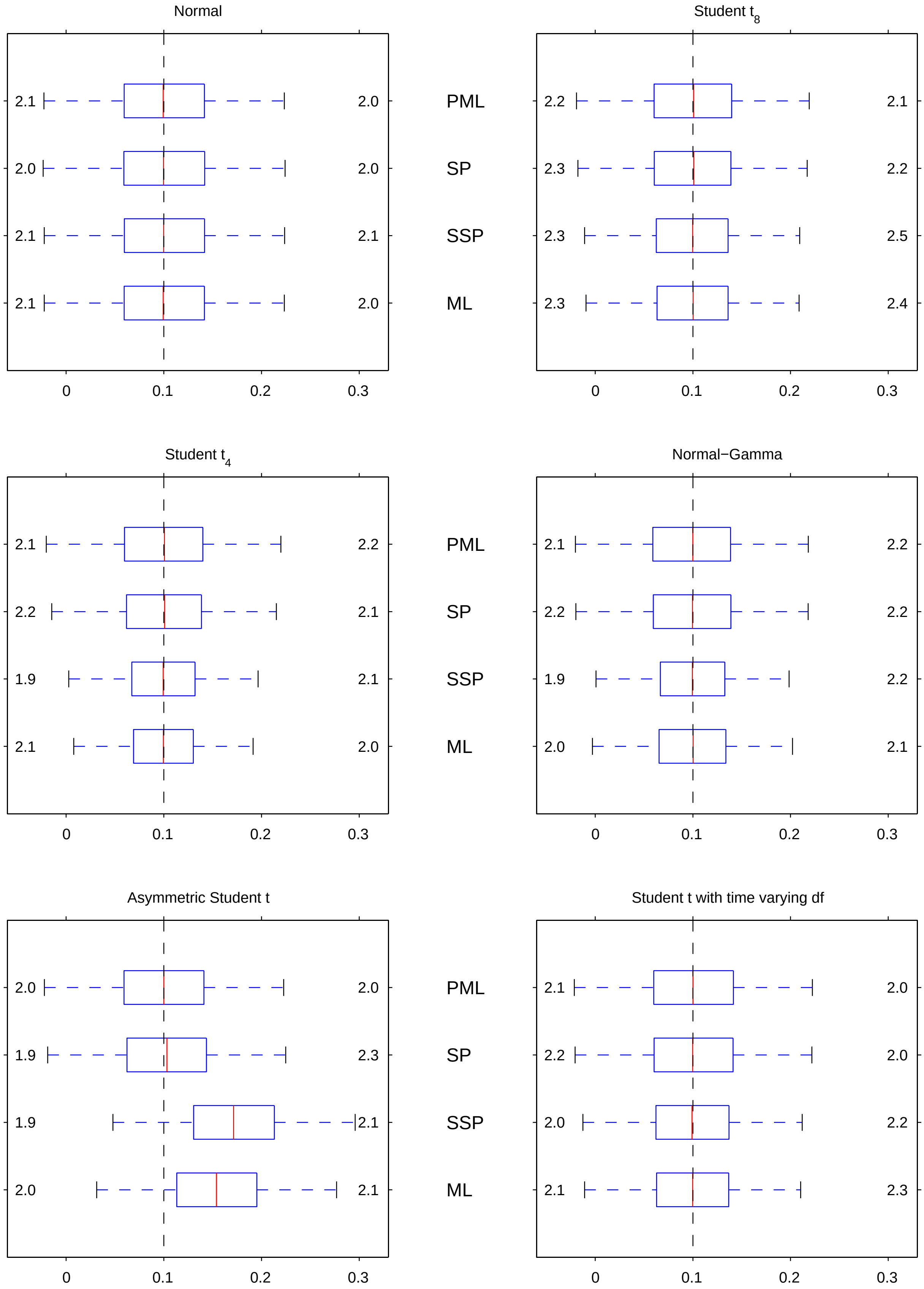
Actual size (%)	t with time-varying df $\bar{\pi}$ Wald	$\bar{\pi}$ LM	$\bar{\gamma}$ Wald	$\bar{\gamma}$ LM
1	1.03	1.09	0.59	0.65
5	4.90	5.08	4.10	4.25
10	10.3	10.3	9.55	9.83

Semiparametric

Actual size (%)	asymmetric t $\bar{\pi}$ Wald	$\bar{\pi}$ LM	$\bar{\gamma}$ Wald	$\bar{\gamma}$ LM
1	100.	50.8	99.9	0.37
5	100.	100.	100.	99.8
10	100.	100.	100.	99.9

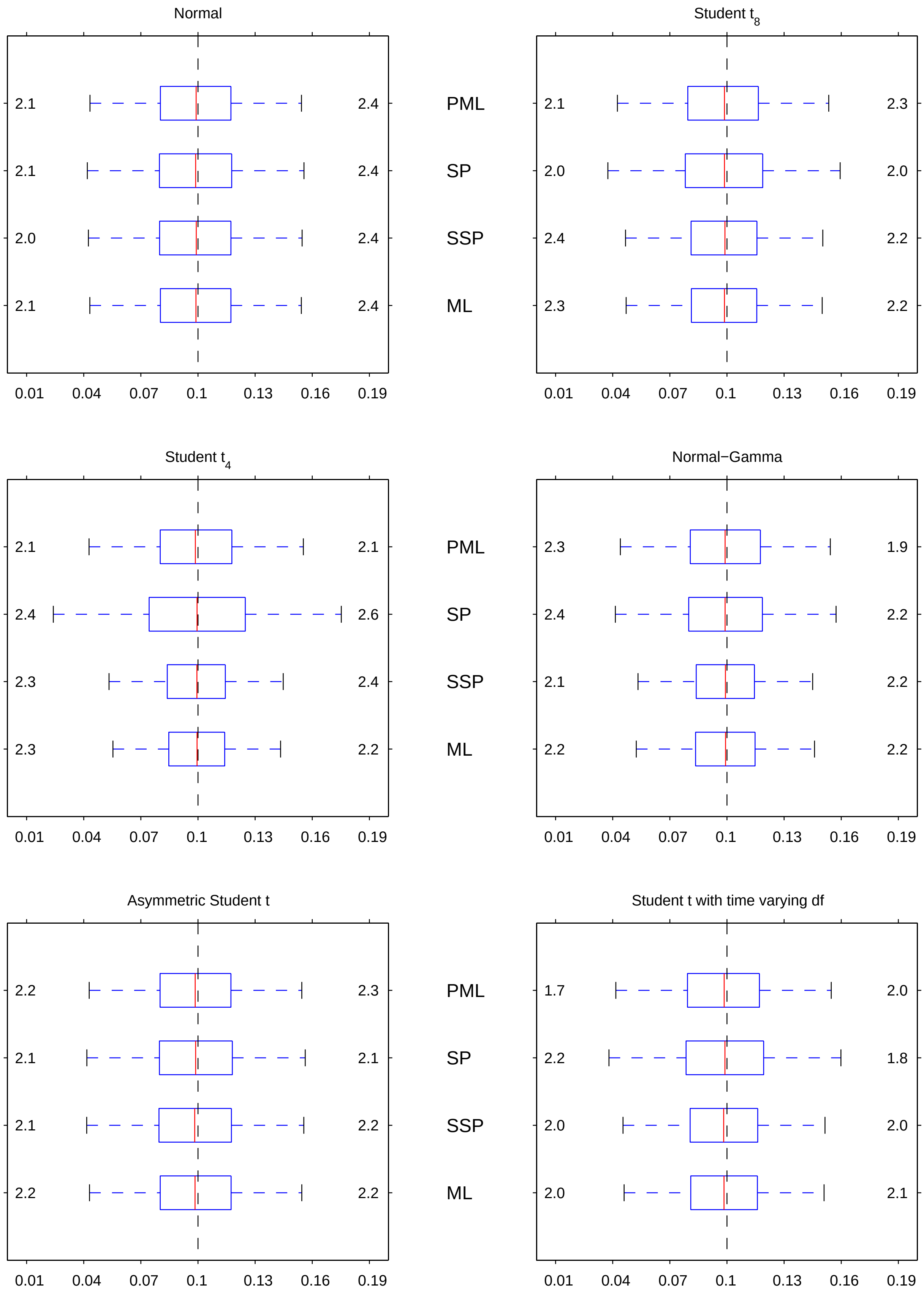
Actual size (%)	t with time-varying df $\bar{\pi}$ Wald	$\bar{\pi}$ LM	$\bar{\gamma}$ Wald	$\bar{\gamma}$ LM
1	0.94	0.85	0.98	0.63
5	5.06	5.10	5.07	4.56
10	10.2	9.71	9.37	9.19

Figure 1A: Monte Carlo distributions of estimators of unconditional mean



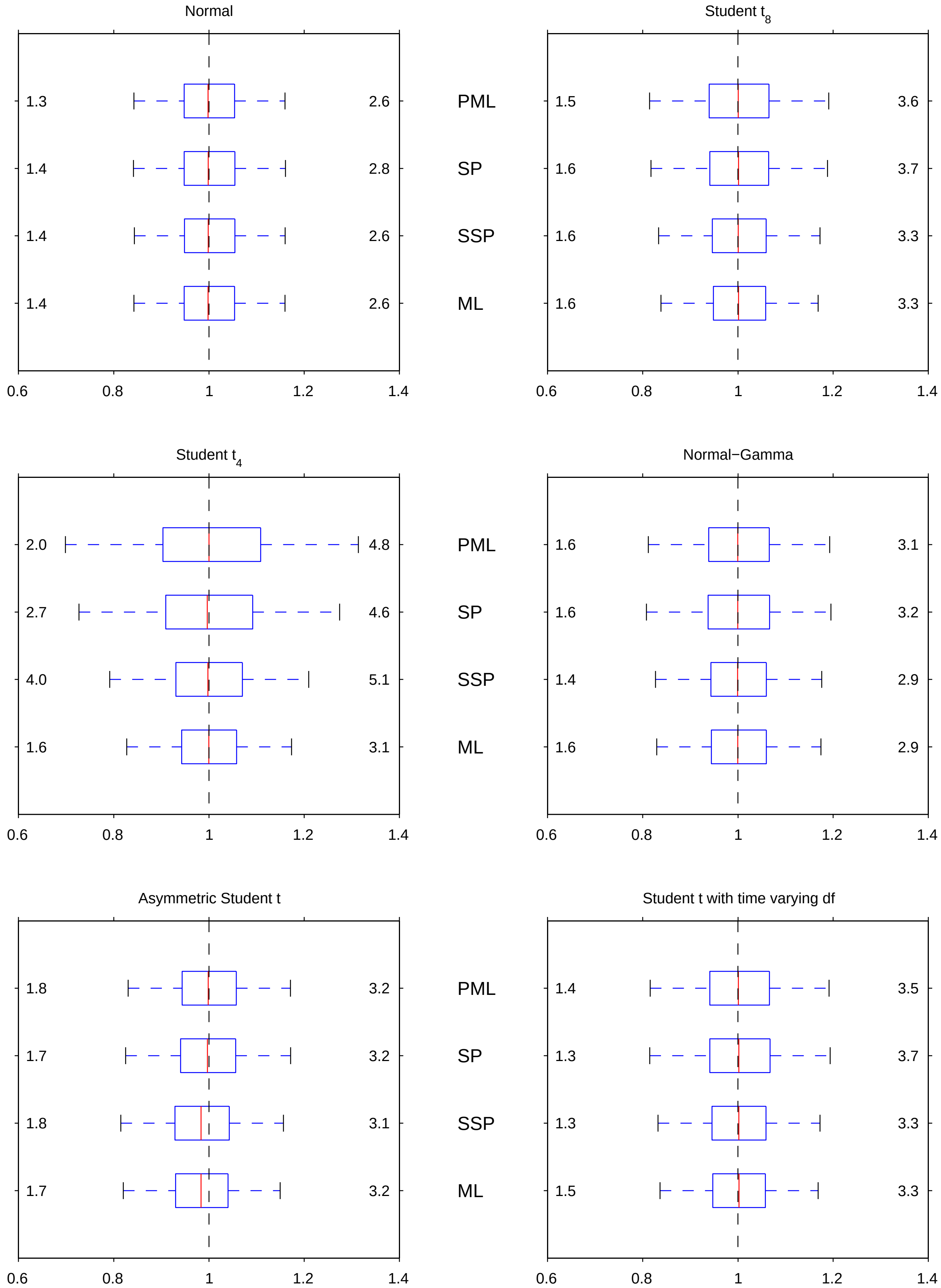
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1B: Monte Carlo distributions of estimators of autoregressive coefficient



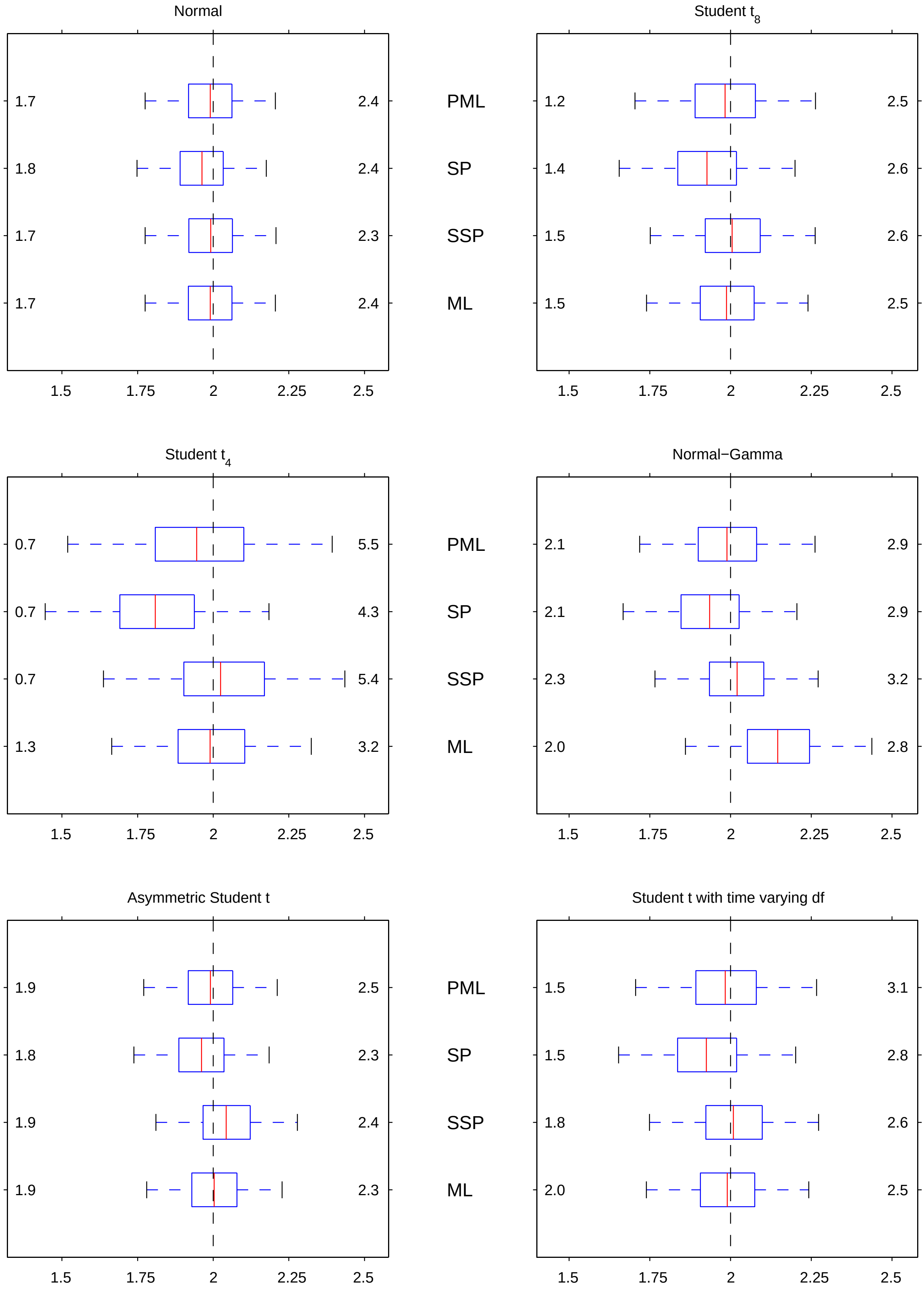
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1C: Monte Carlo distributions of estimators of normalised factor loadings



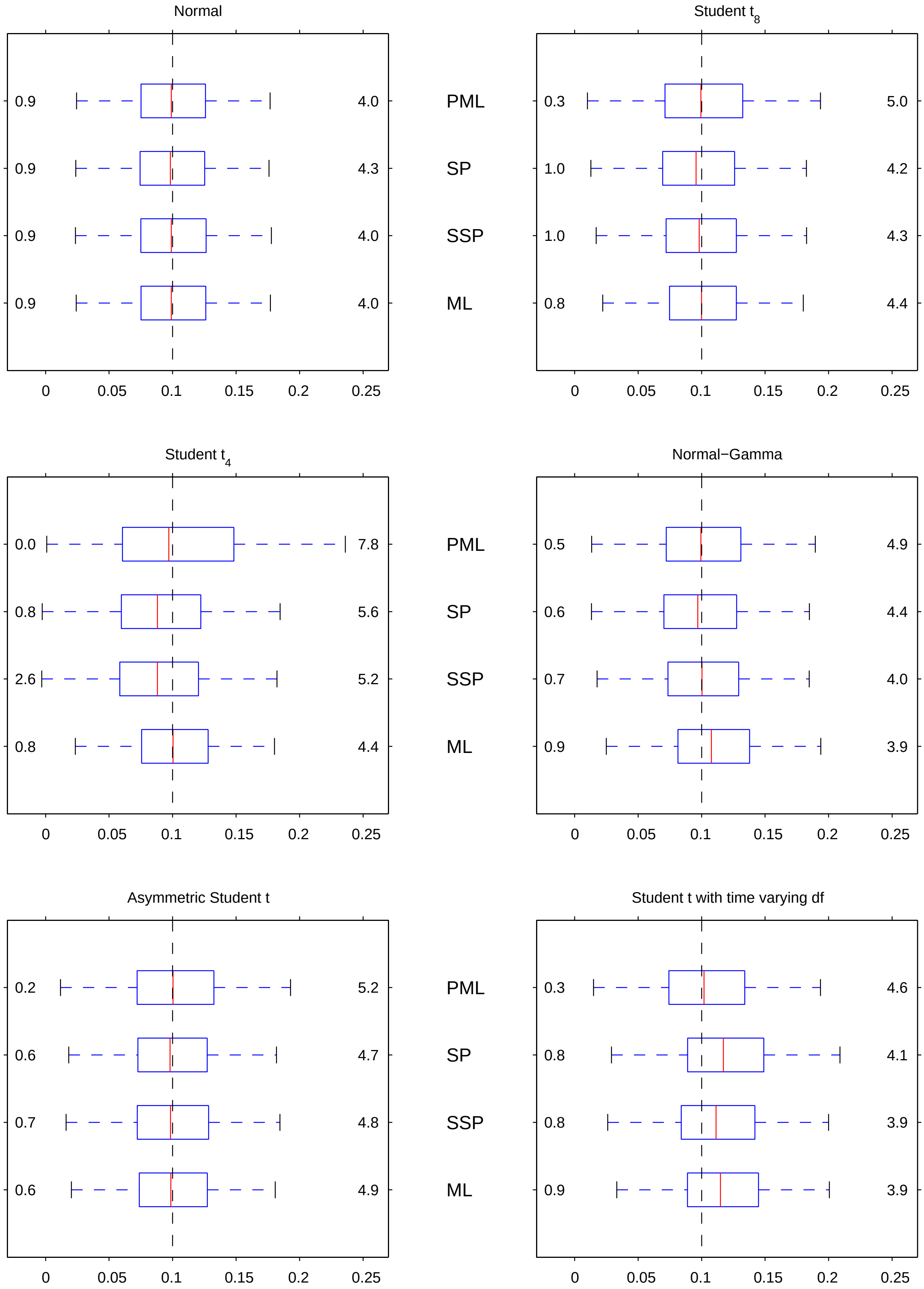
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1D: Monte Carlo distributions of estimators of idyosincratic variances



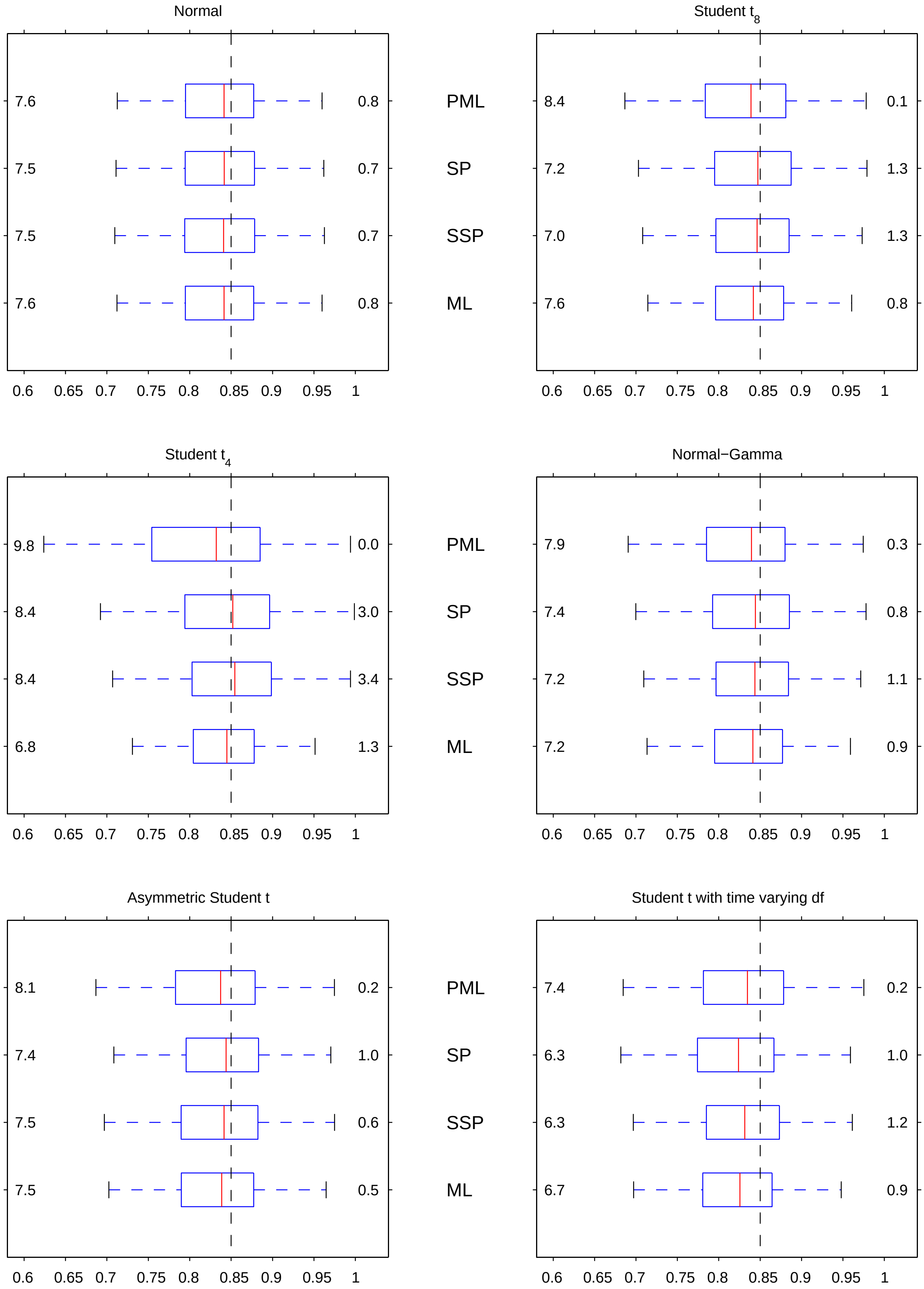
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1E: Monte Carlo distributions of estimators of ARCH coefficent



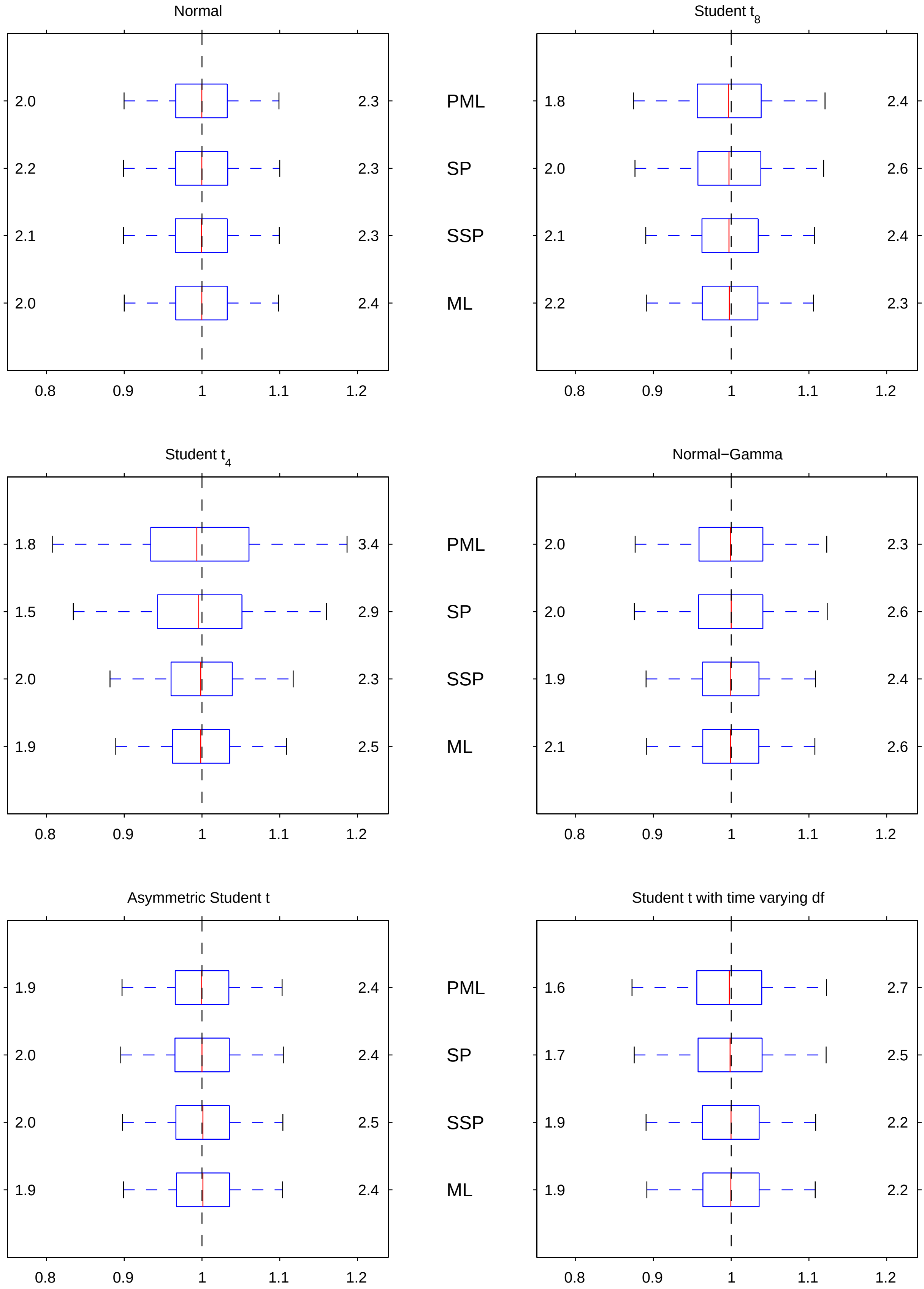
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1F: Monte Carlo distributions of estimators of GARCH coefficent



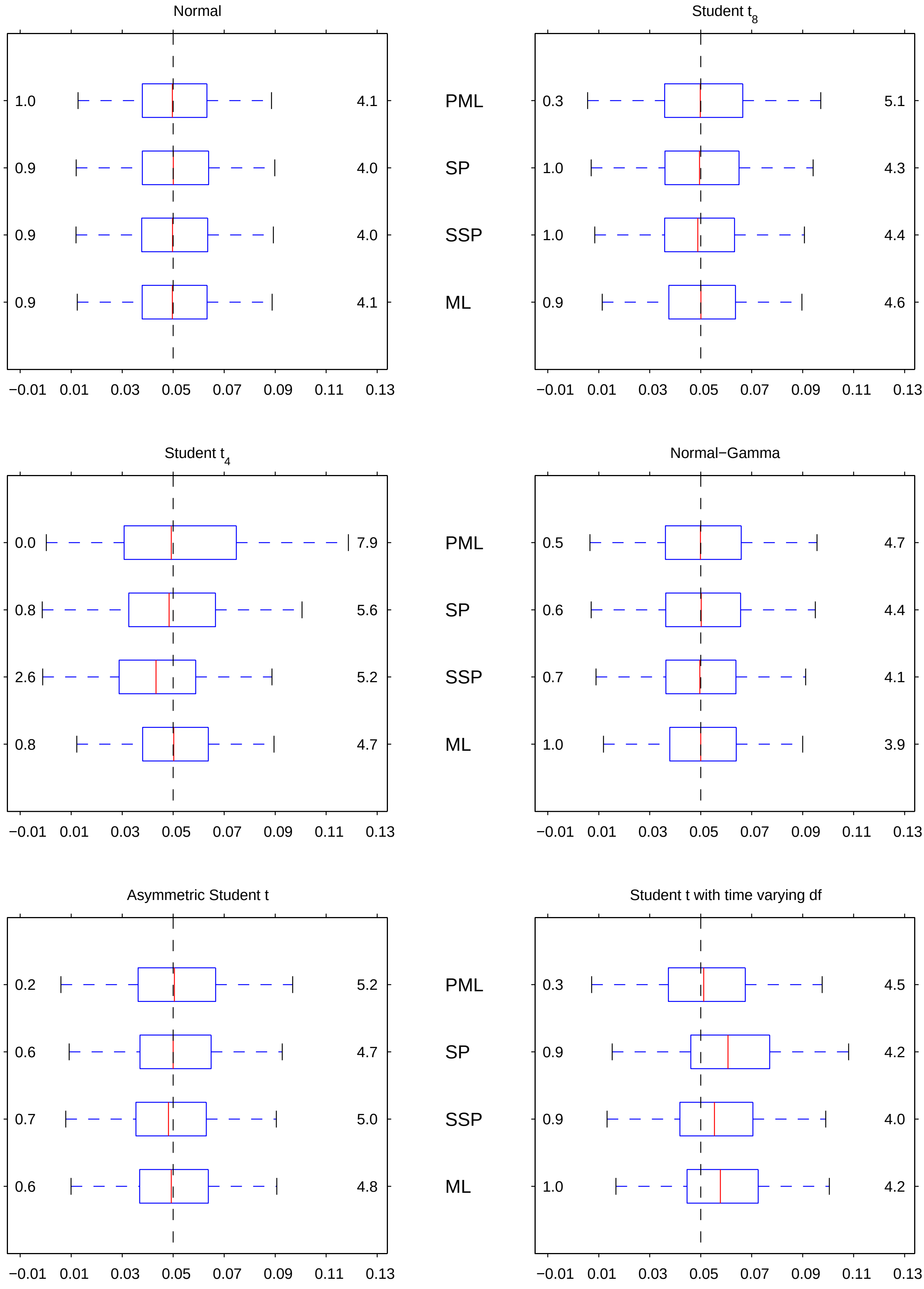
The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1G: Monte Carlo distributions of estimators of re−scaled idyosincratic variances



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.

Figure 1H: Monte Carlo distributions of estimators of re−scaled ARCH coefficent



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. PML means Gaussian−based maximum likelihood estimator, ML Student t−based maximum likelihood estimator, SSP elliptically symmetric semiparametric estimator and SP unrestricted semiparametric estimator.