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

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

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

February 2003
Revised: 29 April 2005

Preliminary and incomplete

Abstract

We compare the efficiency of ML and Gaussian PML estimators of the parameters characterising the mean and variance of conditionally heteroskedastic dynamic regression models with multivariate t innovations. We also propose two alternative sequential estimators of the degrees of freedom parameter based on those PML estimators, and assess their efficiency relative to their ML counterpart. In addition, we derive 1-sided and 2-sided LM tests of conditional homoskedasticity against ARCH(p) and GARCH(1,1) alternatives, and quantify the power gains relative to the tests based on Gaussian PML estimators. Finally, we present finite sample comparisons based on Monte Carlo simulations.

Keywords: Inequality Constraints, ARCH, Financial Returns, Sequential Estimators.

JEL: C13, C12, C51, C52

1 Introduction

Many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering effects (see e.g. Bollerslev, Chou and Kroner (1992) for a survey). This has been long realised, and two main alternative inference approaches have been proposed. The first one uses a "robust" estimation method, such as the Gaussian pseudo-maximum likelihood (PML) procedure advocated by Bollerslev and Wooldridge (1992), which remains consistent for the parameters of the conditional mean and variance functions even if the assumption of conditional normality is violated, or the semiparametric maximum likelihood (SPML) procedures proposed by Engle and Gonzalez-Rivera (1991) and Drost and Klaassen (1997) among others. The second approach, in contrast, specifies a parametric leptokurtic distribution for the standardised innovations, such as the student t employed by Bollerslev (1987), and estimates the conditional mean and variance parameters jointly with the parameters characterising the shape of the assumed distribution by maximum likelihood (ML). While the second approach will often yield more efficient estimators than the first one if the assumed conditional distribution is correct, it has the disadvantage that it may end up sacrificing consistency when it is not (see Newey and Steigerwald (1997)). To the best of our knowledge, though, the trade-offs between efficiency and inconsistency that an empirical researcher faces in practice remain largely unmeasured. One of the objectives of this paper is to quantify the efficiency losses from using PML estimators instead of ML ones in the context of multivariate conditionally heteroskedastic dynamic regression models with student t innovations. We look in detail at two particular examples: one in which the true conditional distribution is in fact Gaussian, and another one in which the model of interest is a univariate nonlinear regression model with ARCH disturbances. In this respect, our work is closely related to Gonzalez-Rivera and Drost (1999), who analysed the asymptotic efficiency of Gaussian PML estimators in univariate GARCH models. However, while they employed as benchmarks the SPML estimator mentioned above and an infeasible ML estimator that maximises the correct likelihood function when the shape parameters are known, we use the more realistic ML procedure that simultaneously estimates the two sets of parameters.

In fact, a non-Gaussian distribution may be indispensable when we are interested in features of the distribution of asset returns, such as its quantiles, which go beyond its conditional mean and variance. For instance, empirical researchers and financial market practitioners are often interested in the so-called Value at Risk of an asset, which is the positive threshold value V such that the probability of the asset suffering a reduction in wealth larger than V equals some pre-specified level $\varkappa < 1/2$. Similarly, in the context of multiple financial assets, one may be interested in the probability of the joint occurrence of several extreme events, which is regularly underestimated by the multivariate normal distribution, especially in larger dimensions. For that reason, we propose two alternative sequential estimators of the degrees of freedom parameter, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators, and assess their asymptotic efficiency relative to their ML counterpart. In particular, we consider a sequential ML estimator, and a sequential method of moments (MM) estimator based on the coefficient of multivariate excess kurtosis.

We also assess the relative power of asymptotic tests based on PML estimators vis-a-vis those based on ML ones. Specifically, we develop a Lagrange Multiplier (LM) test for ARCH effects in univariate nonlinear regressions when the model is estimated by maximising a student t log-likelihood function under the null of conditional homoskedasticity, and evaluate the power gains of our proposed test relative to the standard LM test of Engle (1982). In addition, we also derive and compare computationally simple LM tests against GARCH(1,1) alternatives. Importantly, in all cases we take into account the one-sided nature of the alternative hypotheses to derive more powerful Kuhn-Tucker (KT) multiplier tests, which are asymptotically equivalent to the Likelihood Ratio (LR) and Wald (W) tests.

The rest of the paper is organised as follows. In Section 2, we present closed-form expressions for the score vector and the conditional information matrix of the log-likelihood function based on the student t, and introduce the two alternative sequential estimators of the degrees of freedom parameter. Then, we compare the efficiency of the estimators of the parameters characterising the conditional mean and variance functions and the tail thickness in Sections 3 and 4, respectively. The two score tests mentioned above are discussed in Section 5. A Monte Carlo evaluation of the different parameter estimators and testing procedures can be found in Section 6, followed by an illustrative empirical application to 26 U.K. sectorial stock returns in Section 7. Finally, our conclusions can be found in Section 8. Proofs and auxiliary results are gathered in an Appendix.

2 Theoretical background

2.1 The model

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

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

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

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

To complete the model, we need to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. In principle, we could assume that conditional on $\mathbf{z}_t$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as any particular member of the spherical family, which will be characterised by some additional shape parameters $\boldsymbol{\eta}$ (see the appendix). However, for the sake of concreteness we shall follow Bollerslev (1987) in a univariate context, and Harvey, Ruiz and Sentana (1992) in a multivariate one, and assume that $\boldsymbol{\varepsilon}_t^*$ follows a standardised multivariate t with ν_0 degrees of freedom, or *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ for short. As is well known, the multivariate student t approaches the multivariate normal as $\nu_0 \to \infty$, but has generally fatter tails. For that reason, it is often more convenient to use the reciprocal of the degrees of freedom parameter, $\eta_0 = 1/\nu_0$, as a measure of tail thickness, which will always remain in the finite range $0 \leq \eta_0 < 1/2$ under our assumptions.

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

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \eta)'$ denote the $p+1$ parameters of interest, which we assume variation free. The log-likelihood function of a sample of size T (ignoring initial conditions) takes the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\eta) + g\,[\varsigma_t(\boldsymbol{\theta}), \eta]$, where:

$$d_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$$

corresponds to the Jacobian,

$$c(\eta) = \ln\left[\Gamma\left(\frac{N\eta+1}{2\eta}\right) - \ln\left(\frac{1}{2\eta}\right)\right] - \frac{N}{2}\ln\left(\frac{1-2\eta}{\eta}\right) - \frac{N}{2}\ln 2\pi$$

to the constant of integration of the density, and

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

to its kernel , $\Gamma(.)$ is Euler's gamma function, $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$. Not surprisingly, it can be readily verified that $L_T(\boldsymbol{\theta},0)$ collapses to a conditionally Gaussian log-likelihood.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\eta t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and η, respectively. Fiorentini, Sentana and Calzolari (2003) show that for $\eta > 0$

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

and

$$\begin{aligned} s_{\eta t}(\boldsymbol{\phi}) = \frac{\partial c(\eta)}{\partial\eta} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\eta} &= \frac{N}{2\eta\left(1-2\eta\right)} - \frac{1}{2\eta^2}\left[\psi\left(\frac{N\eta+1}{2\eta}\right)-\psi\left(\frac{1}{2\eta}\right)\right] \\ &-\frac{N\eta+1}{2\eta\left(1-2\eta\right)}\frac{\varsigma_t(\boldsymbol{\theta})}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} + \frac{1}{2\eta^2}\ln\left[1+\frac{\eta}{1-2\eta}\varsigma_t(\boldsymbol{\theta})\right] = e_{rt}(\boldsymbol{\phi}),\end{aligned}$$

where

$$\begin{aligned}\mathbf{e}_{lt}(\boldsymbol{\phi}) &= -2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma_t}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\\ \mathbf{e}_{st}(\boldsymbol{\phi}) &= -vec\left\{\mathbf{I}_N + 2\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma_t}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} = vec\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right],\end{aligned}$$

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

$\psi(x) = \partial\ln\Gamma(x)/\partial x$ is the so-called di-gamma function (or Gauss' psi function; see Abramowitz and Stegun (1964)), and the Jacobian matrices $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}'$ depend on the particular specification adopted.[1]

Taking limits as $\eta \to 0$ from above, they also show that

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

while $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992) because

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

[1] Note that while both $\mathbf{Z}_t(\boldsymbol{\theta})$ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ depend on the specific choice of square root matrix $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ does not, a property that inherits from $l_t(\boldsymbol{\phi})$. The same result is not generally true for non-elliptical distributions (see Haffner and Rombouts (2004) or Mencía and Sentana (2004)), in which case we should redefine $\mathbf{Z}_{st}(\boldsymbol{\theta})$ as $\{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]$.

Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}'_{dt}(\boldsymbol{\phi}), e_{rt}(\boldsymbol{\phi})]'$ evaluated at the true parameter values follows a vector martingale difference, and therefore the same is true of the score vector $\mathbf{s}_t(\boldsymbol{\phi})$. His results also imply that, under suitable regularity conditions, which in particular require that $\boldsymbol{\phi}_0$ belongs to the interior of the parameter space, the asymptotic distribution of the feasible ML estimator will be given by the following expression

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \rightarrow N\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right], \tag{1}$$

where

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\phi}_0) &= p\lim_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\mathcal{I}_t(\boldsymbol{\phi}_0),\\
\mathcal{I}_t(\boldsymbol{\phi}) &= V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right] = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\eta)\mathbf{Z}'_t(\boldsymbol{\theta}),\\
\mathbf{h}_t(\boldsymbol{\phi}_0) &= \begin{pmatrix} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) & \mathbf{h}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) \\ \mathbf{h}'_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) & h_{\eta\eta t}(\boldsymbol{\phi}) \end{pmatrix} = \frac{\partial\mathbf{s}_t(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'} = \frac{\partial^2 l_t(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'},\\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix},
\end{aligned}$$

and

$$\mathcal{M}(\eta) = \begin{bmatrix} \mathcal{M}_{dd}(\eta) & \mathcal{M}_{dr}(\eta) \\ \mathcal{M}'_{dr}(\eta) & \mathcal{M}_{rr}(\eta) \end{bmatrix} = V\left\{\left[\begin{array}{c} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ e_{rt}(\boldsymbol{\phi}) \end{array}\right]\middle| \mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right\} = V\left[\mathbf{e}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right].$$

In this context, Proposition 1 in Fiorentini, Sentana and Calzolari (2003) states that

Proposition 1

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= \begin{pmatrix} \mathcal{I}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) \\ \mathcal{I}'_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) & \mathcal{I}_{\eta\eta t}(\boldsymbol{\phi}) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & 1 \end{pmatrix}\\
&\times \begin{pmatrix} \mathcal{M}_{ll}(\eta) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\eta) & \mathcal{M}_{sr}(\eta) \\ \mathbf{0} & \mathcal{M}'_{sr}(\eta) & \mathcal{M}_{rrt}(\eta) \end{pmatrix}\begin{pmatrix} \mathbf{Z}'_{lt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix},
\end{aligned}$$

where

$$\begin{gathered}
\mathcal{M}_{ll}(\eta) = \frac{\nu\,(N+\nu)}{(\nu-2)\,(N+\nu+2)}\mathbf{I}_N,\\
\mathcal{M}_{ss}(\eta) = \frac{(N+\nu)}{(N+\nu+2)}\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right) - \frac{1}{2(N+\nu+2)}vec(\mathbf{I}_N)vec'(\mathbf{I}_N),\\
\mathcal{M}_{sr}(\eta) = -\frac{2\,(N+2)\,\nu^2}{(\nu-2)\,(N+\nu)\,(N+\nu+2)}vec(\mathbf{I}_N),\\
\mathcal{M}_{rr}(\eta) = \frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{N+\nu}{2}\right)\right] - \frac{N\nu^4\left[\nu^2 + N(\nu-4) - 8\right]}{2\,(\nu-2)^2\,(N+\nu)\,(N+\nu+2)},
\end{gathered}$$

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

2.3 Pseudo maximum likelihood estimators and sequential procedures

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

Proposition 2 *Under the regularity conditions A.1 in Bollerslev and Wooldridge (1992),*

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \rightarrow N\left[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)\right],$$

where $\mathcal{C}(\boldsymbol{\phi}_0) = \mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{B}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)$,

$$\mathcal{A}(\boldsymbol{\phi}_0) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\phi}_0\right] = p \lim_{T\rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{A}_t(\boldsymbol{\phi}_0),$$

$$\mathcal{A}_t(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}; 0)|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),$$

$$\mathcal{B}(\boldsymbol{\phi}_0) = V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\phi}_0\right] = p \lim_{T\rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{B}_t(\boldsymbol{\phi}_0),$$

$$\mathcal{B}_t(\boldsymbol{\phi}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}; 0)|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta})$$

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

and

$$\kappa = \frac{E\left[\varsigma_t^2(\boldsymbol{\theta})|\boldsymbol{\phi}\right]}{N(N+2)} - 1 \tag{3}$$

is the population coefficient of multivariate excess kurtosis.

Given that $\kappa = 2/(\nu - 4)$ for the student t distribution (see appendix), it trivially follows that in our case $\mathcal{B}_t(\boldsymbol{\phi})$ reduces to

$$\frac{\partial \boldsymbol{\mu}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\nu - 2}{2(\nu-4)} \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} \left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'}$$
$$+ \frac{1}{2(\nu-4)} \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} vec\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'}.$$

Importantly, a necessary condition for $\mathcal{B}(\boldsymbol{\phi}_0)$ to be bounded is that $\kappa_0 < \infty$, which in the student t case is equivalent to $\nu_0 > 4$. Otherwise, the asymptotic distribution of the PML estimator $\tilde{\boldsymbol{\theta}}_T$ will be non-standard, unlike that of $\hat{\boldsymbol{\theta}}_T$ (see Hall and Yao (2003)).

Nevertheless, in many circumstances we may be interested in features of the distribution of asset returns, such as its quantiles, which go beyond its conditional mean and variance. For instance, empirical researchers and financial market practitioners are often interested in the so-called Value at Risk of an asset, which is the positive threshold value V such that the probability of the asset suffering a reduction in wealth larger than V equals some pre-specified level $\varkappa < 1/2$. Similarly, in the context of multiple financial assets, one may be interested in the probability of the joint occurrence of several extreme events, which is regularly underestimated by the multivariate normal distribution, especially in larger dimensions.

In this respect, we can use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of η as

$$\tilde{\eta}_T = \arg\max_{\eta} L_T(\tilde{\boldsymbol{\theta}}_T, \eta),$$

which will be characterised by the usual first-order KT conditions

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

where $\bar{s}_{\eta T}(\boldsymbol{\theta}, \eta)$ is the sample mean of $s_{\eta t}(\boldsymbol{\theta}, \eta)$. But since they are analogous to the KT conditions of the ML estimator $\hat{\eta}_T$, then $\tilde{\eta}_T = 0$ if and only if $\hat{\eta}_T = 0$, in which case $\tilde{\boldsymbol{\theta}}_T = \hat{\boldsymbol{\theta}}_T$ too.

This sequential ML estimator of η can be given a rather intuitive interpretation. According to Lemma 1 in Fiorentini, Sentana and Calzolari (2003), if $\boldsymbol{\theta}_0$ were known, then the squared Euclidean norm of the standardised innovations, $\varsigma_t(\boldsymbol{\theta}_0)$, would be independently and identically distributed over time, with a density function, $h\left[\varsigma_t(\boldsymbol{\theta}_0); \eta\right]$, which will be that of either $N(\nu_0 - 2)/\nu_0$ times an F variate with N and ν_0 degrees of freedom when $\nu_0 < \infty$, or a chi-square random variable with N degrees of freedom under Gaussianity. Therefore, we could obtain the (infeasible) ML estimator of η by maximising with respect to η the F-based log-likelihood function of the observed $\varsigma_t(\boldsymbol{\theta}_0)'s$, $\sum_{t=1}^{T} \ln h\left[\varsigma_t(\boldsymbol{\theta}_0); \eta\right]$. Although in practice the standardised residuals are usually unobservable, it turns out that $\tilde{\eta}_T$ is the estimator so obtained when we treat $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$ as if they were really observed.

The asymptotic distribution of the sequential ML estimator of η is stated in the following result:

Proposition 3 *Under the regularity conditions A.1 in Bollerslev and Wooldridge (1992),*

$$\sqrt{T}(\hat{\eta}_T - \eta_0) \to N\left[0, \mathcal{F}(\boldsymbol{\phi}_0)\right],$$

where

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

Importantly, since $\mathcal{C}(\phi_0)$ will become unbounded in the limit as $\nu_0 \to 4$ from above, the asymptotic distribution of $\tilde{\eta}_T$ will also be non-standard when $2 < \nu_0 \leq 4$, unlike that of the feasible ML estimator $\hat{\eta}_T$.

An alternative sequential method of moments (MM) estimator of η, $\breve{\eta}_T$ say, can be obtained from Mardia's (1970) sample coefficient of multivariate excess kurtosis of the estimated standardised residuals

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

by exploiting the theoretical relationship $\kappa = 2/(\nu - 4)$. Specifically, if we define the estimating function

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

and its sample mean as $\bar{m}_{\eta T}(\boldsymbol{\theta}, \eta)$, then the first-order KT conditions characterising $\breve{\eta}_T$ will be

$$\bar{m}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T) \leq 0; \quad \breve{\eta}_T \geq 0; \quad \bar{m}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T) \cdot \breve{\eta}_T = 0 \ ,$$

from where we obtain

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

The asymptotic distribution of the sequential MM estimator $\breve{\eta}_T$ is stated in the following result:

Proposition 4 *If $\nu_0 > 8$, then under the regularity conditions A.1 in Bollerslev and Wooldridge (1992)*

$$\sqrt{T}(\breve{\eta}_T - \eta_0) \to N\left[0, \mathcal{G}(\phi_0)\right]$$

where

$$\mathcal{G}(\phi_0) = \frac{\mathcal{E}(\phi_0) + \mathcal{F}'(\phi_0)\mathcal{C}(\phi_0)\mathcal{F}(\phi_0) + 2\mathcal{F}'(\phi_0)\mathcal{A}^{-1}(\phi_0)\mathcal{D}(\phi_0)}{\mathcal{N}^2(\phi_0)},$$

$$\begin{aligned}
\mathcal{D}(\phi_0) &= \mathbf{Z}_s(\boldsymbol{\theta}_0) \times \frac{4(\nu_0-2)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-6)} vec(\mathbf{I}_N), \\
\mathcal{E}(\phi_0) &= \frac{(\nu_0-2)^2}{(\nu_0-4)^4}\left[\frac{(N+6)(N+4)}{N(N+2)}\frac{(\nu_0-2)(\nu_0-4)}{(\nu_0-6)(\nu_0-8)} - 1\right], \\
\mathcal{F}(\phi_0) &= \mathbf{Z}_s(\boldsymbol{\theta}_0) \times \frac{4(\nu_0-2)}{N(\nu_0-4)} vec(\mathbf{I}_N), \\
\mathcal{N}(\phi_0) &= \frac{\partial m_{\eta t}(\boldsymbol{\theta}_0, \eta_0)}{\partial \eta} = \frac{2\nu_0^2}{(\nu_0-4)^2},
\end{aligned}$$

and

$$\mathbf{Z}_s(\boldsymbol{\theta}_0) = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{st}(\boldsymbol{\theta}_0).$$

In this respect, note that since $\mathcal{G}(\boldsymbol{\phi}_0)$ will diverge to infinity as ν_0 converges to 8 from above, the asymptotic distribution of the sequential MM estimator of η, $\breve{\eta}_T$, will be non-standard for $4 < \nu_0 \leq 8$. Moreover, since the coefficient of excess kurtosis is infinity for $2 < \nu_0 \leq 4$, $\breve{\eta}_T$ will not even be root-T consistent in that case.

In this respect, note that since

$$\bar{m}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0) = \frac{4}{N(N+2)}\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T, 0) + \frac{2}{N}\left[\bar{\varsigma}_T(\tilde{\boldsymbol{\theta}}_T) - N\right],$$

where $\bar{\varsigma}_T(\boldsymbol{\theta})$ is the sample mean of $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$, then $\breve{\eta}_T$ and $\tilde{\eta}_T$ will not necessarily be 0 simultaneously in any given sample. However, we can use a generalisation of Proposition 1 in Fiorentini, Sentana and Calzolari (2004) to show that

$$\left[\bar{\varsigma}_T(\tilde{\boldsymbol{\theta}}_T) - N\right] = o_p(T^{-1/2})$$

if and only $\varsigma_t^2(\boldsymbol{\theta}_0) - N$ can be written as an exact, time-invariant, linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$, in which case zero estimates of $\breve{\eta}_T$ and $\tilde{\eta}_T$ will tend to happen together in large samples. As an extreme example, $\bar{\varsigma}_T(\tilde{\boldsymbol{\theta}}_T)$ will be identically N in nonlinear regression models with conditionally homoskedastic disturbances estimated by Gaussian PML, in which the covariance matrix of the innovations, $\boldsymbol{\Sigma}$, is unrestricted and does not affect the conditional mean, and the conditional mean parameters, $\boldsymbol{\delta}$ say, and the elements of $vech(\boldsymbol{\Sigma})$ are variation free. More generally, Fiorentini, Sentana and Calzolari (2004) establish that the aforementioned condition is indeed satisfied by many univariate ARCH models, including the GARCH-M family analysed by Hentschel (1995), as well as the Quadratic GARCH-M model of Sentana (1995), but not by all.

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

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

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

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

where

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}|\eta t}(\boldsymbol{\theta},\eta) &= \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\eta) - \frac{\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)}{\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\eta) \\ &= \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\begin{array}{c}\mathbf{0} \\ \mathcal{M}_{sr}(\eta_0)/\mathcal{M}_{rr}(\eta_0)\end{array}\right] e_{rt}(\boldsymbol{\phi}),\end{aligned} \tag{4}$$

with $\mathbf{Z}_d(\boldsymbol{\theta}) = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{dt}(\boldsymbol{\phi})$, 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 η, $s_{\eta t}(\boldsymbol{\phi}_0)$. This residual score $\mathbf{s}_{\boldsymbol{\theta}|\eta t}(\boldsymbol{\theta},\eta)$ is sometimes called the parametric efficient score, and its variance,

$$\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \frac{\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)}{\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)}$$

$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \frac{4(N+2)^2\nu^4}{(\nu-2)^2(N+2)^2(N+\nu+2)^2\mathcal{M}_{ss}(\eta_0)}\mathbf{Z}_d(\boldsymbol{\theta})\left[\begin{array}{c}\mathbf{0} \\ vec(\mathbf{I}_N)\end{array}\right]\left[\begin{array}{cc}\mathbf{0}' & vec'(\mathbf{I}_N)\end{array}\right]\mathbf{Z}'_d(\boldsymbol{\theta}),$$

the marginal information matrix of $\boldsymbol{\theta}$, or the feasible parametric efficiency bound. In this respect, note that the inverse of this matrix coincides with the first block of $\mathcal{I}^{-1}(\boldsymbol{\phi}_0)$, $\mathcal{I}^{\theta\theta}(\boldsymbol{\phi}_0)$, which gives us the asymptotic variance of the feasible ML estimator, $\hat{\boldsymbol{\theta}}_T$. Therefore, unless $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ and $s_{\eta t}(\boldsymbol{\phi}_0)$ are uncorrelated, the asymptotic variance of $\hat{\boldsymbol{\theta}}_T$ will generally be larger than $\mathcal{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$, which is the asymptotic variance of the infeasible ML estimator considered by Engle and Gonzalez-Rivera (1991), Gonzalez-Rivera and Drost (1999) and Hafner and Rombouts (2004). As we shall see in Section 3.2, though, it is possible to find situations in which the asymptotic distribution of some elements of $\hat{\boldsymbol{\theta}}_T$ is unaffected by the estimation of η (see also Lange, Little and Taylor (1989)).

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

It is worth noting that the last summand of (4) coincides with $\mathbf{Z}_d(\boldsymbol{\theta}_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on (the linear span of) $e_{rt}(\boldsymbol{\phi}_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)$. Such an interpretation immediately suggests alternative estimators of $\boldsymbol{\theta}$ that replace our parametric assumption on the shape of the distribution of the standardised innovations $\boldsymbol{\varepsilon}^*_t$ by non-parametric or semi-parametric alternatives. In this section, we shall consider two such estimators.

The first one is fully non-parametric, and therefore replaces the linear span of $e_{rt}(\boldsymbol{\phi})$ 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 $\mathbf{e}_{dt}(\boldsymbol{\theta}, 0)$. The following proposition, which generalises Propositions 2, 3 and 4 in Hafner and Rombouts (2004) to models in which the conditional mean is not identically zero, describes the resulting semiparametric efficient score and the corresponding efficiency bound:

Proposition 5 *The semiparametric efficient score is given by the following expression:*

$$\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{K}(0)\,\mathcal{K}^+(\kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta},0)\right], \tag{5}$$

where + denotes Moore-Penrose inverses, while

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0)\left[\mathcal{M}_{dd}(\eta_0) - \mathcal{K}(0)\,\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta}_0)$$

is the semiparametric efficiency bound.

Although these formulae are actually valid for any elliptical distribution, if we exploit the expressions for κ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ that we have derived before for the case of the student t, then it is straightforward to prove that in our case the semiparametric efficient score will be

$$\begin{aligned}
&\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{e}_{lt}(\boldsymbol{\phi}) - \mathbf{Z}_l(\boldsymbol{\theta})\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} - 1\right]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}) \\
&+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{e}_{st}(\boldsymbol{\phi}) - \mathbf{Z}_s(\boldsymbol{\theta})vec\left\{\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} - \frac{1-4\eta}{1-2\eta}\right]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right. \\
&\left.+\frac{2\eta}{1-2\eta}\left[\frac{1-4\eta}{2(N-2)\eta+2}\cdot\frac{\varsigma_t(\boldsymbol{\theta})-N}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} - 1\right]\mathbf{I}_N\right\},
\end{aligned}$$

while

$$\mathcal{M}_{dd}(\eta_0) - \mathcal{K}(0)\,\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0) = \frac{1}{N(\nu\text{-}2)(N+\nu+2)(N+\nu\text{-}2)}$$
$$\times\begin{bmatrix} 2(N+2)(\nu\text{-}2)\mathbf{I}_N & 0 \\ 0 & .5N(N+4)(N+\nu\text{-}2)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right) - N\,(N+6\text{-}\nu)\,vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{bmatrix}.$$

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

For this reason, several authors have suggested to limit the admissible distributions to the class of spherically symmetric ones. As a consequence, the 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 $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)$. The following proposition, which corrects and extends Proposition 7 in Hafner and Rombouts (2004), provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 6 *The elliptically symmetric semiparametric efficient score is given by the following expression:*

$$\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},0)\right], \tag{6}$$

while

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0)\left[\mathring{\mathcal{M}}_{dd}(\eta_0) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa_0)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta}_0)$$

is the elliptically symmetric semiparametric efficiency bound, where $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) = E\left[\mathbf{e}_{dt}(\boldsymbol{\phi})\,|\varsigma_t(\boldsymbol{\theta})\right]$, $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},0) = E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},0)\,|\varsigma_t(\boldsymbol{\theta})\right]$,

$$\begin{aligned}\mathring{\mathcal{K}}(\kappa) &= \frac{(N+2)\kappa+2}{N}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{pmatrix},\\ \mathring{\mathcal{M}}_{dd}(\eta_0) &= \frac{4}{N^2}V\left\{\left.\frac{\partial \ln h\left[\varsigma_t(\boldsymbol{\theta}_0);\eta_0\right]}{\partial\varsigma_t(\boldsymbol{\theta}_0)}\varsigma_t(\boldsymbol{\theta}_0)\right|\boldsymbol{\phi}_0\right\}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{pmatrix},\end{aligned}$$

and $h\left[\varsigma_t(\boldsymbol{\theta}_0);\eta_0\right]$ *is the density function of* $\varsigma_t(\boldsymbol{\theta}_0)$.

Although these formulae are again valid for any elliptical distribution, if we exploit the expressions for κ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ that we have derived before for the case of the student t, then it is easy to see that in our case the ellyptically symmetric semiparametric efficient score will be

$$\begin{aligned}&\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{e}_{lt}(\boldsymbol{\phi})\\ &+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{e}_{st}(\boldsymbol{\phi}) - \mathbf{Z}_s(\boldsymbol{\theta})\left\{\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} - \frac{(1-4\eta)}{1+(N-2)\eta}\right]\frac{\varsigma_t(\boldsymbol{\theta})}{N} - \frac{(N+2)\eta}{1+(N-2)\eta}\right\}vec(\mathbf{I}_N),\end{aligned}$$

while

$$\mathring{\mathcal{M}}_{dd}(\eta_0) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa_0)\,\mathring{\mathcal{K}}(0) = \frac{8(N+2)}{N(N+\nu+2)(N+\nu-2)}\begin{bmatrix}\mathbf{0}\\ vec(\mathbf{I}_N)\end{bmatrix}\begin{bmatrix}\mathbf{0}' & vec'(\mathbf{I}_N)\end{bmatrix}.$$

But once again, in practice $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the joint density of $\boldsymbol{\varepsilon}_t^*$. However, the elliptical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h\left[\varsigma_t(\boldsymbol{\theta}_0);\eta\right]$, avoiding in this way the curse of dimensionality.

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

In the previous sections we have effectively considered five different estimators of $\boldsymbol{\theta}$: (1) the infeasible ML estimator, whose implementation requires knowledge of ν_0; (2) the feasible ML estimator, which simultaneously estimates η; (3) the elliptically symmetric semiparametric estimator, which restricts $\varepsilon_t^*(\boldsymbol{\theta}_0)$ to be spherical, but does not impose any structure on the distribution of $\varsigma_t(\boldsymbol{\theta}_0)$; (4) the unrestricted semiparametric estimator, which does not constrain $\varepsilon_t^*(\boldsymbol{\theta}_0)$ in any sense; and (5) the PML estimator, which imposes $\eta = 0$ even though the true conditional distribution of $\varepsilon_t^*(\boldsymbol{\theta}_0)$ may not be Gaussian. The following proposition ranks the inverses of the asymptotic variances of those five estimators:

Proposition 7 *The infeasible parametric efficiency bound is at least as large as the feasible parametric efficiency bound, which in turn is at least as high as the elliptically symmetric semiparametric efficiency bound, which is not lower than the semiparametric efficiency bound, which in turn weakly dominates the Gaussian pseudo-maximum likelihood efficiency bound* $\mathcal{C}^{-1}(\boldsymbol{\phi}_0)$.

Under correct specification, the ML estimator $\hat{\boldsymbol{\theta}}_T$ is efficient, which implies that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \to N\left[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right],$$

where $\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ is a positive semidefinite matrix. However, the magnitude of the difference between these two matrices is unclear. In the rest of this section, we shall compare the relative asymptotic efficiency of $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$ in two situations of practical interest.

But before, it is important to note that the difference between $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$ immediately suggests a Hausman specification test of the model, which will be given by the quadratic form:

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

Under correct specification of the conditional distribution of $\boldsymbol{\varepsilon}_t$, $H_{\boldsymbol{\theta}T}$ will be asymptotically distributed as a chi-square with degrees of freedom equal to the rank of the matrix $\left[\mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right]$.

3.1 Estimation under conditional normality

Suppose that we decide to maximise the student t likelihood when, in fact, the true conditional distribution is Gaussian. As shown by Fiorentini, Sentana and Calzolari (2003), it turns out that the information matrix is block-diagonal between $\boldsymbol{\theta}$ and η when $\eta_0 = 0$, which means that as far as $\boldsymbol{\theta}$ is concerned, there is no asymptotic efficiency loss in estimating η. More formally:

Proposition 8 *If $\eta_0 = 0$, then*

$$V\left[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right] = \begin{bmatrix} V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right] & \mathbf{0} \\ \mathbf{0}' & N(N+2)/2 \end{bmatrix}$$

where

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

In fact, in large samples the ML estimator of $\boldsymbol{\theta}$ will be numerically identical to the PML estimator approximately half the time because $\eta = 0$ lies at the boundary of the admissible parameter space (see e.g. Andrews (1999) and the references therein). More specifically, when $\eta_0 = 0$, $\sqrt{T}\hat{\eta}_T$ will have an asymptotic normal distribution with mean 0 and variance $2/\left[N(N+2)\right]$ censored from below at 0. Consequently, $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T)$ will be identically 0 with probability approaching 1/2, and $o_p(1)$ the rest of the time. As a result, the Hausman test $H_{\boldsymbol{\theta}T}$ will also be identically 0 with probability approaching 1/2. However, since $\mathcal{C}(\hat{\boldsymbol{\phi}}_T) - \mathcal{I}^{\boldsymbol{\theta\theta}}(\hat{\boldsymbol{\phi}}_T)$ will generally be different from 0 when $\hat{\boldsymbol{\theta}}_T$ is not equal to $\tilde{\boldsymbol{\theta}}_T$, $H_{\boldsymbol{\theta}T}$ might be numerically unstable the rest of the time.

3.2 Estimation of the parameters of a univariate ARCH(q) nonlinear regression model

Consider the following univariate model:

$$
\begin{aligned}
y_t &= \mu_t(\boldsymbol{\delta}_0) + \sigma_t(\boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\varepsilon_t^* \\
\mu_t(\boldsymbol{\delta}) &= \mu\left(\mathbf{z}_t, I_{t-1}; \boldsymbol{\delta}\right) \\
\sigma_t^2(\boldsymbol{\theta}) &= \omega + \sum_{j=1}^{q} \alpha_j \left[y_{t-j} - \mu_{t-j}(\boldsymbol{\delta})\right]^2 \\
\varepsilon_t^* | \mathbf{z}_t, I_{t-1} &\sim i.i.d.\ t(0, 1, \nu_0)
\end{aligned}
$$

Define $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$ and $\boldsymbol{\gamma} = (\omega, \boldsymbol{\alpha}')'$, and partition $\boldsymbol{\theta}$ conformably as $\left(\boldsymbol{\delta}', \omega, \boldsymbol{\alpha}'\right)' = (\boldsymbol{\delta}', \boldsymbol{\gamma}')'$. Given that $\sigma_t^2(\boldsymbol{\theta})$ is symmetric in $\varepsilon_{t-j}(\boldsymbol{\theta})$, and the distribution of ε_t^* conditional on $\mathbf{z}_t$ and I_{t-1} is also symmetric, we can use an argument similar to the one in the proof of Theorem 4 in Engle (1982) to show that

$$
E\left[\left.\frac{1}{\sigma_t^4(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\gamma}'} \right| \boldsymbol{\phi}_0\right] = \mathbf{0}.
$$

But since $\partial \mu_t(\boldsymbol{\delta})/\partial \boldsymbol{\gamma} = 0$ by assumption, this means that $\mathcal{I}_{\boldsymbol{\delta\gamma}}(\boldsymbol{\phi}_0) = \mathcal{A}_{\boldsymbol{\delta\gamma}}(\boldsymbol{\phi}_0) = \mathcal{B}_{\boldsymbol{\delta\gamma}}(\boldsymbol{\phi}_0) = \mathcal{C}_{\boldsymbol{\delta\gamma}}(\boldsymbol{\phi}_0) = \mathbf{0}$, so that both the ML and PML estimators of the conditional mean and variance parameters ($\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, respectively) are asymptotically independent. A similar argument shows that

$$
E\left[\left.\frac{1}{\sigma_t^2(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}} \right| \boldsymbol{\phi}_0\right] = \mathbf{0},
$$

which implies that $\mathcal{I}_{\boldsymbol{\delta}\eta}(\boldsymbol{\phi}_0) = \mathbf{0}$, and thus, that the ML estimators of the conditional mean parameters are also independent from the ML estimator of the tail thickness index η.

But despite the block diagonality of $\mathcal{I}(\boldsymbol{\phi}_0)$ and $\mathcal{C}(\boldsymbol{\phi}_0)$, the inefficiency ratios for $\tilde{\boldsymbol{\theta}}_T$ are rather complicated to obtain, since in general there are no closed form expressions for the unconditional information matrix, and one has to resort to numerical methods (cf. Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999)). Nevertheless, apart from the case in which $\eta_0 = 0$ discussed in the previous subsection, another important exception arises when $\boldsymbol{\alpha}_0 = \mathbf{0}$. In that case, we can use Proposition 1 in Demos and Sentana (1998)[2] to show that if $\nu_0 > 4$, then

$$
\mathcal{C}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) = \begin{bmatrix} \mathcal{C}_{\boldsymbol{\delta\delta}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \omega_0^2 \left[q + 2\left(\nu_0 - 1\right) / \left(\nu_0 - 4\right)\right] & -\omega_0 \boldsymbol{\iota}_q' \\ \mathbf{0} & -\omega_0 \boldsymbol{\iota}_q & \mathbf{I}_q \end{bmatrix}
$$

where

$$
\mathcal{C}_{\boldsymbol{\delta\delta}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) = \left\{ p \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\omega_0} \frac{\partial \mu_t'(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} \frac{\partial \mu_t(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'} \right\}^{-1}
$$

[2]There is a missing scalar term in front of the expression for $\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}_0)$ in their paper.

and $\boldsymbol{\iota}_q$ is a vector of q ones.

As for the ML estimator, we can establish the following result:

Proposition 9 *If $\boldsymbol{\alpha}_0 = 0$, then*

$$\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) = \begin{bmatrix} \mathcal{I}^{-1}_{\boldsymbol{\delta\delta}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathcal{I}^{\omega\omega}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) & \mathcal{I}^{\omega\boldsymbol{\alpha}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) \\ \mathbf{0} & \mathcal{I}^{\omega\boldsymbol{\alpha}\prime}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) & \mathcal{I}^{\boldsymbol{\alpha\alpha}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) \end{bmatrix}$$

where

$$\mathcal{I}_{\boldsymbol{\delta\delta}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) = \frac{\nu_0(\nu_0+1)}{(\nu_0-2)(\nu_0+3)} \left\{ p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\omega_0} \frac{\partial \mu_t'(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} \frac{\partial \mu_t(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'} \right\}$$

and

$$\mathcal{I}^{\boldsymbol{\alpha\alpha}}(\boldsymbol{\delta}_0, \omega_0, \mathbf{0}, \eta_0) = \frac{(\nu_0+3)(\nu_0-4)}{\nu_0(\nu_0-1)} \mathbf{I}_q$$

Therefore, under conditional homoskedasticity, the efficiency ratio of the ML estimator of the conditional mean parameters $\boldsymbol{\delta}$ can be characterised by the scalar quantity

$$\frac{(\nu_0-2)(\nu_0+3)}{\nu_0(\nu_0+1)}$$

while the efficiency ratio of the ML estimator of the q ARCH parameters $\boldsymbol{\alpha}$ is given by the scalar quantity

$$\frac{(\nu_0+3)(\nu_0-4)}{\nu_0(\nu_0-1)} \tag{7}$$

Both these ratios are monotonically increasing in ν_0, and approach 1 from below as $\nu_0 \to \infty$. For $\nu_0 = 9$, for instance, they take the value of 14/15 and 5/6 respectively, while for $\nu_0 = 5$, their values are only 4/5 and 2/5.

4 The relative power of testing procedures based on ML and pseudo-ML estimators of $\boldsymbol{\theta}$

4.1 An LM test for ARCH(q) in univariate regression models

Let us consider again the univariate ARCH(q) nonlinear regression model analyzed in Section 3.2. Given that the score with respect to the ARCH parameters $\boldsymbol{\alpha}$ is

$$s_{\alpha_j t}(\boldsymbol{\phi}) = \frac{1}{2} \left[\frac{(\eta+1)\,\varepsilon_t^{*2}(\boldsymbol{\theta})}{1-2\eta+\eta\varepsilon_t^{*2}(\boldsymbol{\theta})} - 1 \right] \varepsilon_{t-j}^{*2}(\boldsymbol{\theta}) \quad (j = 1, \ldots, q),$$

and that $\mathcal{I}^{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}_0)$ is proportional to the identity matrix under the null from Proposition 4, the information matrix version of a two-sided LM test for ARCH will be given by the following expression

$$\frac{(1+3\bar{\eta}_T)(1-4\bar{\eta}_T)}{(1-\bar{\eta}_T)} \sum_{j=1}^{q} \left\{ \frac{\sqrt{T}}{T} \frac{1}{2} \sum_{t=1}^{T} \left[\frac{(\bar{\eta}_T+1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)}{1-2\bar{\eta}_T+\bar{\eta}_T\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)} - 1 \right] \varepsilon_{t-j}^{*2}(\bar{\boldsymbol{\theta}}_T) \right\}^2 \tag{8}$$

where $\bar{\boldsymbol{\phi}}_T = (\bar{\boldsymbol{\delta}}'_T, \bar{\omega}_T, \mathbf{0}', \bar{\eta}_T) = \arg\max_{\delta,\omega,\eta} L_T(\boldsymbol{\delta}, \omega, \mathbf{0}, \eta)$ are the student t-based restricted ML estimators of $\boldsymbol{\phi}$ obtained under the null of conditional homoskedasticity. Under suitable regularity conditions (see e.g. Crowder (1976)), the asymptotic distribution of (8) will be χ^2_q when $\boldsymbol{\alpha}_0 = \mathbf{0}$.

As usual, there are asymptotically equivalent versions of (8) which are simpler to compute. In particular, given that the first order condition defining $\bar{\omega}_T$ will be

$$\sum_{t=1}^{T} \left[\frac{(\bar{\eta}_T + 1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)}{1 - 2\bar{\eta}_T + \bar{\eta}_T \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)} - 1 \right] = 0$$

the LM test for Arch in (23) is proportional to the sum from $j = 1$ to q of the squares of the sample covariances between $(\bar{\eta}_T + 1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T) / \left[1 - 2\bar{\eta}_T + \bar{\eta}_T \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)\right]$ and $\varepsilon_{t-j}^{*2}(\bar{\boldsymbol{\theta}}_T)$. But since these samples covariances are asymptotically independent under the null in view of Proposition 4, an asymptotically equivalent test can be computed as the sum from $j = 1$ to q of the square t ratios in the regression of $(\bar{\eta}_T + 1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T) / \left[1 - 2\bar{\eta}_T + \bar{\eta}_T \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)\right]$ on a constant and $\varepsilon_{t-j}^{*2}(\bar{\boldsymbol{\theta}}_T)$ $(j = 1, \ldots, q)$.

Not surprisingly, when $\bar{\eta}_T = 0$ the test statistic in (8) numerically coincides with the information matrix version of the Gaussian-based two-sided LM test for Arch(q) derived by Engle (1982). However, that version of Engle's test is incorrectly sized if the conditional distribution is not normal (cf. Koenker (1981)). Since the conditional distribution of ε_t^* is homokurtic, the correct "studentised" version of the Gaussian-based LM test for Arch(q) is

$$\left[\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \mathbf{s}'_{\alpha t}(\breve{\boldsymbol{\phi}}_T) \right] \mathcal{A}^{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}_0)\, \mathcal{C}^{-1}_{\alpha\alpha}(\boldsymbol{\phi}_0)\, \mathcal{A}^{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}_0) \frac{\sqrt{T}}{T} \sum_{t=1}^{T} \mathbf{s}_{\alpha t}(\breve{\boldsymbol{\phi}}_T)$$

$$= \frac{(1 - 4\eta_0)^2}{(1 - \eta_0)^2} \sum_{j=1}^{q} \left\{ \frac{\sqrt{T}}{T} \frac{1}{2} \sum_{t=1}^{T} \left[\varepsilon_t^{*2}(\breve{\boldsymbol{\theta}}_T) - 1 \right] \varepsilon_{t-j}^{*2}(\breve{\boldsymbol{\theta}}_T) \right\}^2 \qquad (9)$$

where $\breve{\boldsymbol{\phi}}_T = (\breve{\boldsymbol{\delta}}'_T, \breve{\omega}_T, \mathbf{0}', 0) = \arg\max_{\delta,\omega} L_T(\boldsymbol{\delta}, \omega, \mathbf{0}, 0)$ are the restricted PML estimators of $\boldsymbol{\phi}$ obtained under the null of conditional homoskedasticity (see Demos and Sentana (1998)). In practice, an asymptotically equivalent test to (9) can be computed as the sum from $j = 1$ to q of the square t ratios in the regression of $\varepsilon_t^{*2}(\breve{\boldsymbol{\theta}}_T)$ on a constant and $\varepsilon_{t-j}^{*2}(\breve{\boldsymbol{\theta}}_T)$. Therefore, from a numerical point of view, the main difference between the t-based and Gaussian-based LM tests for Arch(q), apart from the obvious fact that they use different estimators of $\boldsymbol{\delta}$ and ω to evaluate the standardised residuals, is that the dependent variable in the former is $(\bar{\eta}_T + 1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T) / \left[1 - 2\bar{\eta}_T + \bar{\eta}_T \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)\right]$, while the dependent variable in the latter is $\varepsilon_t^{*2}(\breve{\boldsymbol{\theta}}_T)$.

In order to measure the relative power of the LM test obtained under normality vis-a-vis the one obtained using the student t likelihood, we can compare their non-centrality parameters for a sequence of local alternatives. Since the results in Section 3.2 imply that the non-centrality parameter of the Gaussian test for ARCH is given by $T\boldsymbol{\alpha}'\boldsymbol{\alpha}$, while the non-centrality parameter of the student t-based test for ARCH is given by $\{\nu_0(\nu_0-1)/[(\nu_0+3)(\nu_0-4)]\}T\boldsymbol{\alpha}'\boldsymbol{\alpha}$, the ratio of non-centrality parameters will also be given by (7). Intuitively, the power gains accrue from the fact that the variance of $\varepsilon_t^{*2}(\boldsymbol{\theta}_0)$ $(=2(\nu_0-2)/(\nu_0-4))$ is larger than the variance of $(\eta_0+1)\,\varepsilon_t^{*2}(\boldsymbol{\theta}_0)/\left[1-2\eta_0+\eta_0\varepsilon_t^{*2}(\boldsymbol{\theta}_0)\right]$ $(=2\nu_0/(\nu_0+3))$. Since the former increases without bound as $\nu_0\to 4^+$ while the latter remains bounded, those gains could be substantial when the number of degrees of freedom is low.

Finally, note that since the inequality constraints $\alpha_1\geq 0,\ \ldots,\alpha_q\geq 0$ must be satisfied to guarantee nonnegative conditional variances of an ARCH(q) model, an even more powerful test can be obtained if we test $H_0:\alpha_1=0,\ldots,\alpha_q=0$ versus $H_1:\alpha_1\geq 0,\ldots,\alpha_q\geq 0$, with at least one strict inequality. An argument analogous to the one in Demos and Sentana (1998) shows that a version of the KT multiplier test can be simply computed as the sum of the square t-ratios associated with the positive estimated coefficients in the regression of $(\bar{\eta}_T+1)\,\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)/\left[1-2\bar{\eta}_T+\bar{\eta}_T\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)\right]$ on a constant and the first q lags of $\varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)$. The asymptotic distribution of such a test, which coincides with the asymptotic distribution of the LR and W tests, will be given by the following mixture of $q+1$ independent χ^2's

$$\widetilde{LR}\sim\tilde{W}\sim\sum_{i=0}^{q}\frac{\binom{q}{i}}{2^q}\chi_i^2$$

whose critical values for $q=1,\ldots,12$ can be found in Table 1 in that paper.

4.2 Tests for GARCH(1,1)

Let us now consider testing conditional homoskedasticity vs. the GARCH(1,1) specification

$$\sigma_t^2(\boldsymbol{\theta})=\omega+\alpha\left[y_{t-j}-\mu_{t-j}(\boldsymbol{\delta})\right]^2+\beta\sigma_{t-1}^2(\boldsymbol{\theta})$$

under the maintained assumption that mean and variance parameters are variation free. Since σ_t^2 is effectively computed as $\varphi+\alpha\sum_{j=0}^{t-2}\beta^j\varepsilon_{t-j-1}^2+\beta^{t-1}(\sigma_1^2-\varphi)$, where $\varphi=\omega/(1-\beta)$, it is clear that constant conditional variances are obtained if $\alpha=0$ **and** $\beta=0$. But since the inequality constraints $\alpha\geq 0$ and $\beta\geq 0$ must be satisfied to guarantee nonnegative conditional variances under the alternative, we should again consider one-sided tests. In particular, we should test $H_0:\alpha=0,\ \beta=0$ versus $H_1:\alpha\geq 0,\ \beta\geq 0$, with at least one strict inequality.

However, as Bollerslev (1986) noted, one cannot derive the LM test for conditional homoskedasticity versus GARCH(1,1) in the usual way, because the block of the information matrix whose inverse is required is singular under the null. Econometric wisdom suggests that singularity of the information matrix must be somewhat related to parameter unidentifiability under the null. This is indeed the case, at least asymptotically. From the expression for σ_t^2 above, the time-varying conditional variance is simply $\varphi + \beta^{t-1}(\sigma_1^2 - \varphi)$ when $\alpha = 0$. Hence, σ_t^2 converges to φ as $t \to \infty$ for any $\beta \in [0,1)$, although it may take a long time to settle down if β and $\sigma_1^2 - \varphi$ are large. In contrast, if we set $\sigma_1^2 = \varphi$ to start up the recursions, $\sigma_t^2 = \varphi \ \forall t$. In this specific case, we have a testing situation in which the parameter β is only identified under the alternative. Note, though, that since $\sigma_t^2 = \varphi + \alpha \sum_{j=0}^{t-2} \beta^j \varepsilon_{t-j-1}^2$, α has to be positive under the alternative to guarantee nonnegative variances everywhere, we should still test $H_0 : \alpha = 0$ vs. $H_1 : \alpha \geq 0$ even if we knew β.

There are two standard solutions to testing situations with unidentified parameters under the null. The first one involves computing the LM test statistic for many values of β in the range [0,1), which are then combined to construct an overall test statistic. Such a solution was initially suggested by Davies (1977, 1987), who proposed using the $\sup_\beta LR$ test. More recently, Andrews (2001) discusses ways of obtaining critical values for such tests. His procedure is based on regarding the different LR, W and LM statistics as continuous stochastic processes indexed with respect to the parameter β.

The second solution, which is the one we shall follow in this paper, involves choosing an arbitrary value of β, $\bar{\beta}$ say, to carry out a one-sided LM test as $T\tilde{R}^2$ from the regression of either $(\bar{\eta}_T + 1)\, \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)/\left[1 - 2\bar{\eta}_T + \bar{\eta}_T \varepsilon_t^{*2}(\bar{\boldsymbol{\theta}}_T)\right]$ or $\hat{\varepsilon}_t^2$ on a constant and the distributed lag $\sum_{j=0}^{t-2} \bar{\beta}^j \varepsilon_{t-j-1}^2$ (see Demos and Sentana (1998)). Such tests are asymptotically distributed as a $50 : 50$ mixture of χ_0^2 and χ_1^2 irrespective of the value of $\bar{\beta}$. Obviously, the chosen value of $\bar{\beta}$ influences the small sample power of this test, achieving maximum power when $\bar{\beta}$ coincides with its true value, β_0. In this context, an attractive possibility is to choose $\bar{\beta}$ equal to the decay factor recommended by RiskMetrics (1996) to obtain their widely used exponentially weighted average volatility estimates (e.g. $\bar{\beta} = .94$ for daily observations). In this respect, note that since the RiskMetrics volatility measure is proportional to $\sum_{j=0}^{t-2} \bar{\beta}^j \varepsilon_{t-j-1}^2$, in effect our proposed GARCH(1,1) tests differ from the ARCH(q) tests discussed before in that a few lags of the squared residuals are replaced by the RiskMetrics volatility estimate in the auxiliary regressions.

5 The relative efficiency of ML and sequential estimators of η

The asymptotic distribution of the ML estimator of η, $\hat{\eta}_T$, will be given from (1) by the following expression

$$\sqrt{T}(\hat{\eta}_T - \eta_0) \to N\left[0, \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)\right],$$

where

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

Once more, $\mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ can be understood as the inverse of the residual variance in the theoretical regression of the log-likelihood score corresponding to η, $\mathbf{s}_{\eta t}(\boldsymbol{\phi}_0)$, on the log-likelihood score corresponding to $\boldsymbol{\theta}$, $s_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$. As a result, unless $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ and $s_{\eta t}(\boldsymbol{\phi}_0)$ are uncorrelated, the asymptotic variance of $\hat{\eta}_T$ will generally be larger than $\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0)$, which is the asymptotic variance of the F-based ML estimator of η that we could compute if the $\varsigma_t(\boldsymbol{\theta}_0)'s$ were directly observed.

On the basis of well-known results from Durbin (1970), we would expect that $\tilde{\eta}_T$ will be generally inefficient relative to the ML estimator $\hat{\eta}_T$. Similarly, we would also expect the sequential method of moments estimator $\breve{\eta}_T$ to be generally inefficient relative to the ML estimator $\hat{\eta}_T$ (see e.g. Ogaki (1993)).

The following proposition explains the ranking of the asymptotic covariance matrices of the four estimators of η that we have studied:

Proposition 10

$$\left[\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\right]^{-1} \geq \mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0)$$
$$\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0) + \mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0) \geq \left[\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\right]^{-1}$$
$$\frac{\mathcal{E}(\boldsymbol{\phi}_0) + \mathcal{F}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{F}(\boldsymbol{\phi}_0) + 2\mathcal{F}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}(\boldsymbol{\phi}_0)}{\mathcal{N}^2(\boldsymbol{\phi}_0)} \geq \mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0) + \mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)$$

Under correct specification, the ML estimator $\hat{\eta}_T$ is also efficient, which implies that

$$\sqrt{T}(\tilde{\eta}_T - \hat{\eta}_T) \to N\left[0, \mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)\right],$$

where $\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ is a non-negative scalar. As a result, the asymptotic variance of the sequential ML estimator $\tilde{\eta}_T$ will usually be underestimated by $\mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$. However, the magnitude of the difference between these two asymptotic variances is generally unclear. Nevertheless, since $\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0) = 0$ under normality from Proposition 3, it is clear that $\tilde{\eta}_T$ will be as asymptotically efficient as $\hat{\eta}_T$ when $\eta_0 = 0$. Specifically, under conditional normality both estimators will share the same half normal asymptotic distribution, although they would not numerically coincide when they are not both zero.

Once more, the difference between $\tilde{\eta}_T$ and $\hat{\eta}_T$ suggests an alternative Hausman specification test of the model, which will be given by the following expression:

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

where the Moore-Penrose generalised inverse in this scalar case is simply the reciprocal of $\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ if $\mathcal{F}(\boldsymbol{\phi}_0) - \mathcal{I}^{\eta\eta}(\boldsymbol{\phi}_0)$ is positive, and 0 otherwise. Under correct specification of the conditional distribution of $\boldsymbol{\varepsilon}_t$, $H_{\eta T}$ will be asymptotically distributed as a chi-square with one degree of freedom when $\eta_0 > 0$. In contrast, $H_{\eta T}$ will be identically 0 approximately half the time when $\eta_0 = 0$. However, since $\mathcal{F}(\hat{\boldsymbol{\phi}}_T) - \mathcal{I}^{\eta\eta}(\hat{\boldsymbol{\phi}}_T)$ will generally be different from 0 when $\hat{\boldsymbol{\theta}}_T$ is not equal to $\tilde{\boldsymbol{\theta}}_T$, $H_{\eta T}$ might also be numerically unstable the other half.

Finally, it is worth mentioning that the asymptotic distribution of $\breve{\eta}_T$ will also tend to be half normal as the sample size increases when $\eta_0 = 0$, since $\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)$ is root-T consistent for κ, which is 0 in that case. In fact, it is possible to prove that $\breve{\eta}_T$ will be as efficient as both $\hat{\eta}_T$ and $\tilde{\eta}_T$ under conditional Gaussianity.

6 A Monte Carlo comparison

In this section, we assess the finite sample performance of the different estimators and testing procedures discussed above by means of an extensive Monte Carlo exercise, with an experimental design borrowed from Bollerslev and Wooldridge (1992). Specifically, the model that we simulate and estimate is given by the following equations:

$$\begin{aligned} y_t &= \mu_t(\boldsymbol{\delta}_0) + \sigma_t(\boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\varepsilon_t^* \\ \mu_t(\boldsymbol{\delta}) &= \pi + \rho y_{t-1} \\ \sigma_t^2(\boldsymbol{\delta}, \boldsymbol{\gamma}) &= \omega + \alpha\left[y_{t-1} - \mu_{t-1}(\boldsymbol{\delta})\right]^2 + \beta\sigma_{t-1}^2(\boldsymbol{\delta}, \boldsymbol{\gamma}) \\ \varepsilon_t^*|I_{t-1} &\sim i.i.d.\ t(0, 1, \nu_0) \end{aligned}$$

where $\boldsymbol{\delta}' = (\pi, \rho)$, $\boldsymbol{\gamma}' = (\omega, \alpha, \beta)$, $\pi_0 = 1$, $\rho_0 = .5$, $\omega_0 = .05$, $\alpha_0 = .15$ and $\beta_0 = .8$. As for η_0, we consider three different values: 0, .125 and .25, which correspond to the Gaussian limit, and two student t's with 8 and 4 degrees of freedom respectively. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 250$ and $T = 1,000$ observations based on 10,000 Monte Carlo replications.

Following Fiorentini, Sentana and Calzolari (2003), our estimation procedure employs the following mixed approach: initially, we use a scoring algorithm with a fairly large tolerance criterion; then, after "convergence" is achieved, we switch to a Newton-Raphson algorithm to refine the solution. Both stages are implemented by means of the NAG Fortran 77 Mark 19 library E04LBF routine (see Numerical Algorithm Group 2001 for details), with the analytical expressions derived in Section 2 of that paper. Standard errors are computed using analytical derivatives based on the expressions in Bollerslev and Wooldridge (1992) in the Gaussian case, and Proposition 1 in the case of the t.

Figures 1a, 1b and 1c display kernel estimates of the sampling distributions of the ML (solid) and PML (dashed) estimators of the autoregressive coefficient ρ for $\nu_0 = \infty$, 8 and 4, respectively, constructed with the automatic bandwidth choice given in expression (3.28) of

Silverman (1986). As expected from Proposition 3, the distribution of the two estimators is almost identical under normality, even for the smaller sample size, which is not very surprising given that they are numerically identical over half the time. However, they progressively differ as the degrees of freedom decrease. In this respect, it is important to mention that since the asymptotic distributions of $\hat{\boldsymbol{\delta}}_T$ and $\tilde{\boldsymbol{\delta}}_T$ are independent from the asymptotic distributions of the remaining parameters, as discussed in Section 3.2, the distribution of the PML estimator of ρ, $\tilde{\rho}_T$, remains Gaussian even when $\nu_0 = 4$.

Figures 2a-c and 3a-c display the corresponding kernel estimates of the sampling distributions of the ML and PML estimators of the Arch and Garch parameters α and β, respectively. Again, there is no noticeable difference between ML and PML estimators in the Gaussian case, but the differences become apparent as the distribution of the standardised innovations becomes more leptokurtic. In fact, when $\nu_0 = 4$ the shape of the distribution of the PML estimators $\tilde{\alpha}_T$ and $\tilde{\beta}_T$ remains non-normal even for $T = 1,000$, as discussed in Section 3. Nevertheless, since we have imposed during estimation the usual inequality restrictions on α and β that are compatible with positive, non-explosive variances, those distributions remain bounded between 0 and 1.

Given the large number of parameters involved, we summarise the performance of the estimators of the asymptotic covariance matrix of the estimators of the conditional mean and variance parameters $\boldsymbol{\theta}$ by computing the experimental distribution of a very simple W test statistic. In particular, the null hypothesis that we test is that all five parameters are equal to their true values. Importantly, the asymptotic distribution of such a test will be χ^2_5 regardless of whether or not $\eta_0 = 0$. Our results are summarised in Figures 4a-4c using Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. Not surprisingly, the ML standard errors are more reliable than the PML ones, specially when $\eta_0 = 0$. In fact, while the performance of the ML standard errors seems not to be too sensitive to the value of η_0, the performance of the "robust" PML standard errors notoriously deteriorates as η_0 increases, which is not surprising given that $\mathcal{B}_{\gamma\gamma}(\boldsymbol{\phi}_0)$ becomes unbounded as $\nu \to 4$ from above.

Finally, Figures 4a, 4b and 4c display kernel estimates of the sampling distributions of the ML (solid), sequential ML (dashed) and sequential MM (dash-dotted) estimators of η when $\nu_0 = \infty$, 8 and 4, respectively, together with the fraction of parameter values estimated at the lower bound of 0. Given that there is considerable probability mass on or near the origin, we have used the reflection methods discussed by Silverman (1986) to construct those densities in order to guarantee that they integrate to 1. As can be seen, the proportions of zero estimates of η usually exceed the theoretical values of 1/2, 0 and 0 for $\eta_0 = 0$, .125 and .25, respectively, especially for the smaller sample size. Although the three estimators behave similarly under Gaussianity, they are radically different in the other two cases. As explained in Section 4, while $\hat{\eta}_T$ is asymptotically normally distributed in those two cases, $\breve{\eta}_T$ is not when $\nu_0 = 8$, and not even root-T consistent when $\nu_0 = 4$, in which case $\tilde{\eta}_T$ is not asymptotically normal either.

7 Conclusions

In the context of the general multivariate dynamic regression model with time-varying variances and covariances considered by Bollerslev and Wooldridge (1992), our main contributions are:

1. We compare the relative efficiency of ML and PML estimators of the conditional mean and variance parameters in order to assess the trade-off between efficiency and robustness facing a researcher who is only interested in estimating those parameters. In this respect, we show that there are no efficiency gains or losses in simultaneously estimating the degrees of freedom of a student t model when in fact the conditional distribution is Gaussian. At the same time, in the context of a univariate ARCH nonlinear regression model, we show that the efficiency gains could be substantial, but only if the number of degrees of freedom is low.

2. We propose two computationally simple estimators of the reciprocal of the degrees of freedom parameter of the student t, which can be easily obtained from the Euclidean norm of the standardised residuals evaluated at the PML estimators. In particular, we consider a sequential ML estimator, and a sequential MM estimator based on the coefficient of multivariate excess kurtosis. We also assess the efficiency of these estimators relative to their ML counterpart.

3. We derive an LM test for ARCH(q) in univariate nonlinear regression models, and measure its asymptotic power gains against sequences of local alternatives relative to the Gaussian-based test introduced by Engle (1982). We show that those gains could be important when the conditional distribution is rather leptokurtic.

4. We also derive a simple LM test for GARCH(1,1) in the same context, which uses as regressor the Exponentially Weighted Moving Average volatility estimate popularised by RiskMetrics, and widely used by practitioners.

Appendix

Proofs and auxiliary results

Some useful distribution results

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

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

and

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

respectively.

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

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

where

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

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

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

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

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

then we can write

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

where we have used the definition of the beta function in terms of the gamma function given above.

Proposition 2

The proof is based on a straightforward application of Proposition 1 in Bollerslev and Wooldridge (1992) to the spherically symmetric case. Since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)$, and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)$ is a vector martingale difference sequence, then to obtain $\mathcal{B}_t(\boldsymbol{\phi}_0)$ we only need to compute $V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}_0]$. But since

$$\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\theta}_0,0) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0,0) \end{bmatrix} = \begin{pmatrix} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N\right] \end{pmatrix} = \begin{bmatrix} e_t\mathbf{u}_t \\ vec(e_t^2\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N) \end{bmatrix}$$

for any spherical distribution, then it follows from (A1) and (A2) that

$$V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}_0] = \begin{bmatrix} \mathbf{I}_N & 0 \\ 0 & (\kappa_0+1)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)+\kappa_0 vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix} = \mathcal{K}(\kappa_0). \quad \text{(A3)}$$

As for $\mathcal{A}_t(\boldsymbol{\phi}_0)$, we know that its formula, which is valid regardless of the exact nature of the true conditional distribution, coincides with $\mathcal{B}_t(\boldsymbol{\phi}_0)$ when $\kappa_0=0$ by the (conditional) information matrix equality. □

Proposition 3

The first-order conditions that jointly define the PML estimator of $\boldsymbol{\theta}$ and the sequential ML estimator of η (assuming an interior solution) are:

$$\sqrt{T}\begin{pmatrix} \bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T,0) \\ \bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T,\tilde{\eta}_T) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix},$$

where the overbar denotes the corresponding sample mean. If we linearise these conditions around $\boldsymbol{\phi}_0$, then we obtain using standard arguments that

$$\sqrt{T}\begin{pmatrix} \tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0 \\ \tilde{\eta}_T-\eta_0 \end{pmatrix} = -\begin{pmatrix} \bar{\mathbf{h}}_{\boldsymbol{\theta\theta}T}(\boldsymbol{\theta}_0,0) & \mathbf{0} \\ \bar{\mathbf{h}}'_{\boldsymbol{\theta}\eta T}(\boldsymbol{\theta}_0,\eta_0) & \bar{h}_{\eta\eta}(\boldsymbol{\theta}_0,\eta_0) \end{pmatrix}^{-1}\sqrt{T}\begin{pmatrix} \bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0,0) \\ \bar{s}_{\eta T}(\boldsymbol{\theta}_0,\eta_0) \end{pmatrix} + o_p(1),$$

where we have used the fact that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$ does not vary with η when regarded as the estimating function for $\tilde{\boldsymbol{\theta}}_T$. In addition, given that we can write

$$s_{\eta t}(\boldsymbol{\phi}_0)=-\frac{\nu_0^2}{2\eta_0^2}\left\{\ln\left(1-\frac{\zeta_t}{\xi_t+\zeta_t}\right)-\left[\psi\left(\frac{\nu_0+N}{2}\right)-\psi\left(\frac{\nu_0}{2}\right)\right]\right\}$$
$$-\frac{\nu_0^2\,(\nu_0+N)}{2(\nu_0-2)}\left[\frac{\zeta_t}{\xi_t+\zeta_t}-\frac{N}{(\nu_0+N)}\right],$$

(see Fiorentini, Sentana and Calzolari (2003)), we can use the properties of the beta distribution and the martingale difference nature of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)$ and $s_{\eta l}(\boldsymbol{\theta}_0,\eta_0)$ to prove that

$$E\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)s_{\eta l}(\boldsymbol{\theta}_0,\eta_0)\right|\boldsymbol{\phi}_0]=0\quad\forall t,l,$$

which confirms that $\mathbf{e}_{dt}(\boldsymbol{\theta},0)$ is conditionally orthogonal to $e_{rt}(\boldsymbol{\phi})$. Finally, the expression for $\mathcal{F}(\boldsymbol{\phi}_0)$ follows from the definitions of $\mathcal{A}(\boldsymbol{\phi}_0)$, $\mathcal{B}(\boldsymbol{\phi}_0)$, $\mathcal{I}_{\eta\eta}(\boldsymbol{\phi}_0)$, and $\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)$. □

Proposition 4

In this case, the first-order conditions that jointly define the PML estimator of $\boldsymbol{\theta}$ and the sequential MM estimator of η (assuming an interior solution) are:

$$\sqrt{T}\begin{pmatrix}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\tilde{\boldsymbol{\theta}}_T,0)\\ \bar{m}_{\eta T}(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\0\end{pmatrix}.$$

A Taylor expansion analogous to the one used in the proof of Proposition 3 leads to

$$\sqrt{T}\begin{pmatrix}\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \breve{\eta}_T-\eta_0\end{pmatrix}=-\begin{pmatrix}\bar{\mathbf{h}}_{\boldsymbol{\theta\theta}T}(\boldsymbol{\theta}_0,0) & \mathbf{0}\\ \partial\bar{m}_{\eta T}(\boldsymbol{\theta}_0,\eta_0)/\partial\boldsymbol{\theta}' & \partial\bar{m}_{\eta T}(\boldsymbol{\theta}_0,\eta_0)/\partial\eta\end{pmatrix}^{-1}\sqrt{T}\begin{pmatrix}\bar{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)\\ \bar{m}_{\eta t}(\boldsymbol{\theta}_0,\eta_0)\end{pmatrix}+o_p(1),$$

where

$$\begin{aligned}\frac{\partial m_{\eta t}(\boldsymbol{\theta},\eta)}{\partial\eta}&=-\frac{2}{(1-4\eta)^2}\\ \frac{\partial m_{\eta t}(\boldsymbol{\theta},\eta)}{\partial\boldsymbol{\theta}'}&=\frac{2\varsigma_t(\boldsymbol{\theta})}{N(N+2)}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},\\ \frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=-2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})-\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]vec\left[\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\right]\end{aligned}$$

But since

$$\varsigma_t(\boldsymbol{\theta}_0)=(\nu_0-2)\zeta_t/\xi_t,$$

we can write

$$m_{\eta t}(\boldsymbol{\theta}_0,\eta_0)=\frac{(\nu_0-2)^2\,\zeta_t^2}{N(N+2)\xi_t^2}-\frac{\nu_0-2}{\nu_0-4},$$

and

$$\frac{\partial\varsigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}=-2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\sqrt{\frac{(\nu_0-2)\,\zeta_t}{\xi_t}}\mathbf{u}_t$$
$$-\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\frac{(\nu_0-2)\,\zeta_t}{\xi_t}vec\left(\mathbf{u}_t\mathbf{u}_t'\right).$$

Then, we can combine again the martingale difference character of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$ and the serial independent nature of $\varsigma_t(\boldsymbol{\theta}_0)$ with the properties of the beta distribution to prove that

$$\begin{aligned} E\left[\left.\frac{\partial m_{\eta t}(\boldsymbol{\theta},\eta)}{\partial \boldsymbol{\theta}}\right| \mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0\right] &= -\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} vec\left(\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right)\frac{2(\nu_0-2)}{N(\nu_0-4)}, \\ E\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)\cdot m_{\eta t}(\boldsymbol{\theta}_0,\eta_0)\right| \mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0] &= \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} vec\left(\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right)\frac{2(\nu_0-2)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}, \end{aligned}$$

and

$$E\left[m_{\eta t}(\boldsymbol{\theta}_0,\eta_0)m_{\eta l}(\boldsymbol{\theta}_0,\eta_0)\right|\boldsymbol{\phi}_0] = I(t=l)\frac{(\nu_0-2)^2}{(\nu_0-4)^4}\left[\frac{(N+6)(N+4)}{N(N+2)}\frac{(\nu_0-2)(\nu_0-4)}{(\nu_0-6)(\nu_0-8)}-1\right],$$

where $I(.)$ denotes the usual indicator function. Finally, the expression for $\mathcal{G}(\boldsymbol{\phi}_0)$ follows from the definitions of $\mathcal{A}(\boldsymbol{\phi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_0)$. □

Proposition 5

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

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

from where we can immediately see that

$$E\left[\mathbf{e}_{dt}(\boldsymbol{\theta},0)\mathbf{e}'_{dt}(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}\right] = \mathcal{K}(0) \tag{A4}$$

regardless of the conditional distribution of $\boldsymbol{\varepsilon}_t^*$.

Therefore, it trivially follows from (A3) and (A4) that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathcal{K}(0)\mathcal{K}^+(\kappa_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)\right]\mathbf{e}'_{dt}(\boldsymbol{\theta}_0,0)\,|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0\right\} = \mathbf{0}$$

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

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathcal{K}(0)\mathcal{K}^+(\kappa_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)\right]|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0\right\} = \mathbf{0}.$$

Hence, the second summand of (5), which can be interpreted as $\mathbf{Z}_d(\boldsymbol{\theta}_0)$ times the residual from the theoretical regression of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on a constant and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)$, 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 $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)$.

Now, if we write (5) as

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

we can use the law of iterated expectations to show that the semiparametric efficient score (5) evaluated at the true parameter values will be unconditionally orthogonal to the unrestricted tangent set because so is $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)$ and $E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0\right] = \mathbf{0}$.

Finally, the expression for the semiparametric efficiency bound will be

$$E\left[\begin{array}{c}\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathcal{K}(0)\mathcal{K}^+(\kappa_0)\,\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)\right]\} \\ \times\{\mathbf{e}_{dt}(\boldsymbol{\phi}_0)'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathbf{e}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{e}'_{dt}(\boldsymbol{\theta}_0, 0)\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})\}\end{array}\middle|\boldsymbol{\phi}_0\right]$$
$$= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right]$$
$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\left[\mathbf{e}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{e}'_{dt}(\boldsymbol{\theta}_0, 0)\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$-E\left\{\mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathcal{K}(0)\mathcal{K}^+(\kappa_0)\,\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)\right]\mathbf{e}_{dt}(\boldsymbol{\phi}_0)'\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$+E\left\{\mathbf{Z}_d(\boldsymbol{\theta})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathcal{K}(0)\mathcal{K}^+(\kappa_0)\,\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0)\right]\left[\mathbf{e}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{e}'_{dt}(\boldsymbol{\theta}_0, 0)\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0)\left[\mathcal{M}_{dd}(\eta_0) - \mathcal{K}(0)\mathcal{K}^+(\kappa_0)\,\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta}_0)$$

by virtue of (A3), (A4) and the law of iterated expectations. □

Proposition 6

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

$$\begin{aligned}\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0, 0) &= E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta}_0, 0) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0, 0)\end{array}\middle|\varsigma_t(\boldsymbol{\theta}_0)\right] = E\left\{\begin{array}{c}\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0) - \mathbf{I}_N\right]\end{array}\middle|\varsigma_t(\boldsymbol{\theta}_0)\right\} \\ &= E\left[\begin{array}{c}e_t\mathbf{u}_t \\ vec(e_t^2\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N)\end{array}\middle|e_t^2\right] = \left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\left[\begin{array}{c}\mathbf{0} \\ vec(\mathbf{I}_N)\end{array}\right],\end{aligned} \tag{A5}$$

and

$$\begin{aligned}\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) = E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi}_0) \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0)\end{array}\middle|\varsigma_t(\boldsymbol{\theta}_0)\right] &= -E\left\{\begin{array}{c}2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \eta_0\right]/\partial\varsigma_t\cdot\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0) \\ vec\left[\mathbf{I}_N + 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \eta_0\right]/\partial\varsigma_t\cdot\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)\right]\end{array}\middle|\varsigma_t(\boldsymbol{\theta}_0)\right\} \\ &= -E\left\{\begin{array}{c}2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \eta_0\right]/\partial\varsigma_t\cdot e_t\mathbf{u}_t \\ vec(\mathbf{I}_N + 2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \eta_0\right]/\partial\varsigma_t\cdot e_t^2\mathbf{u}_t\mathbf{u}'_t)\end{array}\middle|e_t^2\right\} = \\ &\quad -\left\{\frac{2\partial g\left[\varsigma_t(\boldsymbol{\theta}_0), \eta_0\right]}{\partial\varsigma_t}\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} + 1\right\}\left[\begin{array}{c}\mathbf{0} \\ vec(\mathbf{I}_N)\end{array}\right],\end{aligned} \tag{A6}$$

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

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

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

and

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

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

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

and

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

In this respect, we can use (3) to show that the latter is simply $[(N+2)\kappa+2]/N$, so that

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

As for the former, we can use expression (2.21) in Fang, Kotz and Ng (1990) to write the density function of $\varsigma_t(\boldsymbol{\theta})$ as

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

where $exp\left\{g[\varsigma_t(\boldsymbol{\theta}),\eta]\right\}$ gives us the kernel of the density of $\boldsymbol{\varepsilon}_t^*$ as a function of $\varsigma_t(\boldsymbol{\theta})$ and the tail parameter(s) η, while $\exp\left[c(\eta)\right]$ is the corresponding constant of integration. We can then exploit this relationship to show that

$$\left[\frac{2\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma_t}\frac{\varsigma_t(\boldsymbol{\theta})}{N}+1\right]=\frac{2}{N}\left[1+\frac{\partial\ln h\left[\varsigma_t(\boldsymbol{\theta})\right]}{\partial\varsigma_t}\varsigma_t(\boldsymbol{\theta})\right]. \tag{A7}$$

On this basis, we can use the fact that we have standardised the spherical variable in such a way that

$$E[\varsigma_t(\boldsymbol{\theta})|\boldsymbol{\phi}]=\int_0^\infty\varsigma_t(\boldsymbol{\theta})h\left[\varsigma_t(\boldsymbol{\theta})\right]=N$$

to show that

$$E\left[\varsigma_t(\boldsymbol{\theta})\frac{\partial\ln h\left[\varsigma_t(\boldsymbol{\theta})\right]}{\partial\varsigma_t(\boldsymbol{\theta})}\middle|\boldsymbol{\phi}\right]=\int_0^\infty\varsigma_t(\boldsymbol{\theta})\frac{\partial h\left[\varsigma_t(\boldsymbol{\theta})\right]}{\partial\varsigma_t}=-1.$$

Similarly, we can also use the definition (3) to show that

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

As a result, we will have that for any spherically symmetric distribution

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

so that

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

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

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

$$\begin{aligned}&E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa_0)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]\mathbf{e}'_{dt}(\boldsymbol{\theta}_0,0)\,|\mathbf{z}_t,I_{t-1},\boldsymbol{\phi}_0\right\}\\&=E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^+(\kappa_0)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta}_0,0)\,|\mathbf{z}_t,I_{t-1},\boldsymbol{\phi}_0\right\}=\mathbf{0}\end{aligned}$$

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

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

Hence, even though $\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]$ is the residual from the theoretical regression of $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})$ on a constant and $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},0)$, it turns out that the second summand of (5) belongs to the restricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\varsigma_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,0)$.

Now, if write (6) as

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi})-\mathbf{Z}_{d}(\boldsymbol{\theta}_0)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)+\mathbf{Z}_{d}(\boldsymbol{\theta}_0)\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\eta_{0}\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0),$$

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

Finally, the expression for the semiparametric efficiency bound will be

$$E\left[\begin{array}{c}\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathbf{Z}_{d}(\boldsymbol{\theta})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]\right\}\\ \times\left\{\mathbf{e}_{dt}(\boldsymbol{\phi}_0)'\mathbf{Z}_{dt}'(\boldsymbol{\theta})-\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi}_0)-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta}_0,0)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\theta})\right\}\end{array}\middle|\boldsymbol{\phi}_0\right]$$
$$=E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}_{dt}'(\boldsymbol{\phi}_0)\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right]$$
$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi}_0)-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta}_0,0)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$-E\left\{\mathbf{Z}_{d}(\boldsymbol{\theta})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]\mathbf{e}_{dt}(\boldsymbol{\phi}_0)'\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$+E\left\{\mathbf{Z}_{d}(\boldsymbol{\theta})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,0)\right]\left[\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi}_0)-\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta}_0,0)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right\}$$
$$=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{Z}_{d}(\boldsymbol{\theta}_0)\left[\mathring{\mathcal{M}}_{dd}\left(\eta_{0}\right)-\mathring{\mathcal{K}}\left(0\right)\mathring{\mathcal{K}}^{+}\left(\kappa_{0}\right)\mathring{\mathcal{K}}\left(0\right)\right]\mathbf{Z}_{d}'(\boldsymbol{\theta}_0)$$

by virtue of the law of iterated expectations, and the definitions of $\mathring{\mathcal{M}}_{dd}\left(\eta_{0}\right)$, $\mathring{\mathcal{K}}\left(0\right)$ and $\mathring{\mathcal{K}}\left(\kappa_{0}\right)$. In this sense, note that

$$\mathring{\mathcal{M}}\left(\eta_{0}\right)=E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0\right]$$

because of (A7) and the expression for $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0)$ in (A6). □

Proposition 7

The proof that $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$ is at least as large as $\left[\mathcal{I}^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)\right]^{-1}$ in the positive semidefinite matrix sense follows trivially from the fact that

$$\frac{4(N+2)^2\nu^4}{(\nu-2)^2(N+2)^2(N+\nu+2)^2\mathcal{M}_{rr}(\eta)}\left[\begin{array}{c}\mathbf{0}\\ vec(\mathbf{I}_N)\end{array}\right]\left[\begin{array}{cc}\mathbf{0}' & vec'(\mathbf{I}_N)\end{array}\right]$$

is a positive semidefinite matrix because

$$\mathcal{M}_{rr}(\eta)=\frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{N+\nu}{2}\right)\right]-\frac{N\nu^4\left[\nu^2+N(\nu-4)-8\right]}{2\left(\nu-2\right)^2\left(N+\nu\right)\left(N+\nu+2\right)}>0$$

for any $\nu > 2$.

As for the difference between the feasible parametric efficiency bound, and the elliptically symmetric semiparametric efficiency bound, it suffices to prove that

$$\frac{4(N+2)^2\nu^4}{(\nu-2)^2(N+2)^2(N+\nu+2)^2\mathcal{M}_{rr}(\eta_0)} - \frac{8(N+2)}{N(N+\nu+2)(N+\nu-2)}$$

is non-negative for any $0 \leq \eta < 1/2$.

Now, the elliptically symmetric semiparametric efficiency bound will be larger than the regular one if

$$\frac{N(N+4)(N+\nu-2)}{2N(\nu-2)(N+\nu+2)(N+\nu-2)}(\mathbf{I}_{N^2}+\mathbf{K}_{NN})$$
$$-\frac{N(N+6-\nu)}{N(\nu-2)(N+\nu+2)(N+\nu-2)}vec(\mathbf{I}_N)vec'(\mathbf{I}_N) - \frac{8(N+2)}{N(N+\nu+2)(N+\nu-2)}vec(\mathbf{I}_N)vec'(\mathbf{I}_N)$$

is positive semidefinite.

Finally, the positive semidefiniteness of

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta})\left[\mathcal{M}_{dd}(\eta) - \mathcal{K}(0)\mathcal{K}^+(\kappa)\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}) - \mathcal{A}(\boldsymbol{\phi})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\phi})$$

can be proved along the same lines as Proposition 3 in Hafner and Rombouts (2004). □

Proposition 9

Since

$$\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\delta}} = -2\sum_{j=1}^{q}\alpha_j\left[y_{t-j}-\mu_{t-j}(\boldsymbol{\delta})\right]\frac{\partial\mu_{t-j}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}$$

then $\partial\sigma_t^2(\boldsymbol{\delta},\omega,\mathbf{0})/\partial\boldsymbol{\delta} = \mathbf{0}$. Hence, the expression for $\mathcal{I}_{\boldsymbol{\delta\delta}}(\boldsymbol{\phi})$ trivially follows from Proposition 1, and the fact that $\sigma_t^2(\boldsymbol{\delta},\omega,\mathbf{0}) = \omega$. In view of the block diagonality of the conditional information matrix between $\boldsymbol{\delta}$ and the remaining parameters, we can concentrate on $\omega, \boldsymbol{\alpha}$ and η in what follows without loss of generality.

Given that

$$\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\gamma} = \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}, \varepsilon_{t-1}^{*2}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-q}^{*2}(\boldsymbol{\theta})\right]'$$

it is straightforward to prove that

$$E\left[\mathcal{I}_{\gamma\gamma t}(\boldsymbol{\delta},\omega,\mathbf{0},\eta)\right|\boldsymbol{\delta},\omega,\mathbf{0},\eta] = \frac{\nu}{2(\nu+3)}\begin{bmatrix}\omega^{-2} & \omega^{-1}\boldsymbol{\iota}_q' \\ \omega^{-1}\boldsymbol{\iota}_q & 2(\nu-1)(\nu-4)^{-1}\mathbf{I}_q + \boldsymbol{\iota}_q\boldsymbol{\iota}_q'\end{bmatrix}$$

since the $\varepsilon_t(\theta_0)'s$ constitute a serially independent sequence under conditional homoskedasticity. Similarly, we can show that

$$E\left[\mathcal{I}_{\gamma\eta t}(\boldsymbol{\delta},\omega,\mathbf{0},\eta)\right|\boldsymbol{\delta},\omega,\mathbf{0},\eta] = \frac{3\nu^2}{(\nu-2)(\nu+1)(\nu+3)}\begin{pmatrix}\omega^{-1} \\ \boldsymbol{\iota}_q\end{pmatrix}$$

Finally, the partitioned inverse formula yields the required expression for $\mathcal{I}^{\gamma\gamma}(\boldsymbol{\delta},\omega,\mathbf{0},\eta)$. □

References

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

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

Andrews, D.W.K. (2001): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica* 69, 683-734.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* 31, 307-327.

Bollerslev, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics* 69, 542-547.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992): "ARCH modeling in finance: a review of the theory and empirical evidence", *Journal of Econometrics* 52, 5-59.

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

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

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

Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika* 64, 247-254.

Davies, R. B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika* 74, 33-43.

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

Diebold, F.X., Gunther, A.G., and Tay, A.S. (1998), "Evaluating density forecasts, with applications to financial risk management", *International Economic Review* 39, 863-883.

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

Durbin, J., 1970, Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, *Econometrica* 38, 410-421.

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

Engle, R.F. and Gonzalez-Rivera, G. (1991): "Semiparametric ARCH models", *Journal of*

Business and Economic Statistics 9, 345-360.

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

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

Hafner, nd rombouts (2004):

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

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

Hentschel, L. (1995): "All in the family: nesting symmetric and asymmetric GARCH models", *Journal of Financial Economics* 39, 71-104.

Koenker, R. (1981): "A note on studentizing a test for heteroskedasticity", *Journal of Econometrics* 17, 107-112.

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.

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

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.

Ogaki, M. (1993): "Generalized method of moments: econometric applications", in Maddala, G.S., C.R. Rao and H.D. Vinod (eds.), *Handbook of Statistics* vol. 11, Elsevier Science Publishers.

RiskMetrics Group (1996): *RiskMetrics Technical Document.*

Sentana, E. (1991): “Quadratic Arch models: a potential re-interpretation of Arch models”, Financial Markets Group Discussion Paper 122, London School of Economics.

Sentana, E. (1995): “Quadratic Arch Models”, *Review of Economic Studies* 62, 639-661.

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