



Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" - Viale Morgagni 59 - 50134 Firenze - www.ds.unifi.it

W O R K I N G P A P E R 2 0 1 2 / 1 0

Sequential estimation
of shape parameters in
multivariate dynamic models

D. Amengual, G. Fiorentini,
E. Sentana



Università degli Studi
di Firenze

Sequential estimation of shape parameters in multivariate dynamic models*

Dante Amengual

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

<amengual@cemfi.es>

Gabriele Fiorentini

Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy

<fiorentini@ds.unifi.it>

Enrique Sentana

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

<sentana@cemfi.es>

February 2012

Abstract

Sequential maximum likelihood and GMM estimators of distributional parameters obtained from the standardised innovations of multivariate conditionally heteroskedastic dynamic regression models evaluated at Gaussian PML estimators preserve the consistency of mean and variance parameters while allowing for realistic distributions. We assess the efficiency of those estimators, and obtain moment conditions leading to sequential estimators as efficient as their joint maximum likelihood counterparts. We also obtain standard errors for the quantiles required in VaR and CoVaR calculations, and analyse the effects on these measures of distributional misspecification. Finally, we illustrate the small sample performance of these procedures through Monte Carlo simulations.

Keywords: Elliptical Distributions, Efficient Estimation, Systemic Risk, Value at Risk.

JEL: C13, C32, G11

*We would like to thank Manuel Arellano, Christian Bontemps, Olivier Faugeras, Javier Mencía, Francisco Peñaranda, Marcos Sanso and David Veredas, as well as audiences at CEMFI, CREST, Princeton, Rimini, Toulouse, the Finance Forum (Granada, 2011) the Symposium of the Spanish Economic Association and the Conference in honour of M. Hashem Pesaran (Cambridge, 2011) for useful comments and suggestions. Of course, the usual caveat applies. Amengual and Sentana gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grants ECO 2008-00280 and 2011-26342.

1 Introduction

Nowadays it is well documented and widely recognised that the distribution of returns on financial assets such as stocks, bonds or currencies is usually rather leptokurtic, even after controlling for volatility clustering effects. Nevertheless, many empirical researchers continue to use the Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) among others because they remain root- T consistent for the conditional mean and variance parameters so long as those moments are correctly specified and the fourth moments are bounded.

However, academics and financial market participants are often interested in features of the distribution of asset returns beyond its conditional mean and variance. In particular, the Basel Capital Adequacy Accord forced banks and other financial institutions to develop models to quantify all their risks accurately. In practice, most institutions chose the so-called Value at Risk (VaR) framework in order to determine the capital necessary to cover their exposure to market risk. As is well known, the VaR of a portfolio of financial assets is defined as the positive threshold value V such that the probability of the portfolio suffering a reduction in wealth larger than V over some fixed time interval equals some pre-specified level $\lambda < 1/2$. Similarly, the recent financial crisis has highlighted the need for systemic risk measures that assess how an institution is affected when another institution, or indeed the entire financial system, is in distress. Given that the probability of the joint occurrence of several extreme events is regularly underestimated by the multivariate normal distribution, any such measure should definitely take into account the non-linear dependence induced by the non-normality of financial returns.

A rather natural modelling strategy is to specify a parametric leptokurtic distribution for the standardised innovations of the vector of asset returns, such as the multivariate Student t , and to estimate the conditional mean and variance parameters jointly with the parameters characterising the shape of the assumed distribution by maximum likelihood (ML) (see for example Pesaran, Schleicher and Zaffaroni (2009) and Pesaran and Pesaran (2010)). Elliptical distributions such as the multivariate t are attractive in this context because they relate mean-variance analysis to expected utility maximisation (see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997)). Moreover, they generalise the multivariate normal distribution, but at the same time they retain its analytical tractability irrespective of the number of assets. However, the problem with non-Gaussian ML estimators is that they often achieve efficiency gains under cor-

rect specification at the risk of returning inconsistent parameter estimators under distributional misspecification, as shown by Newey and Steigerwald (1997). Unfortunately, semiparametric estimators of the joint density of the innovations suffer from the curse of dimensionality, which severely limits their use when the number of assets under consideration is moderately large. Another possibility would be semiparametric methods that impose the assumption of ellipticity, which retain univariate nonparametric rates regardless of the cross-sectional dimension of the data, but asymmetries in the true distribution will again contaminate the resulting estimators of conditional mean and variance parameters.

Sequential estimators of shape parameters that use the Gaussian PML estimators of the mean and variance parameters as first step estimators offer an attractive compromise because they preserve the consistency of the first two conditional moments under distributional misspecification, while allowing for more realistic conditional distributions. The focus of our paper is precisely the econometric analysis of sequential estimators obtained from the standardised innovations evaluated at the Gaussian PML estimators. Specifically, we consider not only sequential ML estimators, but also sequential generalised method of moments (GMM) estimators based on certain functions of the standardised innovations. Although we could easily extend our results to any multivariate distribution, to keep the exposition simple we focus on elliptical distributions. We illustrate our general results with several examples that nest the normal, including the Student t and the original Kotz (1975) distribution, as well as some rather flexible families such as scale mixtures of normals and polynomial expansions of the multivariate normal density, both of which could form the basis for a proper nonparametric procedure.

We explain how to compute asymptotic standard errors of sequential estimators that take into account the sampling variability of the Gaussian PML estimators on which they are based. We also exploit the expressions of the standard errors to assess the relative efficiency of sequential estimators, and obtain the optimal moment conditions that lead to sequential MM estimators which are as efficient as their joint ML counterparts. Although we explicitly consider multivariate conditionally heteroskedastic dynamic regression models, our results obviously apply in univariate contexts as well as in static ones.

We then analyse the use of our sequential estimators in the computation of commonly used risk management measures such as VaR, and recently proposed systemic risk measures such as Conditional Value at Risk (CoVaR) (see Adrian and Brunnermeier (2011)). In particular,

we compare our sequential estimators to nonparametric estimators, both when the parametric conditional distribution is correctly specified and also when it is misspecified. Our analytical and simulation results indicate that the use of sequential ML estimators of flexible parametric families of distributions in estimating those risk measures offers substantial efficiency gains, while incurring in small biases.

The rest of the paper is organised as follows. In section 2, we introduce the model, present closed-form expressions for the score vector and the conditional information matrix, and derive the asymptotic variances of the ML and Gaussian PML estimators. Then, in section 3 we introduce the sequential ML and GMM estimators, and compare their efficiency. In section 4, we study the influence of those estimators on risk measures under both correct specification and misspecification, and derive asymptotically valid standard errors. A Monte Carlo evaluation of the different parameter estimators and risk measures can be found in section 5. Finally, we present our conclusions in section 6. Proofs and auxiliary results are gathered in appendices.

2 Theoretical background

2.1 The dynamic econometric model

Discrete time models for financial time series are usually characterised by an explicit dynamic regression model with time-varying variances and covariances. Typically the N dependent variables, \mathbf{y}_t , are assumed to be generated as:

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

where $\boldsymbol{\mu}(\cdot)$ and $\text{vech}[\boldsymbol{\Sigma}(\cdot)]$ are $N \times 1$ and $N(N+1)/2 \times 1$ vector functions known up to the $p \times 1$ vector of true parameter values $\boldsymbol{\theta}_0$, \mathbf{z}_t are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of \mathbf{y}_t and \mathbf{z}_t , $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular ‘‘square root’’ matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t^*$ is a martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$. Hence,

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

To complete the model, we need to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. We shall initially assume that, conditional on \mathbf{z}_t and I_{t-1} , $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as some particular member of the spherical family with a well defined density, or $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d. s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ for short, where $\boldsymbol{\eta}$ are some q additional shape parameters.

2.2 Elliptical distributions

A spherically symmetric random vector of dimension N , $\boldsymbol{\varepsilon}_t^*$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^* = e_t \mathbf{u}_t$, 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 $\boldsymbol{\varepsilon}_t^*$. The variables e_t and \mathbf{u}_t are referred to as the generating variate and the uniform base of the spherical distribution. Often, we shall also refer to $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime} \boldsymbol{\varepsilon}_t^*$, which trivially coincides with e_t^2 . Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^*$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^*) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*) = \mathbf{I}_N$. If we further assume that $E(e_t^4) < \infty$, then Mardia's (1970) coefficient of multivariate excess kurtosis

$$\kappa = E(\varsigma_t^2)/[N(N+2)] - 1, \quad (2)$$

will also be bounded.

Some examples of elliptical distributions that we use to illustrate our general results are:

Gaussian: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t} \mathbf{u}_t$ is distributed as a standardised multivariate normal if and only if ς_t is a chi-square random variable with N degrees of freedom. Since this involves no additional parameters, we shall identify the normal distribution with $\boldsymbol{\eta}_0 = \mathbf{0}$.

Student t : $\boldsymbol{\varepsilon}_t^* = \sqrt{\nu - 2} \times \sqrt{\zeta_t / \xi_t} \mathbf{u}_t$ is distributed as a standardised multivariate Student t if and only if ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν and variance 2ν , with \mathbf{u}_t , ζ_t and ξ_t mutually independent. Therefore, ς_t will be proportional to an F random variable with N and ν degrees of freedom. In this case, we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions.

Kotz: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t} \mathbf{u}_t$ is distributed as a standardised Kotz if and only if ς_t is a gamma random variable with mean N and variance $N[(N+2)\kappa+2]$, so that the coefficient of multivariate excess kurtosis itself is the shape parameter.

Discrete scale mixture of normals: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t} \mathbf{u}_t$ is distributed as a DSMN if and only if

$$\varsigma_t = \frac{s_t + (1 - s_t)\varkappa}{\alpha + (1 - \alpha)\varkappa} \zeta_t$$

where s_t is an independent Bernoulli variate with $P(s_t = 1) = \alpha$, \varkappa is the variance ratio of the two components, which for identification purposes we restrict to be in the range $(0, 1]$ and ζ_t is an independent chi-square random variable with N degrees of freedom. Effectively, in this case ς_t will be a two-component scale mixture of χ_N^2 s, with shape parameters α and \varkappa .

Polynomial expansion: $\varepsilon_t^* = \sqrt{\varsigma_t} \mathbf{u}_t$ is distributed as a J^{th} -order PE of the multivariate normal if and only if ς_t has a density defined by

$$h(\varsigma_t) = h_o(\varsigma_t) \cdot P_J(\varsigma_t),$$

where

$$h_o(\varsigma_t) = \frac{1}{2^{N/2} \Gamma(N/2)} \varsigma_t^{N/2-1} \exp\left(-\frac{1}{2}\varsigma_t\right)$$

denotes the density function of a χ^2 with N degrees of freedom, and

$$P_J(\varsigma_t) = \left[1 + \sum_{j=2}^J c_j p_{N/2-1,j}^g(\varsigma_t) \right]$$

is a J^{th} order polynomial written in terms of the generalised Laguerre polynomial of order j and parameter $N/2 - 1$, $p_{N/2-1,j}^g(\cdot)$. For example, the second and third order standardised Laguerre polynomials are:

$$\begin{aligned} p_{N/2-1,2}^g(\varsigma) &= \sqrt{\frac{2}{N(N+2)}} \left[\frac{N(N+2)}{4} - \left(\frac{N+2}{2}\right)\varsigma + \frac{1}{4}\varsigma^2 \right], \text{ and} \\ p_{N/2-1,3}^g(\varsigma) &= \sqrt{\frac{12}{N(N+2)(N+4)}} \\ &\times \left[\frac{N(N+2)(N+4)}{24} - \frac{(N+2)(N+4)}{8}\varsigma + \frac{N+4}{8}\varsigma^2 - \frac{1}{24}\varsigma^3 \right]. \end{aligned}$$

As a result, the $J - 1$ shape parameters will be given by c_2, c_3, \dots, c_J . The problem with polynomial expansions is that $h(\varsigma_t)$ will not be a proper density unless we restrict the coefficients so that $P_J(\varsigma)$ cannot become negative. For that reason, in Appendix C.1 we explain how to obtain restrictions on the c_j 's that guarantee the positivity of $P_J(\varsigma)$ for all ς . Figure 1 describes the region in (c_2, c_3) space in which densities of a 3^{rd} -order PE are well defined for all $\varsigma \geq 0$.

As we mentioned in the introduction, the multivariate Gaussian and Student t , which approaches the former as $\eta \rightarrow 0$ (or $\nu \rightarrow \infty$) but has otherwise fatter tails, have been by far the two most popular choices made by empirical researchers to model the conditional distribution of asset returns. But the other examples are more flexible even though they continue to nest the Gaussian distribution. In particular, the Kotz distribution reduces to the normal for $\kappa = 0$, but it can be either platykurtic ($\kappa < 0$) or leptokurtic ($\kappa > 0$). However, the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view. As for the DSMN, it approaches the multivariate normal when $\varkappa \rightarrow 1$, $\alpha \rightarrow 1$ or $\alpha \rightarrow 0$, although near those limits the distributions can be radically different (see Amengual and Sentana (2011))

for further details). Nevertheless, the distribution of any scale mixture of normals is necessarily leptokurtic.¹ Finally, the PE reduces to the spherical normal when $c_j = 0$ for all $j \in \{2, \dots, J\}$. Interestingly, while the distribution of $\boldsymbol{\varepsilon}_t^*$ is leptokurtic for a 2^{nd} order expansion, it is possible to generate platykurtic random variables with a 3^{rd} order expansion.

Figure 2 plots the densities of a normal, a Student t , a platykurtic Kotz distribution, a DSMN and a 3^{rd} -order PE in the bivariate case. Although they all have concentric circular contours because we have standardised and orthogonalised the two components, their densities can differ substantially in shape, and in particular, in the relative importance of the centre and the tails.

They also differ in the degree of cross-sectional “tail dependence” between the components, the normal being the only example in which lack of correlation is equivalent to stochastic independence. In this regard, Figure 3 plots the so-called exceedance correlation between uncorrelated marginal components (see Longin and Solnik, 2001). As can be seen, the distributions we consider have the flexibility to generate very different exceedance correlations, which will be particularly important for systemic risk measures.

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

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*'} \otimes \boldsymbol{\varepsilon}_t^*) &= \mathbf{0}, \\ E(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*'} \otimes \boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*'}) &= E[\text{vec}(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*'}) \text{vec}'(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*'})] = (\kappa + 1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \text{vec}(\mathbf{I}_N) \text{vec}'(\mathbf{I}_N)], \end{aligned}$$

where \mathbf{K}_{mn} is the commutation matrix of orders m and n and κ is defined in (2). An alternative characterisation can be based on the higher order moment parameter of spherical random variables introduced by Berkane and Bentler (1986), $\tau_m(\boldsymbol{\eta})$, which Maruyama and Seo (2003) relate to higher order moments as

$$E[\zeta_t^m | \boldsymbol{\eta}] = [1 + \tau_m(\boldsymbol{\eta})] E[\zeta_t^m | \mathbf{0}] \text{ where } E[\zeta_t^m | \mathbf{0}] = 2^m \prod_{j=1}^m (N/2 + j - 1). \quad (3)$$

For the elliptical examples mentioned above, we derive expressions for $\tau_m(\boldsymbol{\eta})$ in Appendix C.2.

A noteworthy property of these examples is that their moments are always bounded, with the

¹Dealing with discrete scale mixtures of normals with multiple components would be tedious but fairly straightforward. As is well known, multiple component mixtures can arbitrarily approximate the more empirically realistic continuous mixtures of normals such as symmetric versions of the hyperbolic, normal inverse Gaussian, normal gamma mixtures, Laplace, etc. The same is also true of polynomial expansions.

exception of the Student t . Appendix C.3 contains the moment generating functions for the Kotz, the DSMN and the 3rd-order PE.

2.3 The log-likelihood function, its score and information matrix

Let $\phi = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$ denote the $p + q$ parameters of interest, which we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\phi) = \sum_{t=1}^T l_t(\phi)$, with $l_t(\phi) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln |\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

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

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\phi) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varsigma} \frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})] \begin{bmatrix} \mathbf{e}_{lt}(\phi) \\ \mathbf{e}_{st}(\phi) \end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\phi), \\ \mathbf{s}_{\boldsymbol{\eta}t}(\phi) &= \partial c(\boldsymbol{\eta})/\partial \boldsymbol{\eta} + \partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial \boldsymbol{\eta} = \mathbf{e}_{rt}(\phi),\end{aligned}\quad (4)$$

where

$$\begin{aligned}\partial d_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})\text{vec}(\mathbf{I}_N) \\ \partial \varsigma_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\text{vec}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta})]\}, \\ \mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}), \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial \text{vec}'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial \boldsymbol{\theta} \cdot [\boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta})], \\ \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \\ \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \text{vec}\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta}) - \mathbf{I}_N\},\end{aligned}\quad (5)$$

$$\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \text{vec}\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta}) - \mathbf{I}_N\}, \quad (6)$$

$\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ and $\partial \text{vec}[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial \boldsymbol{\theta}'$ depend on the particular specification adopted, and

$$\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial \varsigma \quad (7)$$

can be understood as a damping factor that reflects the kurtosis of the specific distribution assumed for estimation purposes (see Appendix D.1 for further details). But since $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ is equal to 1 under Gaussianity, it is straightforward to check that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

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

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

$$\begin{aligned}\mathcal{I}_t(\boldsymbol{\phi}) &= V[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}'_t(\boldsymbol{\theta}) = -E[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}], \\ \mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},\end{aligned}$$

$\mathbf{h}_t(\boldsymbol{\phi})$ denotes the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ and $\mathcal{M}(\boldsymbol{\phi}) = V[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}]$.

The following result, which reproduces Proposition 2 in Fiorentini and Sentana (2010), contains the required expressions to compute the information matrix of the ML estimators:

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

$$\begin{aligned}\mathcal{M}(\boldsymbol{\eta}) &= \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}, \\ \mathcal{M}_{ll}(\boldsymbol{\eta}) &= V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathcal{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \\ \mathcal{M}_{ss}(\boldsymbol{\eta}) &= V[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathcal{M}_{ss}(\boldsymbol{\eta})(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + [\mathcal{M}_{ss}(\boldsymbol{\eta}) - 1]\text{vec}(\mathbf{I}_N)\text{vec}'(\mathbf{I}_N), \\ \mathcal{M}_{sr}(\boldsymbol{\eta}) &= E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\{\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\boldsymbol{\phi}\} = \text{vec}(\mathbf{I}_N)\mathcal{M}_{sr}(\boldsymbol{\eta}), \\ \mathcal{M}_{rr}(\boldsymbol{\eta}) &= V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E[\partial\mathbf{e}_{rt}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\boldsymbol{\phi}], \\ \mathcal{M}_{ll}(\boldsymbol{\eta}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \frac{\varsigma_t(\boldsymbol{\theta})}{N} \middle| \boldsymbol{\phi}\right\} = E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial\varsigma} \frac{\varsigma_t(\boldsymbol{\theta})}{N} + \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \middle| \boldsymbol{\phi}\right\}, \\ \mathcal{M}_{ss}(\boldsymbol{\eta}) &= \frac{N}{N+2} \left[1 + V\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \frac{\varsigma_t}{N} \middle| \boldsymbol{\phi}\right\}\right] = E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial\varsigma} \frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)} \middle| \boldsymbol{\phi}\right\} + 1, \\ \mathcal{M}_{sr}(\boldsymbol{\eta}) &= E\left[\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1\right\} \mathbf{e}'_{rt}(\boldsymbol{\phi}) \middle| \boldsymbol{\phi}\right] = -E\left\{\frac{\varsigma_t(\boldsymbol{\theta})}{N} \frac{\partial\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'} \middle| \boldsymbol{\phi}\right\}.\end{aligned}$$

Fiorentini, Sentana and Calzolari (2003) provide the relevant expressions for the multivariate standardised Student t , while the expressions for the Kotz distribution and the DSMN are given in Amengual and Sentana (2010).²

2.4 Gaussian pseudo maximum likelihood estimators of $\boldsymbol{\theta}$

If the interest of the researcher lied exclusively in $\boldsymbol{\theta}$, which are the parameters characterising the conditional mean and variance functions, then one attractive possibility would be to estimate a restricted version of the model in which $\boldsymbol{\eta}$ is set to zero. Let $\tilde{\boldsymbol{\theta}}_T = \arg \max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote

²The expression for $\mathcal{M}_{ss}(\boldsymbol{\kappa})$ for the Kotz distribution in Amengual and Sentana (2010) contains a typo. The correct value is $(N\boldsymbol{\kappa} + 2)/[(N + 2)\boldsymbol{\kappa} + 2]$.

such a PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root- T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. The proof is based on the fact that in those circumstances, the pseudo log-likelihood score, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$, is a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from $\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})$. The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result, which reproduces Proposition 3.2 in Fiorentini and Sentana (2010):³

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

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

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

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

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

3 Sequential estimators of the shape parameters

3.1 Sequential ML estimator of $\boldsymbol{\eta}$

Unfortunately, the normality assumption does not guarantee consistent estimators of other features of the conditional distribution of asset returns, such as its quantiles. Nevertheless, we can use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of $\boldsymbol{\eta}$ as $\tilde{\boldsymbol{\eta}}_T = \arg \max_{\boldsymbol{\eta}} L_T(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$.⁴

Interestingly, these sequential ML estimators of $\boldsymbol{\eta}$ can be given a rather intuitive interpretation. If $\boldsymbol{\theta}_0$ were known, then the squared Euclidean norm of the standardised innovations, $\varsigma_t(\boldsymbol{\theta}_0)$, would be *i.i.d.* over time, with density function

$$h(\varsigma_t; \boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)} \varsigma_t^{N/2-1} \exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})] \quad (8)$$

³Throughout this paper, we use the high level regularity conditions in Bollerslev and Wooldridge (1992) because we want to leave unspecified the conditional mean vector and covariance matrix in order to maintain full generality. Primitive conditions for specific multivariate models can be found for instance in Ling and McAleer (2003).

⁴In some cases there will be inequality constraints on $\boldsymbol{\eta}$, but for simplicity of exposition we postpone the details to Appendix D.1.

in view of expression (2.21) in Fang, Kotz and Ng (1990). Therefore, we could obtain the infeasible ML estimator of $\boldsymbol{\eta}$ by maximising with respect to this parameter the log-likelihood function of the observed $\varsigma_t(\boldsymbol{\theta}_0)'s$, $\sum_{t=1}^T \ln h[\varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\eta}]$. Although in practice the standardised residuals are usually unobservable, it is easy to prove from (8) that $\tilde{\boldsymbol{\eta}}_T$ is the estimator so obtained when we treat $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$ as if they were really observed.

Durbin (1970) and Pagan (1986) are two classic references on the properties of sequential ML estimators. A straightforward application of their results to our problem allows us to obtain the asymptotic distribution of $\tilde{\boldsymbol{\eta}}_T$, which reflects the sample uncertainty in $\tilde{\boldsymbol{\theta}}_T$:

Proposition 3 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \phi_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_0) \rightarrow N[0, \mathcal{F}(\phi_0)]$, where*

$$\mathcal{F}(\phi_0) = \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\phi_0) + \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\phi_0) \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\phi_0) \mathcal{C}(\phi_0) \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\phi_0) \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\phi_0). \quad (9)$$

Importantly, since $\mathcal{C}(\phi_0)$ will become unbounded as $\kappa_0 \rightarrow \infty$, the asymptotic distribution of $\tilde{\boldsymbol{\eta}}_T$ will also be non-standard in that case, unlike that of the feasible ML estimator $\hat{\boldsymbol{\eta}}_T$.

Expression (9) suggests that $\mathcal{F}(\phi_0)$ depends on the specification adopted for the conditional mean and variance functions. However, it turns out that the asymptotic dependence between estimators of $\boldsymbol{\theta}$ and estimators of $\boldsymbol{\eta}$ is generally driven by a scalar parameter, in which case $\mathcal{F}(\phi_0)$ does not depend on the functional form of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ or $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. To clarify this point, it is convenient to introduce the following reparametrisation:

Reparametrisation 1 *A homeomorphic transformation $\mathbf{r}(\cdot) = [\mathbf{r}'_1(\cdot), r'_2(\cdot)]'$ of the conditional mean and variance parameters $\boldsymbol{\theta}$ into an alternative set of parameters $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_1, \vartheta'_2)'$, where ϑ_2 is a scalar, and $\mathbf{r}(\boldsymbol{\theta})$ is twice continuously differentiable with $\text{rank}[\partial \mathbf{r}'(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that*

$$\left. \begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1) \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \vartheta_2 \boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1) \end{aligned} \right\} \quad \forall t, \quad (10)$$

with

$$E[\ln |\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)| | \phi_0] = k \quad \forall \boldsymbol{\vartheta}_1. \quad (11)$$

Expression (10) simply requires that one can construct pseudo-standardised residuals

$$\varepsilon_t^\circ(\boldsymbol{\vartheta}_1) = \boldsymbol{\Sigma}_t^{\circ-1/2}(\boldsymbol{\vartheta}_1)[\mathbf{y}_t - \boldsymbol{\mu}_t^\circ(\boldsymbol{\vartheta}_1)]$$

which are i.i.d. $s(\mathbf{0}, \vartheta_2 \mathbf{I}_N, \boldsymbol{\eta})$, where ϑ_2 is a global scale parameter, a condition satisfied by most static and dynamic models.⁵ But given that we can multiply this parameter by some scalar

⁵The only exceptions would be restricted models in which the overall scale is effectively fixed, or in which it is not possible to exclude ϑ_2 from the mean. In the first case, the information matrix will be block diagonal between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, so no correction is necessary, while in the second case the general expression in Proposition 3 applies.

positive smooth function of $\boldsymbol{\vartheta}_1$, $k(\boldsymbol{\vartheta}_1)$ say, and divide $\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)$ by the same function without violating (10), condition (11) simply provides a particularly convenient normalisation.⁶

We can then show the following result:

Proposition 4 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, and reparametrisation (1) is admissible, then*

$$\mathcal{F}(\boldsymbol{\phi}_0) = \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) + \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) \mathcal{M}'_{sr}(\boldsymbol{\eta}_0) \mathcal{M}_{sr}(\boldsymbol{\eta}_0) \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) \cdot \left(\frac{N}{2\vartheta_{20}} \right)^2 \mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_0) \quad (12)$$

where

$$\mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) = \frac{\{2(\kappa+1) + N\kappa\}}{4} \frac{4\vartheta_2^2}{N}$$

is the asymptotic variance of the feasible PML estimator of ϑ_2 , while the asymptotic variance of the feasible ML estimator of $\boldsymbol{\eta}$ is

$$\mathcal{I}^{\boldsymbol{\eta}}(\boldsymbol{\phi}_0) = \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) + \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) \mathcal{M}'_{sr}(\boldsymbol{\eta}_0) \mathcal{M}_{sr}(\boldsymbol{\eta}_0) \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) \cdot \left(\frac{N}{2\vartheta_{20}} \right)^2 \mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0), \quad (13)$$

with

$$\mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}) = \frac{1}{2\mathcal{M}_{ss}(\boldsymbol{\eta}) + N[\mathcal{M}_{ss}(\boldsymbol{\eta}) - 1 - \mathcal{M}_{sr}(\boldsymbol{\eta}) \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}) \mathcal{M}'_{sr}(\boldsymbol{\eta})]} \frac{4\vartheta_2^2}{N}.$$

In general, $\boldsymbol{\vartheta}_1$ or ϑ_2 will have no intrinsic interest. Therefore, given that $\tilde{\boldsymbol{\eta}}_T$ is numerically invariant to the parametrisation of conditional mean and variance, it is not really necessary to estimate the model in terms of those parameters for the above expressions to apply as long as it would be conceivable to do so. In this sense, it is important to stress that neither (12) nor (13) effectively depend on ϑ_2 , which drops out from those formulas.

It is easy to see from (9) that $\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\phi}_0) \leq \mathcal{I}^{\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \leq \mathcal{F}(\boldsymbol{\phi}_0)$ regardless of the distribution, with equality between $\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\phi}_0)$ and $\mathcal{F}(\boldsymbol{\phi}_0)$ if and only if $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) = \mathbf{0}$, in which case the sequential ML estimator of $\boldsymbol{\eta}$ will be $\boldsymbol{\theta}$ -adaptive, or in other words, as efficient as the infeasible ML estimator of $\boldsymbol{\eta}$ that we could compute if the $\varsigma_t(\boldsymbol{\theta}_0)$'s were directly observed. This condition simplifies to $\mathcal{M}_{sr}(\boldsymbol{\eta}_0) = \mathbf{0}$ when reparametrisation (1) is admissible.

A more interesting question in practice is the relationship between $\mathcal{I}^{\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ and $\mathcal{F}(\boldsymbol{\phi}_0)$. Theorem 5 in Pagan (1986) implies that

$$\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \hat{\boldsymbol{\eta}}_T) \rightarrow N[0, \mathcal{Y}(\boldsymbol{\phi}_0)],$$

⁶Bickel (1982) exploited this parametrisation in his study of adaptive estimation in the *iid* elliptical case, and so did Linton (1993) and Hodgson and Vorkink (2003) in univariate and multivariate GARCH-M models, respectively. As Fiorentini and Sentana (2010) show, in multivariate dynamic models with elliptical innovations (10) provides a general sufficient condition for the partial adaptivity of the ML estimators of $\boldsymbol{\vartheta}_1$ under correct specification, and for their consistency under distributional misspecification.

where

$$\mathcal{Y}(\phi_0) = \mathcal{I}_{\eta\eta}^{-1}(\phi_0)\mathcal{I}'_{\theta\eta}(\phi_0) \left\{ \mathcal{C}(\phi_0) - [\mathcal{I}_{\theta\theta}(\phi_0) - \mathcal{I}_{\theta\eta}(\phi_0)\mathcal{I}_{\eta\eta}^{-1}(\phi_0)\mathcal{I}'_{\theta\eta}(\phi_0)]^{-1} \right\} \mathcal{I}_{\theta\eta}(\phi_0)\mathcal{I}_{\eta\eta}^{-1}(\phi_0).$$

Therefore, the sequential ML estimator will be asymptotically as efficient as the joint ML estimator if and only if $\mathcal{Y}(\phi_0) = 0$. If reparametrisation (1) is admissible, the scalar nature of ϑ_2 implies that the only case in which $\mathcal{I}^{\eta\eta}(\phi_0) = \mathcal{F}(\phi_0)$ with $\mathcal{I}_{\theta\eta}(\phi_0) \neq \mathbf{0}$ will arise when the Gaussian PMLE of ϑ_2 is as efficient as the joint ML.⁷ Otherwise, there will be an efficiency loss.

3.2 Sequential GMM estimators of η

If we can compute the expectations of $L \geq q$ functions of ζ_t , $\mathbf{v}(\cdot)$ say, then we can also compute a sequential GMM estimator of η by minimising the quadratic form $\bar{\mathbf{n}}_T'(\tilde{\theta}_T, \eta)\mathbf{\Omega}\bar{\mathbf{n}}_T(\tilde{\theta}_T, \eta)$, where $\mathbf{\Omega}$ is a positive definite weighting matrix, and $\mathbf{n}_t(\theta, \eta) = \mathbf{v}[\zeta_t(\theta)] - E\{\mathbf{v}[\zeta_t(\theta)]|\phi\}$. When $L > q$, Hansen (1982) showed that if the long-run covariance matrix of the sample moment conditions has full rank, then its inverse will be the ‘‘optimal’’ weighting matrix, in the sense that the difference between the asymptotic covariance matrix of the resulting GMM estimator and an estimator based on any other norm of the same moment conditions is positive semidefinite.

This optimal estimator is infeasible unless we know the optimal matrix, but under additional regularity conditions, we can define an asymptotically equivalent but feasible two-step optimal GMM estimator by replacing it with an estimator evaluated at some initial consistent estimator of ϕ . An alternative way to make the optimal GMM estimator feasible is by explicitly taking into account in the criterion function the dependence of the long-run variance on the parameter values, as in the single-step Continuously Updated (CU) GMM estimator of Hansen, Heaton and Yaron (1996). As we shall see below, in our parametric models we can often compute these GMM estimators using analytical expressions for the optimal weighting matrices, which we would expect a priori to lead to better performance in finite samples.

Following Newey (1984), Newey (1985) and Tauchen (1985), we can obtain the asymptotic covariance matrix of the sample average of the influence functions evaluated at the Gaussian PML estimator, $\tilde{\theta}_T$, using the expansion

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \mathbf{n}_t(\tilde{\theta}_T, \eta_0) &= \frac{1}{T} \sum_{t=1}^T \mathbf{n}_t(\theta_0, \eta_0) - \mathcal{N}_{\mathbf{n}}\sqrt{T}(\tilde{\theta}_T - \theta_0) + o_p(1) \\ &= (I, -\mathcal{N}_{\mathbf{n}}\mathcal{A}^{-1})\frac{1}{T} \sum_{t=1}^T \begin{bmatrix} \mathbf{n}_t(\theta_0, \eta_0) \\ \mathbf{s}_{\theta}(\theta_0; \mathbf{0}) \end{bmatrix} + o_p(1) \end{aligned}$$

⁷The Kotz distribution provides a noteworthy example in which both $M_{sr}(\eta_0) = 0$ and $\mathcal{C}_{\vartheta_2\vartheta_2}(\phi_0) = \mathcal{I}^{\vartheta_2\vartheta_2}(\phi_0)$.

where

$$\mathcal{N}_{\mathbf{n}} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left(-\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta}'} \middle| \boldsymbol{\phi}_0 \right).$$

Hence, we immediately get that

$$\lim_{T \rightarrow \infty} V \left(\frac{\sqrt{T}}{T} \sum_{t=1}^T \mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0) \middle| \boldsymbol{\phi}_0 \right) = (I, -\mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1}) \begin{pmatrix} \mathcal{G}_{\mathbf{n}} & \mathcal{D}_{\mathbf{n}} \\ \mathcal{D}'_{\mathbf{n}} & \mathcal{B} \end{pmatrix} \begin{pmatrix} I \\ -\mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1} \end{pmatrix} = \mathcal{E}_{\mathbf{n}}, \quad (14)$$

where

$$\begin{pmatrix} \mathcal{G}_{\mathbf{n}} & \mathcal{D}_{\mathbf{n}} \\ \mathcal{D}'_{\mathbf{n}} & \mathcal{B} \end{pmatrix} = \lim_{T \rightarrow \infty} V \left(\frac{\sqrt{T}}{T} \sum_{t=1}^T \begin{bmatrix} \mathbf{n}_t(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \\ \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0; \mathbf{0}) \end{bmatrix} \middle| \boldsymbol{\phi}_0 \right).$$

An asymptotically equivalent way of dealing with parameter uncertainty replaces the original influence functions $\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})$ with the residuals from their *IV* regression onto $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$ using $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ as instruments.⁸ More formally:

Proposition 5 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then the optimal sequential GMM estimators based on $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$ and $\mathbf{n}_t^\perp(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$, where*

$$\mathbf{n}_t^\perp(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}),$$

will be asymptotically equivalent.

In those cases in which reparametrisation (1) is admissible, we can obtain a third and much simpler equivalent procedure by using the residuals from the alternative *IV* regression of $\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})$ onto $\varsigma_t(\boldsymbol{\theta})/N - 1$ using $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \varsigma_t(\boldsymbol{\theta})/N - 1$ as instrument. Specifically,

Proposition 6 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, and reparametrisation (1) is admissible, then the asymptotic variance of the sample average of*

$$\mathbf{n}_t^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{N}{2} \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right],$$

where

$$\mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}_2) = \text{Cov} \left[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}_1), \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}_2] \frac{\varsigma_t(\boldsymbol{\theta})}{N} \middle| \boldsymbol{\theta}, \boldsymbol{\eta}_1 \right],$$

is equal to (14), which reduces to

$$\mathcal{G}_{\mathbf{n}} - \frac{N}{2} [\mathbb{k}_{\mathbf{n}}(\mathbf{0}) \mathbb{k}'_{\mathbf{n}}(\boldsymbol{\eta}) + \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}) \mathbb{k}'_{\mathbf{n}}(\mathbf{0})] + \left[\frac{N}{2} + \frac{N(N+2)\kappa}{4} \right] \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}) \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta})'.$$

Finally, it is worth mentioning that when the number of moment conditions L is strictly larger than the number of shape parameters q , one could use the overidentifying restrictions statistic to test if the distribution assumed for estimation purposes is the true one.

⁸See Bontemps and Meddahi (2011) for alternative approaches in moment-based specification tests.

3.2.1 Higher order moments and orthogonal polynomials

The most obvious moments to use in practice to estimate the shape parameters are powers of ς_t . Specifically, we can consider the influence functions:

$$\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\varsigma_t^m(\boldsymbol{\theta})}{2^m \prod_{j=1}^m (N/2 + j - 1)} - [1 + \tau_m(\boldsymbol{\eta})]. \quad (15)$$

But given that for $m = 1$, expression (15) reduces to $\ell_{1t}(\boldsymbol{\theta}) = \varsigma_t(\boldsymbol{\theta})/N - 1$ irrespective of $\boldsymbol{\eta}$, we have to start with $m \geq 2$.

An alternative is to consider influence functions defined by the relevant m^{th} order orthogonal polynomials $p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = \sum_{h=0}^m a_h(\boldsymbol{\eta}) \varsigma_t^h(\boldsymbol{\theta})$.⁹ Again, we have to consider $m \geq 2$ because the first two non-normalised polynomials are always $p_{0t}[\varsigma_t(\boldsymbol{\theta})] = 1$ and $p_{1t}[\varsigma_t(\boldsymbol{\theta})] = \ell_{1t}(\boldsymbol{\theta})$, which do not depend on $\boldsymbol{\eta}$.

Given that $\{p_{1t}[\varsigma_t(\boldsymbol{\theta})], p_{2t}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \dots, p_{Mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}$ is a full-rank linear transformation of $[\ell_{1t}(\boldsymbol{\theta}), \ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \dots, \ell_{Mt}(\boldsymbol{\theta}, \boldsymbol{\eta})]$, the optimal joint GMM estimator of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ based on the first M polynomials would be asymptotically equivalent to the corresponding estimator based on the first M higher order moments. The following proposition extends this result to optimal sequential GMM estimators that keep $\boldsymbol{\theta}$ fixed at its Gaussian PML estimator, $\tilde{\boldsymbol{\theta}}_T$:

Proposition 7 *If $\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $E[\varsigma_t^{2M} | \boldsymbol{\eta}_0] < \infty$, the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, and reparametrisation (1) is admissible, then the optimal sequential estimator of $\boldsymbol{\eta}$ based on the orthogonal polynomials of order 2, 3, ..., M is asymptotically equivalent to the analogous estimator based on the corresponding higher order moments, with an asymptotic variance that takes into account the sample uncertainty in $\tilde{\boldsymbol{\theta}}_T$ given by*

$$\mathcal{G}_{\mathbf{P}} + \left[\frac{N}{2} + \frac{N(N+2)\kappa}{4} \right] \mathbb{k}_{\mathbf{P}}(\boldsymbol{\eta}) \mathbb{k}_{\mathbf{P}}(\boldsymbol{\eta})'$$

where $\mathcal{G}_{\mathbf{P}}$ is a diagonal matrix of order $M - 1$ with representative element

$$V[p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] | \boldsymbol{\phi}] = \sum_{h=0}^m \sum_{k=0}^m \left\{ a_h(\boldsymbol{\eta}) a_k(\boldsymbol{\eta}) [1 + \tau_{h+k}(\boldsymbol{\eta})] 2^{h+k} \prod_{j=1}^{h+k} (N/2 + j - 1) \right\}$$

and $\mathbb{k}_{\mathbf{P}}(\boldsymbol{\eta})$ is an $M - 1$ vector with representative element

$$\text{Cov} \left[p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \frac{\varsigma_t(\boldsymbol{\theta})}{N} \middle| \boldsymbol{\phi} \right] = \sum_{h=1}^m h a_h(\boldsymbol{\eta}) [1 + \tau_h(\boldsymbol{\eta}_0)] \frac{2^{h+1}}{N} \prod_{j=1}^h (N/2 + j - 1).$$

Importantly, these sequential GMM estimators will be not only asymptotically equivalent but also numerically equivalent if we use single-step GMM methods such as CU-GMM.

By using additional moments, we can in principle improve the efficiency of the sequential MM estimators, but the precision with which we can estimate $\tau_m(\boldsymbol{\eta})$ rapidly decreases with m .

⁹ Appendix B contains the expressions for the coefficients of the second and third order orthogonal polynomials of the different examples we consider.

3.2.2 Efficient sequential GMM estimators of η

Our previous GMM optimality discussion applies to a given set of moments. But one could also ask which estimating functions would lead to the most efficient sequential estimators of η taking into account the sampling variability in $\tilde{\theta}_T$. The following result answers this question by exploiting the characterisation of efficient sequential estimators in Newey and Powell (1998):

Proposition 8 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \phi_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \eta_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then the efficient influence function is given by the efficient parametric score of η :*

$$\mathbf{s}_{\eta|\theta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{s}_{\eta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{I}'_{\theta\eta}(\phi_0) \mathcal{I}_{\theta\theta}^{-1}(\phi_0) \mathbf{s}_{\theta t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \quad (16)$$

which is the residual from the theoretical regression of $\mathbf{s}_{\eta t}(\phi_0)$ on $\mathbf{s}_{\theta t}(\phi_0)$.

Importantly, the proof of this statement also implies that the resulting sequential MM estimator of η will achieve the efficiency of the feasible ML estimator, which is the largest possible. The reason is twofold. First, the variance of the efficient parametric score $\mathbf{s}_{\eta|\theta t}(\phi_0)$ in (16) coincides with the inverse of the asymptotic variance of the feasible ML estimator of η , $\hat{\eta}_T$. Second, this matrix is also the expected value of the Jacobian matrix of (16) with respect to η .

In those cases in which reparametrisation (1) is admissible, expression (16) reduces to

$$\mathbf{s}_{\eta|\theta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{s}_{\eta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{M'_{sr}(\boldsymbol{\eta})}{(1 + 2/N)M_{ss}(\boldsymbol{\eta}) - 1} \left[\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] \frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right], \quad (17)$$

which is once again much simpler to compute.

3.3 Efficiency comparisons

3.3.1 An illustration in the case of the Student t

In view of its popularity, it is convenient to illustrate our previous analysis in the case of the multivariate Student t . Given that when reparametrisation (1) is admissible Proposition 7 implies the coincidence between the asymptotic distributions of $\check{\eta}_T$ and $\hat{\eta}_T$, which are the sequential MM estimators of η based on the fourth moment and the second order polynomial, respectively, we first derive the distribution of those estimators in the general case:

Proposition 9 *If $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}, \phi_0$ is i.i.d. $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$, with $\nu_0 > 8$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) hold, then $\sqrt{T}(\check{\eta}_T - \eta_0) \rightarrow N[0, \mathcal{E}_\ell(\phi_0)/\mathcal{H}^2(\phi_0)]$ and*

$\sqrt{T}(\hat{\eta}_T - \eta_0) \rightarrow N[0, \mathcal{E}_p(\phi_0)/\mathcal{H}^2(\phi_0)]$, where

$$\begin{aligned}
\mathcal{E}_\ell(\phi_0) &= \mathcal{G}_\ell(\phi_0) + \mathcal{N}'_\ell(\phi_0)\mathcal{C}(\phi_0)\mathcal{N}_\ell(\phi_0) - 2\mathcal{N}'_\ell(\phi_0)\mathcal{A}^{-1}(\phi_0)\mathcal{D}_\ell(\phi_0), \\
\mathcal{E}_p(\phi_0) &= \mathcal{G}_p(\phi_0) + \mathcal{N}'_p(\phi_0)\mathcal{C}(\phi_0)\mathcal{N}_p(\phi_0), \\
\mathcal{D}_\ell(\phi_0) &= \text{cov}[\mathbf{s}_{\theta t}(\boldsymbol{\theta}_0, 0), \ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\phi_0] = \frac{4(\nu_0 - 2)(N + \nu_0 - 2)}{N(\nu_0 - 4)(\nu_0 - 6)}\mathbf{W}_s(\phi_0), \\
\mathcal{G}_\ell(\phi_0) &= V[\ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\phi_0] = \frac{(\nu_0 - 2)^2}{(\nu_0 - 4)^2} \left[\frac{(N + 6)(N + 4)}{N(N + 2)} \frac{(\nu_0 - 2)(\nu_0 - 4)}{(\nu_0 - 6)(\nu_0 - 8)} - 1 \right], \\
\mathcal{G}_p(\phi_0) &= V\{p_{2t}[\varsigma_t(\boldsymbol{\theta}_0), \eta_0]| \phi_0\} = \mathcal{G}_\ell(\phi_0) - \frac{8(\nu_0 - 2)^2(N + \nu_0 - 2)}{N(\nu_0 - 6)^2(\nu_0 - 4)}, \\
\mathcal{N}'_\ell(\phi_0) &= \text{cov}[\mathbf{s}_{\theta t}(\boldsymbol{\theta}_0, \eta_0), \ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\phi_0] = \frac{4(\nu_0 - 2)}{N(\nu_0 - 4)}\mathbf{W}_s(\phi_0), \\
\mathcal{N}'_p(\phi_0) &= \text{cov}\{\mathbf{s}_{\theta t}(\boldsymbol{\theta}_0, \eta_0), p_{2t}[\varsigma_t(\boldsymbol{\theta}_0), \eta_0]| \phi_0\} = -\frac{8(\nu_0 - 2)}{N(\nu_0 - 4)(\nu_0 - 6)}\mathbf{W}_s(\phi_0), \\
\mathcal{H}(\phi_0) &= \text{cov}[s_{\eta t}(\boldsymbol{\theta}_0, \eta_0), \ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\phi_0] = \text{cov}\{s_{\eta t}(\boldsymbol{\theta}_0, \eta_0), p_{2t}[\varsigma_t(\boldsymbol{\theta}_0), \eta_0]| \phi_0\} = \frac{2\nu_0^2}{(\nu_0 - 4)^2}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{W}_s(\phi_0) &= \mathbf{Z}_d(\phi_0)[\mathbf{0}', \text{vec}'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\phi_0][\mathbf{0}', \text{vec}'(\mathbf{I}_N)]' \\
&= E\left\{ \frac{1}{2} \partial \text{vec}'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)] / \partial \boldsymbol{\theta} \cdot \text{vec}[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)] \Big| \phi_0 \right\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\phi_0] = -E\{ \partial d_t(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} | \phi_0 \}. \quad (18)
\end{aligned}$$

The following proposition compares the efficiency of these estimators of η to the sequential ML estimator:

Proposition 10 *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \phi_0$ is i.i.d. $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ with $\nu_0 > 8$, then $\mathcal{F}(\phi_0) \leq \mathcal{J}(\phi_0)$. If in addition*

$$\mathcal{A}^{-1}(\phi_0)\mathbf{W}_s(\phi_0) = \frac{(N + \nu_0 - 2)}{(\nu_0 - 4)}\mathcal{B}^{-1}(\phi_0)\mathbf{W}_s(\phi_0), \quad (19)$$

then $\mathcal{J}(\phi_0) \leq \mathcal{G}(\phi_0)$, with equality if and only if

$$\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1 \right] - \frac{2(N + \nu_0 - 2)}{N(\nu_0 - 4)}\mathbf{W}'_s(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathbf{s}_{\theta t}(\boldsymbol{\theta}_0, 0) = 0 \quad \forall t. \quad (20)$$

The first part of the proposition shows that sequential ML is always more efficient than sequential MM based on the second order polynomial. Nevertheless, Proposition 8 implies that there is a sequential MM procedure that is more efficient than sequential ML. Condition (19) is trivially satisfied in the limiting case in which the Student t distribution is in fact Gaussian, and in dynamic univariate models with no mean. Also, it is worth mentioning that (20), which in turn implies (19), is satisfied by most dynamic univariate GARCH-M models (see Fiorentini, Sentana and Calzolari (2004)). More generally, condition (20) will hold in any model that satisfies reparametrisation (1).

Given that $\mathcal{I}_{\theta\eta}(\phi_0) = \mathbf{0}$ under normality from Proposition 1, it is clear that $\tilde{\eta}_T$ will be as asymptotically efficient as the feasible ML estimator $\hat{\eta}_T$ when $\eta_0 = \mathbf{0}$, which in turn is as efficient

as the infeasible ML estimator in that case. Moreover, the restriction $\eta \geq 0$ implies that these estimators will share the same half normal asymptotic distribution under conditional normality, although they would not necessarily be numerically identical when they are not zero. Similarly, the asymptotic distributions of $\check{\eta}_T$ and $\hat{\eta}_T$ will also tend to be half normal as the sample size increases when $\eta_0 = 0$, since $\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)$ is root- T consistent for κ , which is 0 in the Gaussian case. However, while $\hat{\eta}_T$ will always be as efficient as $\check{\eta}_T$ under normality because $p_{2t}[\zeta_t(\boldsymbol{\theta}), \eta]$ is proportional to $s_{\eta t}(\boldsymbol{\theta}_0, 0)$, $\check{\eta}_T$ will be less efficient unless condition (20) is satisfied.

Finally, note that since both $\mathcal{G}_\ell(\phi_0)$ and $\mathcal{G}_p(\phi_0)$ will diverge to infinity as ν_0 converges to 8 from above, $\check{\eta}_T$ and $\hat{\eta}_T$ will not be root- T consistent for $4 \leq \nu_0 \leq 8$. Moreover, since κ is infinite for $2 < \nu_0 \leq 4$, $\check{\eta}_T$ and $\hat{\eta}_T$ will not even be consistent in the interior of this range.

3.3.2 Asymptotic standard errors and relative efficiency

Figures 4 to 6 display the asymptotic standard deviation (top panels) and the relative efficiency (bottom panels) of the joint MLE and efficient sequential MM estimator, the sequential MLE, and finally the sequential GMM estimators based on orthogonal polynomials, obtained using the results in Propositions 4 and 7 under the assumption that reparametrisation (1) is admissible, which, as we mentioned before, covers most static and dynamic models.

Figure 4 refers to the Student t distribution. For slight departures from normality ($\eta < .02$ or $\nu > 50$) all estimators behave similarly. As η increases, the GMM estimators become relatively less efficient, with the exactly identified GMM estimator being the least efficient, as expected from Proposition 10. Notice, however, that when ν approaches 12 the GMM estimator based on the second and third orthogonal polynomials converges to the GMM estimator based only on the second one since the variance of the third orthogonal polynomial increases without bound. In turn, the variance of the estimator based on the second order polynomial blows up as ν converges to 8 from above, as we mentioned at the end of the previous subsection. Until roughly that point, the sequential ML estimator performs remarkably well, with virtually no efficiency loss with respect to the benchmark given by either the joint MLE or the efficient sequential MM. For smaller degrees of freedom, though, differences between the sequential and the joint ML estimators become apparent, especially for values of ν between 5 and 4.

The DSMN distribution has two shape parameters. In Figures 5a and 5b we maintain the scale ratio parameter \varkappa equal to .5 and report the asymptotic efficiency as a function of the mixing probability parameter α . In contrast, in Figures 5c and 5d we look at the asymptotic

efficiency of the different estimators fixing the mixing probability at $\alpha = 0.05$. Interestingly, we find that, broadly speaking, the asymptotic standard errors of the sequential MLE and the joint MLE are indistinguishable, despite the fact that the information matrix is not diagonal and the Gaussian PML estimators of $\boldsymbol{\theta}$ are inefficient, unlike in the case of the Kotz distribution. As for the GMM estimators, which in this case are well defined for every combination of parameter values, we find that the use of the fourth order orthogonal polynomial enhances efficiency except for some isolated values of α .

Finally, Figures 6a to 6d show the results for the PE distribution, with $c_2 = 0$ in the first two figures and $c_3 = 0$ in the other two. Again sequential MLE is very efficient with virtually no efficiency loss with respect to the benchmark. The GMM estimators are less efficient, but the use of the fourth order polynomial is very useful in estimating c_2 when $c_3 = 0$ and in estimating c_3 when $c_2 = 0$.

3.4 Misspecification analysis

So far we have maintained the assumption that the true conditional distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ is correctly specified. Although distributional misspecification will not affect the Gaussian PML estimator of $\boldsymbol{\theta}$, the sequential estimators of $\boldsymbol{\eta}$ will be inconsistent if the true distribution of $\boldsymbol{\varepsilon}_t^*$ given \mathbf{z}_t and I_{t-1} does not coincide with the assumed one. To focus our discussion on the effects of distributional misspecification, in the remaining of this section we shall assume that (1) is true.

Let us consider situations in which the true distribution is *i.i.d.* elliptical but different from the parametric one assumed for estimation purposes, which will often be chosen for convenience or familiarity. For simplicity, we define the pseudo-true values of $\boldsymbol{\eta}$ as consistent roots of the expected pseudo log-likelihood score, which under appropriate regularity conditions will maximise the expected value of the pseudo log-likelihood function. We can then prove that:

Proposition 11 *If $\boldsymbol{\varepsilon}_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0$, is *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, and reparametrisation (1) is admissible, then the asymptotic distribution of the sequential ML estimator of $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\eta}}_T$, will be given by*

$$\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_\infty) \rightarrow N \left\{ \mathbf{0}, \mathcal{H}_{rr}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \mathcal{E}_r(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \mathcal{H}_{rr}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \right\},$$

where $\boldsymbol{\eta}_\infty$ solves $E[\mathbf{e}_{rt}(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty) | \boldsymbol{\varphi}_0] = \mathbf{0}$,

$$\mathcal{H}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = \mathcal{M}_{rr}^H(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty; \boldsymbol{\varphi}_0), \quad \mathcal{M}_{rr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E[\partial \mathbf{e}_{rt}(\boldsymbol{\phi}) / \partial \boldsymbol{\eta}' | \boldsymbol{\varphi}],$$

and

$$\begin{aligned} \mathcal{E}_r(\phi_\infty; \varphi_0) &= [\mathcal{O}_{rr}(\phi_\infty; \varphi_0)]^{-1} + \frac{N\{2[\kappa_0 + 1] + N\kappa_0\}}{4} \\ &\quad \times [\mathcal{O}_{rr}(\phi_\infty; \varphi_0)]^{-1} \mathbf{M}_{sr}^O(\phi_\infty; \varphi_0)' \mathbf{M}_{sr}^O(\phi_\infty; \varphi_0) [\mathcal{O}_{rr}(\phi_\infty; \varphi_0)]^{-1}, \end{aligned}$$

with

$$\begin{aligned} \mathbf{M}_{sr}^O(\phi; \varphi) &= E[\{\delta[\varsigma_t(\vartheta), \boldsymbol{\eta}] \cdot [\varsigma_t(\vartheta)/N] - 1\} \mathbf{e}_{rt}(\phi) | \varphi], \\ \mathcal{O}_{rr}(\phi_\infty; \varphi_0) &= \mathcal{M}_{rr}^O(\vartheta_0, \boldsymbol{\eta}_\infty; \varphi_0), \quad \mathcal{M}_{rr}^O(\phi; \varphi) = V[\mathbf{e}_{rt}(\phi) | \varphi]. \end{aligned}$$

4 Application to risk measures

As we mentioned in the introduction, nowadays many institutional investors all over the world regularly use risk management procedures based on the ubiquitous VaR to control for the market risks associated with their portfolios. Furthermore, the recent financial crisis has highlighted the need for systemic risk measures that point out which institutions would be most at risk should another crisis occur. In that sense, Adrian and Brunnermeier (2011) propose to measure the systemic risk of individual institutions by means of the so-called Exposure CoVaR, which they define as the VaR of financial institution i when the entire financial system is in distress. To gauge the usefulness of our results in practice, in this section we focus on the role that the shape parameter estimators play in the reliability of those risk measures.¹⁰

For illustrative purposes, we consider a simple dynamic market model, in which reparametrisation (1) is admissible. Specifically, if r_{Mt} denotes the excess returns on the market portfolio, and r_{it} the excess returns on asset i ($i = 2, \dots, N$), we assume that $\mathbf{r}_t = (r_{Mt}, r_{2t}, \dots, r_{Nt})$ is generated as

$$\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})[\mathbf{r}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})] | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d. s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}),$$

with

$$\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \left[\begin{array}{c} \mu_{Mt} \\ \mathbf{a}_t(\boldsymbol{\theta}) + \mathbf{b}_t(\boldsymbol{\theta})\mu_{Mt} \end{array} \right] \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \left[\begin{array}{cc} \sigma_{Mt}^2 & \sigma_{Mt}\mathbf{b}'_t(\boldsymbol{\theta}) \\ \sigma_{Mt}\mathbf{b}_t(\boldsymbol{\theta}) & \sigma_{Mt}^2\mathbf{b}_t(\boldsymbol{\theta})\mathbf{b}'_t(\boldsymbol{\theta}) + \boldsymbol{\Omega}_t(\boldsymbol{\theta}) \end{array} \right] \\ \sigma_{Mt}^2 &= \sigma_M^2 + \gamma(\varepsilon_{Mt-1}^2 - \sigma_M^2) + \beta(\sigma_{Mt-1}^2 - \sigma_M^2). \end{aligned} \tag{21}$$

In this model, μ_{Mt} and σ_{Mt}^2 denote the conditional mean and variance of r_{Mt} , while $\mathbf{a}_t(\boldsymbol{\theta})$ and $\mathbf{b}_t(\boldsymbol{\theta})$ are the alpha and beta of the other $N - 1$ assets with respect to the market portfolio,

¹⁰Acharya et al. (2010) and Brownlees and Engle (2011) consider a closely related systemic risk measure, the Marginal Expected Shortfall, which they define as the expected loss an equity investor in a financial institution would experience if the overall market declined substantially. It would be tedious but straightforward to extend our analysis to that measure.

respectively, and $\mathbf{\Omega}_t(\boldsymbol{\theta})$ their residual covariance matrix. Given that the portfolio of financial institutions changes every day, a multivariate framework such as this one offers important advantages over univariate procedures because we can compute the different risk management measures in closed form from the parameters of the joint distribution without the need to re-estimate the model.¹¹

4.1 VaR and Exposure CoVaR

Let $W_{t-1} > 0$ denote the initial wealth of a financial institution which can invest in a safe asset with gross returns R_{0t} , and N risky assets with excess returns \mathbf{r}_t . Let $\mathbf{w}_t = (w_{Mt}, w_{2t}, \dots, w_{Nt})'$ denote the weights on its chosen portfolio. The random final value of its wealth over a fixed period of time, which we normalise to 1, will be

$$W_{t-1}R_{wt} = W_{t-1}(R_{0t} + r_{wt}) = W_{t-1}(R_{0t} + \mathbf{w}_t'\mathbf{r}_t).$$

This value contains both a safe component, $W_{t-1}R_{0t}$, and a random component, $W_{t-1}r_{wt}$. Hence, the probability that this institution suffers a reduction in wealth larger than some fixed positive threshold value V_t will be given by the following expression

$$\begin{aligned} \Pr[W_{t-1}(1 - R_{0t}) - W_{t-1}r_{wt} \geq V_t] &= \Pr(r_{wt} \leq 1 - R_{0t} - V_t/W_{t-1}) \\ &= \Pr\left[\frac{r_{wt} - \mu_{wt}}{\sigma_{wt}} \leq \frac{1 - R_{0t} - V_t/W_{t-1} - \mu_{wt}}{\sigma_{wt}}\right] = F\left[\frac{1 - R_{0t} - V_t/W_{t-1} - \mu_{wt}}{\sigma_{wt}}\right], \end{aligned}$$

where $\mu_{wt} = \mathbf{w}_t'\boldsymbol{\mu}_t$ and $\sigma_{wt}^2 = \mathbf{w}_t'\boldsymbol{\Sigma}_t\mathbf{w}_t$ are the expected excess return and variance of r_{wt} , and $F(\cdot)$ is the cumulative distribution function of a zero mean - unit variance random variable within the appropriate elliptical class.¹²

The value of V_t which makes the above probability equal to some pre-specified value λ ($0 < \lambda < 1/2$) is known as the $100(1 - \lambda)\%$ VaR of the portfolio R_{wt} . For convenience, though, the portfolio VaR is often reported in fractional form as $-V_t/W_{t-1}$. Consequently, if we define $q_1(\lambda, \boldsymbol{\eta})$ as the λ^{th} quantile of the distribution of standardised returns, which will be negative for $\lambda < 1/2$, the reported figure will be given by

$$V_t/W_{t-1} = 1 - R_{0t} - \mu_{wt} - \sigma_{wt}q_1(\lambda, \boldsymbol{\eta}).$$

¹¹An attractive property of using parametric methods for VaR and CoVaR estimation is that it guarantees quantiles that do not cross.

¹²Due to the properties of the elliptical distributions (see theorem 2.16 in Fang et al (1990)), the cumulative distribution function $F(\cdot)$ does not depend in any way on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ or the vector of portfolio weights, only on the vector of shape parameters $\boldsymbol{\eta}$.

By definition, the Exposure CoVaR of a financial institution will be very much influenced by the market beta of its portfolio. To isolate tail dependence from the linear dependence induced by correlations, in what follows we focus on the CoVaR of an institution after hedging its market risk component. More formally, if

$$r_{ht} = r_{wt} - \frac{Cov_{t-1}(r_{wt}, r_{Mt})}{V_{t-1}(r_{Mt})} r_{Mt}$$

denotes the idiosyncratic risk component of portfolio R_{wt} , we look at the Exposure CoVaR of r_{ht} . To simplify the exposition, we assume that $\mathbf{a}_t(\boldsymbol{\theta}) = \mathbf{0}$, $\mathbf{b}_t(\boldsymbol{\theta}) = \mathbf{b}$ and $\boldsymbol{\Omega}_t(\boldsymbol{\theta}) = \boldsymbol{\Omega}$, so that the conditional mean of r_{ht} is 0 and its variance $\sigma_h^2 = \sum_{j=2}^N w_{jt}^2 \omega_j$. In this context, the specific Exposure CoVaR, CV_t , will be implicitly defined by

$$q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}) = \frac{1}{\sigma_w^h} \left[1 - R_{0t} - \frac{CV_t}{W_{t-1} \sum_{j=2}^N w_{jt}} \right],$$

where $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$ denotes the λ_2^{th} quantile of the (standardised) distribution of r_{ht} conditional on the market return r_{Mt} being *below* its λ_1^{th} quantile. More formally,

$$\begin{aligned} \lambda_2 &= \Pr [\varepsilon_{ht}^* \leq q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}) | \varepsilon_{Mt}^* \leq q_1(\lambda_1, \boldsymbol{\eta})] \\ &= \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} f_1(\varepsilon_{1t}^*, \boldsymbol{\eta}) \left[\int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} f_{2|1}(\varepsilon_{2t}^*, \varepsilon_{1t}^*; \boldsymbol{\eta}) d\varepsilon_{2t}^* \right] d\varepsilon_{1t}^*, \end{aligned}$$

with

$$q_1(\lambda, \boldsymbol{\eta}) = \frac{1}{\sigma_{Mt}} \left[1 - R_0 - \mu_{Mt} - \frac{V_t}{w_M W_{t-1}} \right].$$

In Appendix C.4 we provide the conditional and marginal cumulative distribution functions required to obtain $q_1(\lambda, \boldsymbol{\eta})$ and $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$ for the multivariate Student t , DSMN and 3^{rd} -order PE, on the basis of which we compute the parametric VaR and CoVaR measures.

4.2 The effect of sampling uncertainty on parametric VaR and CoVaR

In practice, the above expressions will be subject to sampling variability in the estimation of μ_w , σ_w , $q_1(\lambda, \boldsymbol{\eta})$ and $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$. Given that we are assuming that μ_w and σ_w will be evaluated at their Gaussian PML estimators in all cases, in the rest of this section we shall focus on the sampling variability in estimating $q_1(\lambda, \boldsymbol{\eta})$ and $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$.

In parametric models, $q_1(\lambda, \boldsymbol{\eta})$ and $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$ would be known with certainty for all values of λ regardless of the sample size if we assumed we knew the true value of $\boldsymbol{\eta}$, $\boldsymbol{\eta}_0$. More generally, though, we have to take into account the variability in estimating $\boldsymbol{\eta}$. Asymptotic valid standard

errors for those quantiles can be easily obtained by a direct application of the delta method. Appendix C.5 contains the required expressions for $\partial q_1(\lambda, \boldsymbol{\eta})/\partial \boldsymbol{\eta}$ and $\partial q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})/\partial \boldsymbol{\eta}$. On the basis of those expressions, Figure 7 displays confidence bands for parametric VaR and CoVaR computed with the Student t (7a), DSMN (7b) and PE (7c) distributions. To save space, we only look at the 1% and 5% significance levels for the case in which $\lambda_1 = \lambda_2$. The dotted lines represent the 95% confidence intervals based on the asymptotic variance of the sequential ML estimator for a hypothetical sample size of $T = 1,000$ and $N = 5$. As expected, the confidence bands are larger at the 1% level than at the 5% one. They are also larger for CoVaR than for VaR, the intuition being that the number of observations effectively available is smaller. These figures also illustrate that the assumption of Gaussianity could be rather misleading even in situations where the actual DGP has moderate excess kurtosis. This is particularly true for the VaR figures at the 99% level, and especially for the CoVaR numbers at both levels.

4.3 A comparison of parametric and nonparametric VaR figures under correct specification and under misspecification

The so-called historical method is a rather popular way of computing VaR figures employed by many financial institutions all over the world. Some of the most sophisticated versions of this method rely on the empirical quantiles of the returns to the current portfolio over the last T observations after correcting for time-varying expected returns, volatilities and correlations (see Gouriéroux and Jasiak (2009) for a recent survey). Since this is a fully non-parametric procedure, the asymptotic variance of the λ^{th} empirical quantile of the standardised return distribution will be given by

$$\frac{\lambda(1-\lambda)}{f^2[q_1(\lambda)]}, \quad (22)$$

where $f(\cdot)$ denotes the true density function (see p. 72 in Koenker (2005)).

By construction, the empirical quantile ignores any restriction on the distribution of standardised returns. A simple way of imposing symmetry would be to average (the absolute value of) the sample counterparts to $q_1(\lambda)$ and $q_1(1-\lambda)$. Using again the results in Koenker (2005) for the joint asymptotic distribution of two sample quantiles, it is straightforward to prove that the corresponding asymptotic variance will be

$$\frac{\lambda}{2f^2[q_1(\lambda)]},$$

which is lower than (22) for any $\lambda < 1/2$. This way of enforcing symmetry, though, is inefficient,

as can be seen by looking at the limiting case of the median, for which it offers no gains. The most efficient estimator of $q_1(\lambda)$ that imposes symmetry turns out to be the $(1 - 2\lambda)^{th}$ quantile of the empirical distribution of the absolute values of the standardised returns. Given that the density of the absolute values is twice the density of the original variable, the asymptotic variance of this quantile estimator will be given by

$$\frac{\lambda(1 - 2\lambda)}{2f^2[q_1(\lambda)]}.$$

As expected, this variance goes to 0 as $\lambda \rightarrow 1/2$ since we are considering standardised returns, and the symmetry assumption implies that mean and median should coincide.

It is interesting to relate the asymptotic variances of these non-parametric quantile estimators to the asymptotic variance implied by parametric models. In Appendix D.5 we show that the asymptotic variance of $q_1(\lambda, \tilde{\boldsymbol{\eta}}_T)$ can be written as

$$\frac{\lambda(1 - \lambda)}{f^2[q_1(\lambda, \boldsymbol{\eta}_0); \boldsymbol{\eta}_0]} E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta}_0), \boldsymbol{\eta}_0] V[\tilde{\boldsymbol{\eta}}_T | \boldsymbol{\eta}_0] E[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \geq q_1(\lambda, \boldsymbol{\eta}_0), \boldsymbol{\eta}_0], \quad (23)$$

which coincides with (22) multiplied by a damping factor. Importantly, the distribution used to compute the foregoing expectation is the same as the distribution used for estimation purposes. Hence, this expression continues to be valid under misspecification of the conditional distribution, although in that case we must use a robust (sandwich) formula to obtain $V[\tilde{\boldsymbol{\eta}}_T | \boldsymbol{\varphi}_0]$. Specifically, if $\varepsilon_t^* | \mathbf{z}_t, I_{t-1}; \boldsymbol{\varphi}_0$, is *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\theta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then the asymptotic variance of the sequential ML estimator of $q_1(\lambda, \tilde{\boldsymbol{\eta}}_T)$ will still be given by (23), but with $\boldsymbol{\eta}_0$ replaced by the pseudo-true value of $\boldsymbol{\eta}$ defined in Proposition 11, $\boldsymbol{\eta}_\infty$.

The top panels of Figures 8a-c display the 99% VaR numbers corresponding to the Student t , DSMN and PE distributions obtained with the different sequential ML estimators both under correct specification and under misspecification. Asymptotic standard errors for the parametric estimators are shown in the bottom panels. Those figures also contain standard errors for the λ^{th} empirical quantile of the standardised return distribution, and the $(1 - 2\lambda)^{th}$ quantile of the empirical distribution of the absolute values of the standardised returns, which are labeled as NP and SNP, respectively. As can be seen, the two non-parametric quantile estimators are always consistent but largely inefficient. In contrast, the parametric estimators have fairly narrow variation ranges, but they can be sometimes noticeably biased under misspecification, especially

when they rely on the Student t . In contrast, the biases due to distributional misspecification seem to be small when one uses flexible distributions such as DSMNs and PEs.

5 Monte Carlo Evidence

5.1 Design and estimation details

In this section, we assess the finite sample performance of the different estimators and risk measures discussed above by means of an extensive Monte Carlo exercise, with an experimental design based on (21). Specifically, we simulate and estimate a model in which $N = 5$, $\mu_M = 0.1$, $\mathbf{a} = \mathbf{0}$, $\mathbf{b} = (1, 2, 1, 2)$, $\mathbf{\Omega} = I_4$, $\gamma = 0.1$ and $\beta = 0.85$. As for $\boldsymbol{\varepsilon}_t^*$, we consider a multivariate Student t with 10 degrees of freedom, a DSMN with the same kurtosis and $\alpha = 0.05$, and finally a 3rd-order PE also with the same kurtosis and $c_3 = -1$.

Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 2,500$ observations (plus another 100 for initialisation) based on 1,600 Monte Carlo replications. This sample size corresponds roughly to 10 years of daily data. The numerical strategy employed by our estimation procedure is described in Appendix D.2. Given that the Gaussian PML estimators of $\boldsymbol{\theta}$ are unbiased, and they share the same asymptotic distribution under the different distributional assumptions because of their common kurtosis coefficient, we do not report results for $\tilde{\boldsymbol{\theta}}_T$ in the interest of space.

5.2 Sampling distribution of the different estimators of $\boldsymbol{\eta}$

Table 1 presents means and standard deviations of the sampling distributions for three different sequential estimators of the shape parameters under correct specification. Specifically, we consider sequential ML (SML), efficient sequential MM (ESMM), and orthogonal polynomial-based MM estimators that use the 2nd polynomial in the case of the Student t , and the 2nd and 3rd for the other two. The top panel reports results for the Student t , while the middle and bottom panels contain statistics for DSMN and the 3rd-order PE, respectively.

By and large, the behavior of the different estimators is in accordance with what the asymptotic results from Section 3.4 would suggest. In particular, the standard deviations of ESMM and SML essentially coincide, as expected from Figures 4-6. In contrast, the exactly identified orthogonal polynomials-based estimator is clearly inefficient relative to the others, which is also in line with the asymptotic standard errors in Figures 4-6. This is particularly noticeable in the case of the PE, as the sampling standard deviation of the SMM-based estimator of c_3 more

than doubles those of ESMM and SML.

Another thing to note is that the estimators of the DSMN parameters α and \varkappa seem to be slightly upward biased, and that the bias increases when those parameters are estimated using MM orthogonal polynomials. The same comment applies to the 3^{rd} -order PE parameters c_2 and c_3 . In that case, however, the estimators tend to underestimate the true magnitude of the parameters.

5.3 Sampling distribution of VaR and CoVaR measures

Having sequentially estimated the parameters of the three distributions that we are considering in each simulated sample, we then computed parametric VaR and CoVaR measures using the conditional and marginal CDFs in Appendix C.4. In the interest of space, we report results based on the sequential ML estimator of $\boldsymbol{\eta}$ only. As for the historical VaR and CoVaR, we focus on the λ^{th} empirical quantile of the relevant standardised distribution, which we estimate by linear interpolation in order to reduce potential biases in small samples.¹³ The objective of our exercise is twofold: 1) to shed some light on the finite sample performance of parametric and non-parametric VaR and CoVaR estimators; and 2) to assess the effects of distributional misspecification on the latter.

Figure 9a summarises the sampling distribution of the different 99% VaR measures by means of box-plots with a different DGP on each panel. As usual, the central boxes describe the first and third quartiles of the sampling distributions, as well as their median, and we set the maximum length of the whiskers to one interquartile range. Each panel contains five rows with the three SML-based measures, as well as the non-parametric one (denoted by NP) and the Gaussian quantile as a reference.

When the true distribution is Student t , all the parametric VaR measures perform well, in the sense that their sampling distributions are highly concentrated around the true value. In contrast, the sampling uncertainty of the 1% non-parametric quantile is much bigger. The same comments apply when the DGPs are either DSMN or PE distributions, although in those cases, the bias of the misspecified Student t -based VaR is pronounced.

The same general pattern emerges in Figure 9b, which compares the 95% CoVaR measures. For the distributions we use as examples, the effects of distributional misspecification seem to be

¹³Alternatively, we could obtain estimates of the CDF by integrating a kernel density estimator, but the first-order asymptotic properties of the associated quantiles would be the same (see again Koenker (2005)).

minor compared to the potential efficiency gains from using a parametric model for estimating the quantiles. This is particularly true when we use flexible distributions such as DSMNs or PEs to conduct inference.

6 Conclusions

In the context of the general multivariate dynamic regression model with time-varying variances and covariances considered by Bollerslev and Wooldridge (1992), we study the statistical properties of sequential estimators of the shape parameters of the conditional distributions, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators. In particular, we consider sequential ML estimators, as well as sequential GMM estimators. The main advantage of such estimators is that they preserve the consistency of the conditional mean and variance functions, but at the same time allow for a more realistic conditional distribution. We pay special attention to elliptical distributions such as the Student t and Kotz distributions, as well as flexible families like discrete scale mixtures of normals and polynomial expansions, which could form the basis for a proper nonparametric procedure. These results are important in practice because empirical researchers as well as financial market participants often want to go beyond the first two conditional moments, which implies that one cannot simply treat the shape parameters as if they were nuisance parameters.

We explain how to compute asymptotic standard errors of sequential estimators that take into account the sampling variability of the Gaussian PML estimators on which they are based. Further, we exploit the asymptotic variance expressions that we derive to assess the relative efficiency of sequential estimators, and obtain the optimal moment conditions that lead to sequential MM estimators which are as efficient as their joint ML counterparts. Moreover, our theoretical calculations indicate that the efficiency loss of sequential ML estimators is usually very small. From a practical point of view, we also provide simple analytical expressions for the asymptotic variances by exploiting a reparametrisation of the conditional mean and variance functions which covers most dynamic models. Obviously, our results also apply in univariate contexts as well as in static ones.

We then analyse the use of our sequential estimators in the calculation of commonly used risk management measures such as VaR, and recently proposed systemic risk measures such as CoVaR. Specifically, we provide analytical expressions for the asymptotic variances of those

measures. Perhaps not surprisingly, our results indicate that the standard errors are larger for CoVaR than for VaR, and that they increase as we lower the significance level. Our findings also confirm that the assumption of Gaussianity could be rather misleading even in situations where the actual DGP has moderate excess kurtosis. This is particularly true for the VaR figures at low significance levels, and especially for the CoVaR numbers. We also compare our sequential estimators to nonparametric estimators, both under correct specification of the parametric distribution, and also under misspecification. In this sense, our analytical and simulation results indicate that the use of sequential ML estimators of flexible parametric families of distributions offer substantial efficiency gains for those risk measures, while incurring in small biases.

As we mentioned in the introduction, the sequential estimation approach that we have studied could be equally applied models with non-spherical innovations, so it might be useful to derive the different expressions that we have obtained for general multivariate distributions. It might also be interesting to introduce dynamic features in higher-order moments. In this sense, at least two possibilities might be worth exploring: either time varying shape parameters, as in Jondeau and Rockinger (2003), or a regime switching process, following Guidolin and Timmermann (2007). These topics constitute interesting avenues for future research.

Appendix

A Proofs

Proposition 3

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

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

On this basis, the expression for $\mathcal{F}(\boldsymbol{\phi}_0)$ follows from the definitions of $\mathcal{B}(\boldsymbol{\phi}_0)$, $\mathcal{C}(\boldsymbol{\phi}_0)$ and $\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ in Propositions 1 and 2, together with the martingale difference nature of $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, and the fact that $E\{\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\mathbf{e}'_{rt}(\boldsymbol{\phi}) | \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\} = \mathbf{0}$. \square

Proposition 4

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

$$\mathbf{Z}_{dt}(\boldsymbol{\vartheta}) = \begin{Bmatrix} \vartheta_2^{-1/2} [\partial \boldsymbol{\mu}'_t(\boldsymbol{\vartheta}_1) / \partial \boldsymbol{\vartheta}_1] \boldsymbol{\Sigma}_t^{\circ-1/2'}(\boldsymbol{\vartheta}_1) \\ 0 \end{Bmatrix} \\ \frac{1}{2} \left\{ \begin{array}{l} \partial \text{vec}'[\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)] / \partial \boldsymbol{\vartheta}_1 \\ \frac{1}{2} \vartheta_2^{-1} \text{vec}'(\mathbf{I}_N) \end{array} \right\} [\boldsymbol{\Sigma}_t^{\circ-1/2'}(\boldsymbol{\vartheta}_1) \otimes \boldsymbol{\Sigma}_t^{\circ-1/2'}(\boldsymbol{\vartheta}_1)] = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\boldsymbol{\vartheta}_2 s}(\boldsymbol{\vartheta}) \end{bmatrix}. \quad (\text{A1})$$

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

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \\ s_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta}) \mathbf{e}_{lt}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) + \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}) \mathbf{e}_{st}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \\ \mathbf{Z}_{\boldsymbol{\vartheta}_2 s}(\boldsymbol{\vartheta}) \mathbf{e}_{st}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \end{bmatrix},$$

where $\mathbf{e}_{lt}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ are given in (5) and (6), respectively. Specifically,

$$s_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) = \frac{N}{2\vartheta_2} \left[\delta(\varsigma_t, \boldsymbol{\eta}) \frac{\varsigma_t}{N} - 1 \right]. \quad (\text{A2})$$

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ and $s_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ is given by

$$\begin{aligned} & E \left\{ \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}) \end{bmatrix} \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{Z}'_{\boldsymbol{\vartheta}_2 s}(\boldsymbol{\vartheta}) \end{bmatrix} \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} \\ &= \frac{\{2\mathcal{M}_{ss}(\boldsymbol{\eta}) + N[\mathcal{M}_{ss}(\boldsymbol{\eta}) - 1]\}}{2\vartheta_2} E \left\{ \frac{1}{2} \frac{\partial \text{vec}'[\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)]}{\partial \boldsymbol{\vartheta}_1} [\boldsymbol{\Sigma}_t^{\circ-1/2'}(\boldsymbol{\vartheta}_1) \otimes \boldsymbol{\Sigma}_t^{\circ-1/2'}(\boldsymbol{\vartheta}_1)] \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} \text{vec}(\mathbf{I}_N) \\ &= \frac{\{2\mathcal{M}_{ss}(\boldsymbol{\eta}) + N[\mathcal{M}_{ss}(\boldsymbol{\eta}) - 1]\}}{2\vartheta_2} \mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \text{vec}(\mathbf{I}_N) = \frac{\{2\mathcal{M}_{ss}(\boldsymbol{\eta}) + N[\mathcal{M}_{ss}(\boldsymbol{\eta}) - 1]\}}{2\vartheta_2} \mathbf{W}_{\boldsymbol{\vartheta}_1}(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \end{aligned}$$

with $\mathbf{Z}_{\vartheta_1 s}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) = E[\mathbf{Z}_{\vartheta_1 st}(\boldsymbol{\vartheta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$, as well as the law of iterated expectations, together with the results in Proposition 1. In this context, condition (11) implies that $\mathbf{W}_{\vartheta_1}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ will be $\mathbf{0}$, so that (18) reduces to

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = [0 \quad \cdots \quad 0 \quad N/(2\vartheta_2)]'.$$

This condition also implies that the unconditional covariance between $\mathbf{s}_{\vartheta_1 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ and $\mathbf{s}_{\boldsymbol{\eta} t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ will be $\mathbf{0}$ too, so that the information matrix will be block diagonal between $\boldsymbol{\vartheta}_1$ and $(\vartheta_2, \boldsymbol{\eta})$.

In turn, the unconditional variance of $s_{\vartheta_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ will be given by

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

while its covariance with $s_{\boldsymbol{\eta} t}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ will be

$$\mathbf{M}_{sr}(\boldsymbol{\eta}_0) \frac{N}{2\vartheta_2}.$$

Therefore, the asymptotic variance of the feasible ML estimator of $\boldsymbol{\eta}$ is

$$\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) = \left[\mathcal{M}_{rr}(\boldsymbol{\eta}_0) - \mathbf{M}'_{sr}(\boldsymbol{\eta}_0) \mathbf{M}_{sr}(\boldsymbol{\eta}_0) \frac{N}{\{2\mathbf{M}_{ss}(\boldsymbol{\eta}_0) + N[\mathbf{M}_{ss}(\boldsymbol{\eta}_0) - 1]\}} \right]^{-1}.$$

Using the partitioned inverse formula, we can finally obtain (13), with

$$\begin{aligned} \mathcal{I}^{\vartheta_2 \vartheta_2}(\boldsymbol{\phi}_0) &= [\mathcal{I}_{\vartheta_2 \vartheta_2}(\boldsymbol{\phi}_0) - \mathcal{I}_{\vartheta_2 \boldsymbol{\eta}}(\boldsymbol{\phi}_0) \mathcal{I}_{\boldsymbol{\eta} \boldsymbol{\eta}}^{-1}(\boldsymbol{\phi}_0) \mathcal{I}'_{\vartheta_2 \boldsymbol{\eta}}(\boldsymbol{\phi}_0)]^{-1} \\ &= \frac{1}{2\mathbf{M}_{ss}(\boldsymbol{\eta}_0) + N[\mathbf{M}_{ss}(\boldsymbol{\eta}_0) - 1 - \mathbf{M}_{sr}(\boldsymbol{\eta}_0) \mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0) \mathbf{M}'_{sr}(\boldsymbol{\eta}_0)]} \frac{4\vartheta_2^2}{N}. \end{aligned}$$

Analogous algebraic manipulations that exploit the block-triangularity of (A1) and the constancy of $\mathbf{Z}_{\vartheta_2 st}(\boldsymbol{\vartheta})$ show that $\mathcal{A}(\boldsymbol{\phi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_0)$, and therefore $\mathcal{C}(\boldsymbol{\phi}_0)$, will also be block diagonal between $\boldsymbol{\vartheta}_1$ and ϑ_2 when (11) holds. But since

$$s_{\vartheta_2 t}(\boldsymbol{\vartheta}, \mathbf{0}) = \frac{N}{2\vartheta_2} \left[\frac{\zeta_t(\boldsymbol{\vartheta})}{N} - 1 \right],$$

the expression for (12) follows. \square

Proposition 5

It is easy to see that under standard regularity conditions

$$\frac{1}{T} \sum_{t=1}^T \mathbf{n}_t^\perp(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0) = \frac{1}{T} \sum_{t=1}^T \mathbf{n}_t^\perp(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - \mathcal{N}_{\mathbf{n}^\perp} \sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1),$$

where

$$\mathcal{N}_{\mathbf{n}^\perp} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left(- \frac{\partial \mathbf{n}_t^\perp(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta}'} \middle| \phi_0 \right).$$

But since

$$\mathcal{N}_{\mathbf{n}^\perp} = \mathcal{N}_{\mathbf{n}} - \mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1} \mathcal{A} = \mathbf{0},$$

it immediately follows that

$$\lim_{T \rightarrow \infty} V \left(\frac{\sqrt{T}}{T} \sum_{t=1}^T \mathbf{n}_t^\perp(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0) \middle| \phi_0 \right) = \begin{pmatrix} I & -\mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathcal{G}_{\mathbf{n}} & \mathcal{D}_{\mathbf{n}} \\ \mathcal{D}'_{\mathbf{n}} & \mathcal{B} \end{pmatrix} \begin{pmatrix} I \\ -\mathcal{N}_{\mathbf{n}} \mathcal{A}^{-1} \end{pmatrix} = \boldsymbol{\varepsilon}_{\mathbf{n}}.$$

Given that

$$\frac{\partial \mathbf{n}_t^\perp(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}'} = \frac{\partial \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}'},$$

it follows that the efficient sequential GMM estimators based on $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$ and $\mathbf{n}_t^\perp(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$ will be asymptotically equivalent. \square

Proposition 6

Let us prove that

$$\mathbf{n}_t^\perp(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{n}_t^\perp(\boldsymbol{\vartheta}, \boldsymbol{\eta}) = \mathbf{n}_t^\circ(\boldsymbol{\vartheta}, \boldsymbol{\eta}) = \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) - \frac{N}{2} \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right]$$

by proving the first equality, followed by the last one, and finally the middle one.

The chain rule for first derivatives implies that

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \frac{\partial \mathbf{r}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\ \text{Cov}[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}) | \phi] &= \text{Cov}[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] \frac{\partial \mathbf{r}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{aligned}$$

and

$$\text{Cov}[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}) | \phi] = \frac{\partial \mathbf{r}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{Cov}[\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \mathbf{0}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

Therefore,

$$\begin{aligned} \mathbf{n}_t^\perp(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \text{Cov}[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}) | \phi] \text{Cov}^{-1}[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}) | \phi] \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}) \\ &= \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) - \text{Cov}[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] \text{Cov}^{-1}[\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \mathbf{0}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] \mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}, \mathbf{0}) = \mathbf{n}_t^\perp(\boldsymbol{\vartheta}, \boldsymbol{\eta}). \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbf{n}_t^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{\text{Cov}[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \delta(\varsigma_t, \boldsymbol{\eta}) \varsigma_t / N - 1 | \phi]}{\text{Cov}[\delta(\varsigma_t, \boldsymbol{\eta}) \varsigma_t / N - 1, \varsigma_t / N - 1 | \phi]} \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right] \\ &= \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{N}{2} \mathbb{k}_{\mathbf{n}}(\boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right] \end{aligned}$$

by definition of $\mathbb{k}_n(\boldsymbol{\eta})$ since

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

(see Fiorentini and Sentana (2010) for a proof). But since $\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})$ depends on $\boldsymbol{\theta}$ only through $\varsigma_t(\boldsymbol{\theta})$, it will coincide with $\mathbf{n}_t^\circ(\boldsymbol{\vartheta}, \boldsymbol{\eta})$.

In order to prove the middle equality, it is convenient to note that ellipticity implies that

$$E \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] = \mathbf{0}$$

and

$$E \left\{ \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \text{vec}' \left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) - \mathbf{I}_N \right] \middle| \mathbf{z}_t, I_{t-1}; \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} = E \left\{ \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right] \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} \text{vec}'(\mathbf{I}_N)$$

because the only random variable that enters $\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ is ς_t . Therefore

$$\begin{aligned} \text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] &= E \left\{ \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \left[\mathbf{e}'_{1t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \mathbf{Z}'_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta}) + \mathbf{e}'_{st}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \mathbf{Z}'_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}) \right] \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} \\ &= E \left\{ \mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right] \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right\} \mathbf{W}_{\boldsymbol{\vartheta}_1}(\boldsymbol{\vartheta}, \boldsymbol{\eta}). \end{aligned}$$

But since we saw in the proof of Proposition 4 that $\mathbf{W}_{\boldsymbol{\vartheta}_1}(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ is $\mathbf{0}$, it follows that

$$\text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] = \mathbf{0}.$$

Similarly,

$$\text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] = \frac{N}{2\vartheta_2} \text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \delta(\varsigma_t, \boldsymbol{\eta}) \frac{\varsigma_t(\boldsymbol{\vartheta})}{N} - 1 \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] = \frac{N}{2\vartheta_2} \mathbb{k}_n(\boldsymbol{\eta})$$

in view of (A2).

Further, since we also saw in the proof of Proposition 4 that $\mathcal{A} = \text{Cov} \left[\mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \mathbf{0}), \mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right]$ is block-diagonal with (p, p) -element equal to

$$\mathcal{A}_{\boldsymbol{\vartheta}_2 \boldsymbol{\vartheta}_2} = \text{Cov} \left[\mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \mathbf{0}), \mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] = \frac{N^2}{4\vartheta_2^2} E \left\{ \left[\delta(\varsigma_t, \boldsymbol{\eta}) \frac{\varsigma_t}{N} - 1 \right] \left(\frac{\varsigma_t}{N} - 1 \right) \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right\} = \frac{N}{2\vartheta_2^2},$$

it immediately follows that

$$\begin{aligned} &\text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] \text{Cov}^{-1} \left[\mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \mathbf{0}), \mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] \mathbf{s}_{\boldsymbol{\vartheta}_t}(\boldsymbol{\vartheta}, \mathbf{0}) \\ &= \frac{\text{Cov} \left[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right]}{\text{Cov} \left[\mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \mathbf{0}), \mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \boldsymbol{\eta}) \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right]} \mathbf{s}_{\boldsymbol{\vartheta}_2 t}(\boldsymbol{\vartheta}, \mathbf{0}) = \frac{N}{2} \mathbb{k}_n(\boldsymbol{\eta}) \left[\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right] \end{aligned}$$

regardless of the original model.

Finally, the *i.i.d.* nature of ς_t implies that asymptotic variance of the sample average of $\mathbf{n}_t^\circ(\boldsymbol{\vartheta}, \boldsymbol{\eta})$ will be

$$\begin{aligned} V[\mathbf{n}_t^\circ(\boldsymbol{\vartheta}, \boldsymbol{\eta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] &= V[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] - \frac{N}{2} \mathbb{k}_n(\boldsymbol{\eta}) \text{Cov}' \left[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \frac{\varsigma_t}{N} \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] \\ &\quad - \frac{N}{2} \text{Cov}' \left[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \frac{\varsigma_t}{N} \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right] \mathbb{k}_n'(\boldsymbol{\eta}) + \frac{N^2}{4} \mathbb{k}_n(\boldsymbol{\eta}) \mathbb{k}_n(\boldsymbol{\eta})' V \left(\frac{\varsigma_t}{N} - 1 \middle| \boldsymbol{\vartheta}, \boldsymbol{\eta} \right) \\ &= V[\mathbf{n}_t(\boldsymbol{\vartheta}, \boldsymbol{\eta}) | \boldsymbol{\vartheta}, \boldsymbol{\eta}] - \frac{N}{2} [\mathbb{k}_n(\boldsymbol{\eta}) \mathbb{k}_n'(\mathbf{0}) + \mathbb{k}_n(\boldsymbol{\eta}) \mathbb{k}_n'(\mathbf{0})] + \left(\frac{N}{2} + \frac{N(N+2)\kappa}{4} \right) \mathbb{k}_n(\boldsymbol{\eta}) \mathbb{k}_n(\boldsymbol{\eta})' \end{aligned}$$

where we have used the fact that

$$V \left(\frac{\varsigma_t}{N} \right) = \frac{(N+2)\kappa}{N} + \frac{2}{N},$$

which follows from the definition of κ in (2). \square

Proposition 7

In view of Proposition 6, we can easily define adjusted moments that are invariant to the sampling uncertainty surrounding $\tilde{\boldsymbol{\theta}}_T$. Specifically, for $m \geq 1$ we get

$$\begin{aligned} \ell_{mt}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{\text{cov}\{\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}), \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]_{\varsigma_t(\boldsymbol{\theta})/N-1}\}}{\text{cov}\{p_{1t}[\varsigma_t(\boldsymbol{\theta})], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]_{\varsigma_t(\boldsymbol{\theta})/N-1}\}} p_{1t}[\varsigma_t(\boldsymbol{\theta})], \\ p_{mt}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] &= p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] - \frac{\text{cov}\{p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]_{\varsigma_t(\boldsymbol{\theta})/N-1}\}}{\text{cov}\{p_{1t}[\varsigma_t(\boldsymbol{\theta})], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]_{\varsigma_t(\boldsymbol{\theta})/N-1}\}} p_{1t}[\varsigma_t(\boldsymbol{\theta})], \end{aligned}$$

which are such that

$$\ell_{1t}^\circ(\boldsymbol{\theta}) = p_{1t}^\circ[\varsigma_t(\boldsymbol{\theta})] = 0.$$

The bilinearity of the covariance operator applied to (B3) implies that

$$p_{mt}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = \ell_{mt}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) - \sum_{j=1}^{m-1} \frac{\text{cov}\{\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}}{V\{p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}} p_{jt}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}].$$

As a result, we can write $\{p_{2t}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \dots, p_{Mt}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}$ as a full-rank linear transformation of $[\ell_{2t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}), \dots, \ell_{Mt}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta})]$, which confirms the asymptotic equivalence in the case of two-step GMM procedures, and the numerical equivalence for single-step ones.

As a way of example,

$$p_{2t}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = \ell_{2t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{\text{cov}\{\ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{1t}[\varsigma_t(\boldsymbol{\theta})]\}}{V\{p_{1t}[\varsigma_t(\boldsymbol{\theta})]\}} p_{1t}^\circ[\varsigma_t(\boldsymbol{\theta})] = \ell_{2t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta})$$

and

$$\begin{aligned} p_{3t}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] &= \ell_{3t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{\text{cov}\{\ell_{3t}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{2t}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}}{V\{p_{2t}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}} p_{2t}^\circ[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] - \frac{\text{cov}\{\ell_{3t}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{1t}[\varsigma_t(\boldsymbol{\theta})]\}}{V\{p_{1t}[\varsigma_t(\boldsymbol{\theta})]\}} p_{1t}^\circ[\varsigma_t(\boldsymbol{\theta})] \\ &= \ell_{3t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}) - \frac{\text{cov}\{\ell_{3t}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{2t}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}}{V\{p_{2t}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}} \ell_{2t}^\circ(\boldsymbol{\theta}, \boldsymbol{\eta}). \end{aligned}$$

In addition, the expression for $V\{p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]|\boldsymbol{\phi}\}$ follows directly from the expression for the polynomials.

Similarly, it follows from the orthogonality of polynomials that

$$\mathbb{k}_{\mathbf{p}}(\mathbf{0}) = \text{Cov}\{p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \varsigma_t(\boldsymbol{\theta})/N|\boldsymbol{\phi}\} = \mathbf{0}.$$

Finally, in order to derive the expression for $\text{Cov}\{p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N|\boldsymbol{\phi}\}$ it is convenient to note that Lemma 1 in Fiorentini and Sentana (2010) implies that

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

which in turn implies that

$$\begin{aligned} & E \left\{ \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) \left[\delta(\varsigma_t, \boldsymbol{\eta}) \frac{\varsigma_t}{N} - 1 \right] \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right\} \\ &= -\frac{2}{N} E \left\{ \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) \left[1 + \varsigma_t \cdot \frac{\partial \ln h(\varsigma_t, \boldsymbol{\eta})}{\partial \varsigma} \right] \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right\} \\ &= -\frac{2}{N} E \left[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) \cdot \varsigma_t \cdot \frac{\partial \ln h(\varsigma_t, \boldsymbol{\eta})}{\partial \varsigma} \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right] \\ &= \frac{2}{N} E \left\{ \frac{\partial[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) \cdot \varsigma_t]}{\partial \varsigma} \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right\} = \frac{2}{N} E \left[\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \varsigma} \cdot \varsigma_t \middle| \boldsymbol{\theta}, \boldsymbol{\eta} \right]. \end{aligned}$$

This expression yields

$$\mathbb{k}_{\mathbf{p}}(\boldsymbol{\eta}) = \sum_{h=1}^m ha_h(\boldsymbol{\eta}) [1 + \tau_h(\boldsymbol{\eta})] \frac{2^{h+1}}{N} \prod_{j=1}^h (N/2 + j - 1)$$

in the case of orthogonal polynomials. □

Proposition 8

The first thing to note is that

$$\text{cov}[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0) \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})] = \mathbf{0},$$

which means that

$$E \left[\frac{\partial \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}$$

by virtue of the generalised information equality, which in turn implies that the asymptotic distribution of the sample average of $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})$ will be invariant to parameter uncertainty in $\boldsymbol{\theta}$ (see Bontemps and Meddahi (2011) for further discussion of this point).

Following Newey and Powell (1998), if $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is efficient then it will satisfy

$$V[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})] = -E \left[\frac{\partial \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right].$$

But

$$V [\mathbf{s}_{\eta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})] = \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0),$$

which coincides with

$$-E \left[\frac{\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right] = \text{cov} [\mathbf{s}_{\eta t}(\boldsymbol{\theta}, \boldsymbol{\eta}), \mathbf{s}_{\eta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\eta})].$$

In the case in which parametrisation (1) holds, the same steps used in the proof of Proposition 6 yield expression (17). \square

Proposition 9

The linearised influence functions corresponding to $\check{\eta}_T$ and $\hat{\eta}_T$ are

$$\ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathcal{N}'_{\ell}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

and

$$p_{2t}[\varsigma(\boldsymbol{\theta}_0), \boldsymbol{\eta}] - \mathcal{N}'_p(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

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

$$\begin{aligned} \text{cov}[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0), \ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)|\boldsymbol{\phi}_0] &= \mathbf{Z}_d(\boldsymbol{\phi}_0) \cdot E\{E[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0) \cdot \ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)|\varsigma_t; \boldsymbol{\phi}_0]|\boldsymbol{\phi}_0\} \\ &= \mathbf{W}_s(\boldsymbol{\phi}_0)E \left[\left(\frac{\varsigma_t}{N} - 1 \right) \ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \middle| \boldsymbol{\phi}_0 \right] \end{aligned}$$

and

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

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

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

and

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

On the other hand, since $p_{2t}[\varsigma(\boldsymbol{\theta}_0), \eta_0]$ is the residual from the least squares projection of $\ell_{2t}(\boldsymbol{\theta}_0, \eta_0)$ on $\varsigma_t/N - 1$, we can obtain the relevant expressions for $p_{2t}[\varsigma(\boldsymbol{\theta}_0), \eta_0]$ from those of $\ell_{2t}(\boldsymbol{\theta}_0, \eta_0)$ by using the fact that

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

and

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

□

Proposition 10

To prove that $\mathcal{F}(\phi_0) \leq \mathcal{E}_p(\phi_0)/\mathcal{H}^2(\phi_0)$ it is convenient to note that both sides of this inequality can be decomposed into a component that reflects the asymptotic variance of the estimators of η if $\boldsymbol{\theta}_0$ were known, plus a second component that reflects the sample variability in the PML estimator $\tilde{\boldsymbol{\theta}}_T$. With respect to the first component, it is clear that $\mathcal{I}_{\eta\eta}^{-1}(\phi_0) \leq \mathcal{G}_p(\phi_0)/\mathcal{H}^2(\phi_0)$. As for the second component, we must compare

$$\mathcal{I}'_{\boldsymbol{\theta}\eta}(\phi_0)\mathcal{C}(\phi_0)\mathcal{I}_{\boldsymbol{\theta}\eta}(\phi_0)/\mathcal{I}_{\eta\eta}^2(\phi_0) = \left[\frac{2(N+2)\nu^2}{(\nu-2)(N+\nu)(N+\nu+2)\mathcal{I}_{\eta\eta}(\phi_0)} \right]^2 \mathbf{W}'_s(\phi_0)\mathcal{C}(\phi_0)\mathbf{W}_s(\phi_0)$$

with

$$\mathcal{N}'_p(\phi_0)\mathcal{C}(\phi_0)\mathcal{N}_p(\phi_0)/\mathcal{H}^2(\phi_0) = \left[\frac{4(\nu-2)(\nu-4)}{N\nu^2(\nu-6)} \right]^2 \mathbf{W}'_s(\phi_0)\mathcal{C}(\phi_0)\mathbf{W}_s(\phi_0).$$

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

$$\mathcal{I}_{\eta\eta}(\phi_0) - \frac{(N+2)N\nu^4(\nu-6)}{2(\nu-2)^2(\nu-4)(N+\nu)(N+\nu+2)} \geq 0.$$

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

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

$$\begin{aligned} \mathcal{E}_\ell(\phi_0) &= [\mathcal{G}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{D}_\ell(\phi_0)] \\ &+ \{[\mathcal{N}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{A}(\phi_0)]'\mathcal{C}(\phi_0)[\mathcal{N}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{A}(\phi_0)]\}. \end{aligned}$$

In this set up, it is straightforward to prove that

$$[\mathcal{N}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{A}(\phi_0)] = \mathcal{N}_p(\phi_0)$$

if condition (19) holds. As for the first component, since $\mathcal{G}_p(\phi_0)$ is the residual variance in the regression of $\ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ on $\varsigma_t/N - 1$, while $\mathcal{G}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{D}_\ell(\phi_0)$ is the residual variance in the regression of $\ell_{2t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ on $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$, and the Gaussian pseudo-score can be written as $\mathbf{W}_s(\phi_0)(\varsigma_t/N - 1)$ plus an extra term that is orthogonal to ς_t , it is clear that

$$\mathcal{G}_p(\phi_0) \leq \mathcal{G}_\ell(\phi_0) - \mathcal{D}'_\ell(\phi_0)\mathcal{B}^{-1}(\phi_0)\mathcal{D}_\ell(\phi_0),$$

with equality if and only if $(\varsigma_t/N - 1)$ can be written as an exact linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$, as in (20). \square

Proposition 11

To obtain the variance of the elliptically symmetric score of $\boldsymbol{\eta}$ under misspecification, we can follow exactly the same steps as in the proof of Proposition 10 in Fiorentini and Sentana (2010) by exploiting that $E[\mathbf{e}_{rt}(\phi) | \boldsymbol{\varphi}_0] = \mathbf{0}$ holds at the pseudo-true parameter values $\phi_\infty = (\boldsymbol{\vartheta}_0, \boldsymbol{\eta}'_\infty)'$. Specifically, under standard regularity conditions

$$\frac{\sqrt{T}}{T} \sum_{t=1}^T \mathbf{e}_{rt}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_\infty) = \frac{\sqrt{T}}{T} \sum_{t=1}^T \mathbf{e}_{rt}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_\infty) + \frac{1}{T} \sum_{t=1}^T \frac{\partial \mathbf{e}_{rt}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_\infty)}{\partial \boldsymbol{\theta}'} \frac{\sqrt{T}}{T} (\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1),$$

where

$$\mathcal{O}_{sr}(\phi_\infty; \boldsymbol{\varphi}_0) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left(- \frac{\partial \mathbf{e}_{rt}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_\infty)}{\partial \boldsymbol{\theta}'} \Big| \boldsymbol{\varphi}_0 \right),$$

which under reparametrisation (1) reduces to

$$\mathcal{O}_{sr}(\phi_\infty; \boldsymbol{\varphi}_0) = \frac{N}{2\vartheta_2} \mathcal{M}_{sr}^O(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty; \boldsymbol{\varphi}_0),$$

where

$$\mathcal{M}_{sr}^O(\boldsymbol{\phi}; \boldsymbol{\varphi}) = E \left[\left\{ \delta[\varsigma_t(\boldsymbol{\vartheta}), \boldsymbol{\eta}] \cdot \left(\frac{\varsigma_t(\boldsymbol{\vartheta})}{N} - 1 \right) \right\} \mathbf{e}_{rt}(\boldsymbol{\phi}) \Big| \boldsymbol{\varphi} \right].$$

The remaining elements of the usual sandwich formula are

$$\mathcal{O}_{rr}(\phi_\infty; \boldsymbol{\varphi}_0) = \mathcal{M}_{rr}^O(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty; \boldsymbol{\varphi}_0), \quad \mathcal{M}_{rr}^O(\boldsymbol{\phi}; \boldsymbol{\varphi}) = V[\mathbf{e}_{rt}(\boldsymbol{\phi}) | \boldsymbol{\varphi}].$$

and

$$\mathcal{H}_{rr}(\phi_\infty; \boldsymbol{\varphi}_0) = \mathcal{M}_{rr}^H(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty; \boldsymbol{\varphi}_0), \quad \mathcal{M}_{rr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E \left[\frac{\partial \mathbf{e}_{rt}(\boldsymbol{\phi})}{\partial \boldsymbol{\eta}'} \Big| \boldsymbol{\varphi} \right].$$

Finally, the same steps used in the proof of Proposition 6 yield the expression for $\mathcal{E}_r(\phi_\infty; \boldsymbol{\varphi}_0)$. \square

B Orthogonal polynomials

The m^{th} orthogonal polynomial associated to a spherical distribution for $\boldsymbol{\varepsilon}_t^*$ is given by

$$p_{mt}^s[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = \sum_{h=0}^m a_h^s(\boldsymbol{\eta}) \varsigma_t^h(\boldsymbol{\theta}),$$

where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*'}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ and $\boldsymbol{\eta}$ are the shape parameters. The first two non-normalised polynomials are always $p_{0t}[\varsigma_t(\boldsymbol{\theta})] = 1$ and

$$p_{1t}[\varsigma_t(\boldsymbol{\theta})] = \frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1,$$

which do not depend on $\boldsymbol{\eta}$. Subsequent polynomials can be obtained by recursively regressing $\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta})$ in (15) on $p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ for $j = 0, 1, \dots, m-1$. Specifically,

$$p_{mt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = \ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}) - \sum_{j=1}^{m-1} \frac{\text{cov}\{\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}), p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}}{V\{p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\}} p_{jt}[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]. \quad (\text{B3})$$

As a result, the polynomials have zero mean and are orthogonal to each other by construction, although not orthonormal unless we standardise them by their respective standard deviations.

Next, we present the coefficients for the second and third orthogonal polynomials of the distributions we use in Sections 3, 4 and 5 to illustrate our results.

B.1 Orthogonal polynomials for the standardised Student t distribution

In this case the coefficients of the second order polynomial are

$$a_0^t(\eta) = \frac{N(N+2)(\nu-2)^3}{4(\nu-4)(\nu-6)},$$

$$a_1^t(\eta) = -\frac{(N+2)(\nu-2)}{2(\nu-6)}$$

and

$$a_2^t(\eta) = \frac{1}{4},$$

while the third order polynomial coefficients become

$$a_0^t(\eta) = -\frac{N(N+2)(N+4)(\nu-2)^2}{8(\nu-6)(\nu-8)(\nu-10)},$$

$$a_1^t(\eta) = \frac{(N+2)(N+4)(\nu-2)}{8(\nu-8)(\nu-10)},$$

$$a_2^t(\eta) = -\frac{(N+4)(\nu-2)}{8(\nu-10)},$$

and

$$a_3^t(\eta) = \frac{1}{24}.$$

B.2 Orthogonal polynomials for the standardised Kotz distribution

The coefficients of the second order polynomial are

$$a_0^k(\kappa) = \frac{N \{N + [(N + 2)\kappa + 2]\}}{2 [(N + 2)\kappa + 2]},$$

$$a_1^k(\kappa) = -\frac{N + (N + 2)\kappa + 2}{(N + 2)\kappa + 2},$$

and

$$a_2^k(\kappa) = \frac{1}{2 [(N + 2)\kappa + 2]}.$$

In turn, the coefficients of the third order one are

$$a_0^k(\kappa) = \frac{N \{N + [(N + 2)\kappa + 2]\} \{N + 2 [(N + 2)\kappa + 2]\}}{6 [(N + 2)\kappa + 2]^2},$$

$$a_1^k(\kappa) = -\frac{[N + (N + 2)\kappa + 2] \{N + 2 [(N + 2)\kappa + 2]\}}{2 [(N + 2)\kappa + 2]^2},$$

$$a_2^k(\kappa) = \frac{N + 2 [(N + 2)\kappa + 2]}{2 [(N + 2)\kappa + 2]^2},$$

and

$$a_3^k(\kappa) = -\frac{1}{6 [(N + 2)\kappa + 2]^2}.$$

B.3 Orthogonal polynomials for the standardised DSMN distribution

In this case the coefficients of the second order polynomial are

$$a_0^{ds}(\alpha, \varkappa) = \frac{N(N + 2)}{8} \frac{1}{[\alpha(1 - \varkappa) + \varkappa]^4} \{2(1 - \alpha)\varkappa^4 + (N + 4)(1 - \alpha)\alpha\varkappa^3$$

$$- 2(N + 2)(1 - \alpha)\alpha\varkappa^2 + (N + 4)(1 - \alpha)\alpha\varkappa + 2\alpha^2\},$$

$$a_1^{ds}(\alpha, \varkappa) = -\frac{N + 2}{8} \frac{1}{[\alpha(1 - \varkappa) + \varkappa]^3} \{2(1 - \alpha)(4 + N\alpha)\varkappa^3$$

$$- N(1 - \alpha)\alpha\varkappa^2 - N(1 - \alpha)\alpha\varkappa + \alpha [N(1 - \alpha) + 4]\},$$

and

$$a_2^{ds}(\alpha, \varkappa) = \frac{1}{8} \frac{1}{[\alpha(1 - \varkappa) + \varkappa]^2} \{(1 - \alpha)(2 + N\alpha)\varkappa^2 - 2N(1 - \alpha)\alpha\varkappa + [2 + N(1 - \alpha)]\alpha\}.$$

Similarly, the coefficients of the third order polynomial are

$$\begin{aligned}
a_0^{ds}(\alpha, \varkappa) = & -\frac{N(N+2)(N+4)}{192} \frac{1}{[(\alpha(1-\varkappa) + \varkappa)^6]} \\
& \times \{8(1-\alpha)^3 \varkappa^9 + (N+4)(N+6)(1-\alpha)^2 \alpha \varkappa^8 \\
& -4(N+2)(N+6)(1-\alpha)^2 \alpha \varkappa^7 + 6(N+4)^2(1-\alpha)^2 \alpha \varkappa^6 \\
& + (N+6)(1-\alpha)\alpha [(5N+24)\alpha - 4(N+4)] \varkappa^5 \\
& - (N+6)(1-\alpha)\alpha [24\alpha + N(5\alpha-1) - 8] \varkappa^4 \\
& + 6(N+4)^2(1-\alpha)\alpha^2 \varkappa^3 - 4(N+2)(N+6)(1-\alpha)\alpha^2 \varkappa^2 \\
& + (N+4)(N+6)(1-\alpha)\alpha^2 \varkappa + 8\alpha^3\},
\end{aligned}$$

$$\begin{aligned}
a_1^{ds}(\alpha, \varkappa) = & \frac{(N+2)(N+4)}{192} \frac{1}{[(\alpha(1-\varkappa) + \varkappa)^5]} \\
& \times \{(1-\alpha)^2 [N(N+10)\alpha + 24] \varkappa^8 - 2N(N+6)(1-\alpha)^2 \alpha \varkappa^7 \\
& - 2N(N+8)(1-\alpha)^2 \alpha \varkappa^6 \\
& + 2(1-\alpha)\alpha [4N(N+7) - N(3N+14)\alpha + 48] \varkappa^5 \\
& - (N+6)(7N+24)(1-\alpha)\alpha \varkappa^4 \\
& + 2(1-\alpha)\alpha [N(3\alpha N + N + 14(\alpha+1)) + 48] \varkappa^3 \\
& - 2N(N+8)(1-\alpha)\alpha^2 \varkappa^2 - 2N(N+6)(1-\alpha)\alpha^2 \varkappa \\
& + [24 - N(N+10)(\alpha-1)] \alpha^2\},
\end{aligned}$$

$$\begin{aligned}
a_2^{ds}(\alpha, \varkappa) = & -\frac{N+4}{192} \frac{1}{[(\alpha(1-\varkappa) + \varkappa)^4]} \\
& \times \{2(1-\alpha)^2 [N(N+8)\alpha + 12] \varkappa^7 - N(7N+38)(1-\alpha)^2 \alpha \varkappa^6 \\
& + (1-\alpha)\alpha [8(N+2)(N+3) - N(7N+26)\alpha] \varkappa^5 \\
& - 2(N+2)(N+6)(1-\alpha)\alpha \varkappa^4 - 2(N+2)(N+6)(1-\alpha)\alpha \varkappa^3 \\
& + (1-\alpha)\alpha [N(7\alpha N + N + 26\alpha + 14) + 48] \varkappa^2 \\
& - N(7N+38)(1-\alpha)\alpha^2 \varkappa + 2[12 + N(N+8)(1-\alpha)] \alpha^2\}
\end{aligned}$$

and

$$\begin{aligned}
a_3^{ds}(\alpha, \varkappa) &= \frac{1}{192} \frac{1}{[(\alpha(1-\varkappa) + \varkappa)^3]} \\
&\times \left\{ (1/\alpha)^2 [N(N+6)\alpha + 8] \varkappa^6 - 4N(N+4)(1-\alpha)^2 \alpha \varkappa^5 \right. \\
&\quad + (1-\alpha)\alpha [6(N+2)^2 - N(5N+14)\alpha] \varkappa^4 \\
&\quad - 4(N+2)(N+4)(1-\alpha)\alpha \varkappa^3 \\
&\quad + (1-\alpha)\alpha [N(5\alpha N + N + 14\alpha + 10) + 24] \varkappa^2 \\
&\quad \left. - 4N(N+4)(1-\alpha)\alpha^2 \varkappa + [8 - N(N+6)(\alpha-1)] \alpha^2 \right\}.
\end{aligned}$$

B.4 Orthogonal polynomials for the standardised 3^{rd} -order PE distribution

The coefficients of the second order polynomial are

$$\begin{aligned}
a_0^{pe}(c_2, c_3) &= \frac{1}{4} (-8c_2^2 + 2N(N+8)c_2 + N(N(N+2) - 12c_3)), \\
a_1^{pe}(c_2, c_3) &= -\frac{1}{2} [N(N+2) + (N+6)c_2 - 3c_3],
\end{aligned}$$

and

$$a_2^{pe}(c_2, c_3) = \frac{1}{4} (N + 2c_2).$$

In turn, the coefficients of the third order one are

$$\begin{aligned}
a_0^{pe}(c_2, c_3) &= -\frac{1}{24} \left\{ (N+2)^2(N+4)N^3 - 4(N^3 + 32N + 192)c_2^3 \right. \\
&\quad + 6c_2^2 [N(N+4)(N(N+2) + 48) + 4((N-4)N + 48)c_3] \\
&\quad - 12c_3 [(N+2)(7N+48)N^2 + 3(N(N+6) + 24)c_3N - 72c_3^2] \\
&\quad - 6c_2N^2(N+2)(N+4)(N+12) \\
&\quad \left. + 24c_2c_3 [N(N(N+14) + 120) + 6(N-12)c_3] \right\}, \\
a_1^{pe}(c_2, c_3) &= \frac{1}{8} \left\{ N^2(N+4)(N+2)^2 - 12[3(N(N+10) + 32) + 4c_2]c_3^2 \right. \\
&\quad + 2(N+4)c_2 [N(N+2)(3N+32) + c_2(N(3N+14) - 2Nc_2 + 96)] \\
&\quad \left. - 8[3N(N+2)(3N+20) + 2c_2(N(N+14) - Nc_2 + 96)]c_3 \right\}, \\
a_2^{pe}(c_2, c_3) &= \frac{1}{8} \left\{ (N+2)(N+4)N^2 - 36(N+8)c_3^2 \right. \\
&\quad + 2(N+4)c_2 [N(3N+26) + (3N-2c_2+8)c_2] \\
&\quad \left. - 4[15N(N+6) + 2(N-c_2+12)c_2]c_3 \right\},
\end{aligned}$$

and

$$a_3^{pe}(c_2, c_3) = \frac{1}{24} [N^2(N+2) - 4c_2^3 + 6Nc_2^2 + 6N(N+6)c_2 - 12c_3(4N+3c_3)].$$

C Auxiliary results

C.1 Positivity of Laguerre expansions

To identify the region in the parameter space for which $P_J(\varsigma) = 1 + \sum_{j=2}^J c_j \cdot p_j(\varsigma, N) \geq 0$ it is convenient to reparametrise $P_J(\varsigma)$ as $\hat{P}_J(c_2, c_3, \mathbf{t})$, with $\mathbf{t} = (\varsigma, c_4, \dots, c_J)$. For each value of $\mathbf{t} \in \mathbb{R}^{J-2}$, the equation $\hat{P}_J(c_2, c_3, \mathbf{t}) = 0$ defines a straight line in the \mathbf{t} -hyperplane. To determine the set of $\boldsymbol{\eta}$'s as a function of \mathbf{t} such that $\hat{P}_J(c_2, c_3, \mathbf{t})$ remains zero for small variations of \mathbf{t} , we should also impose $\partial \hat{P}_J(c_2, c_3, \mathbf{t}) / \partial \mathbf{t} = 0$. Finally, once this bound is found, we need to determine the subregion in which $P_J(\varsigma) \geq 0$ for $\varsigma \geq 0$.

C.1.1 Second order expansion

In this simple case the positivity region corresponds to those values of c_2 for which the polynomial $1 + c_2 \cdot p_2(t, N)$ is positive. Since the vertex of this quadratic function occurs at $t = N + 2 > 0$, positivity requires that its roots are either complex or double, which holds for $0 \leq c_2 \leq N$.

C.1.2 Third order expansion

For a given ς , the 3^{rd} order polynomial frontier that guarantees positivity must satisfy the following two equations in two unknowns

$$\begin{cases} 1 + c_2 \cdot p_2(t, N) + c_3 \cdot p_3(t, N) = 0 \\ c_2 \cdot \partial p_2(t, N) / \partial t - c_3 \cdot \partial p_3(t, N) / \partial t = 0 \end{cases}$$

whose solution is

$$c_2(t) = \frac{8 + 6N + N^2 - 8t - 2Nt + t^2}{8A(N, t)} \quad \text{and} \quad c_3(t) = \frac{N + 2 - t}{2A(N, t)}$$

with

$$A(N, t) = \frac{N^3t + Nt^3 - 5N^2}{24} + \frac{t^3 - N^3}{12} - \frac{N^4 + t^4}{96} + \frac{Nt - 2N}{3} + \frac{N^2t - Nt^2 - t^2}{4} - \frac{N^2t^2}{16}.$$

The solid (dashed) black line in Figure 1 represents the frontier defined by positive (negative) values of ς . Notice that if we imposed the above conditions for all $\varsigma \in \mathbb{R}$, then $c_3 = 0$ and

$0 \leq c_2 \leq N$. Such a frontier, however, is overly restrictive because it does not take into account the non-negativity of ς . In this sense, the red line represents the tangent of $P_3(\varsigma)$ at $\varsigma = 0$ while the blue line is the tangent of $P_3(\varsigma)$ when $\varsigma \rightarrow +\infty$. The grey area, therefore, defines the admissible set in the (c_2, c_3) space. Focusing on $\varsigma \in \mathbb{R}_+$ only allows for a larger range of (c_2, c_3) with $c_3 < 0$, which is given by the difference between the dashed black line and the blue one.

C.2 Higher order moments

The higher order moment parameter of spherical random variables defined in (3) for the four distributions that we use to illustrate our results are:

(a) Student t distribution with $\nu = 1/\eta$ degrees of freedom:

$$1 + \tau_m^t(\eta) = (1 - 2\eta)^{m-1} \prod_{j=2}^m \frac{1}{(1 - 2j\eta)} \quad \text{when } \eta < (2m)^{-1}.$$

(b) Kotz distribution with excess kurtosis κ :

$$1 + \tau_m^k(\kappa, N) = \left(\frac{(N+2)\kappa + 2}{2} \right)^m \prod_{j=1}^m \frac{N/[(N+2)\kappa + 2] + j - 1}{N/2 + j - 1}.$$

(c) Discrete scale mixture of normals distribution with mixing probability α and variance ratio \varkappa :

$$1 + \tau_m^{ds}(\alpha, \varkappa) = \frac{\alpha + (1 - \alpha)\varkappa^m}{[\alpha + (1 - \alpha)\varkappa]^m}.$$

(d) 3rd-order polynomial expansion distribution with parameters c_2 and c_3 :

$$1 + \tau_m^{pe}(\alpha, \varkappa) = 1 + \frac{2m(m-1)}{N(N+2)}c_2 - \frac{4m[2 + m(m-3)]}{N(N+2)(N+4)}c_3.$$

Derivation of the results:

(a) If ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν and variance 2ν , with ζ_t and ξ_t mutually independent, then the uncentred moments of integer order r of $(\nu/N) \times (\zeta_t/\xi_t)$ are given by

$$E \left[\left(\frac{\zeta_t/N}{\xi_t/\nu} \right)^r \right] = \left(\frac{\nu}{N} \right)^r \frac{r-1+N/2}{-1+\nu/2} \frac{r-2+N/2}{-2+\nu/2} \times \cdots \times \frac{1+N/2}{-(r-1)+\nu/2} \frac{N/2}{-r+\nu/2}$$

(Mood, Graybill and Boes, 1974). Given that $\varsigma_t = (\nu - 2)\zeta_t/\xi_t$, it is straightforward to see that

$$E \left\{ \left[(\nu - 2) \frac{\zeta_t}{\xi_t} \right]^m \right\} = \frac{N}{2} \left[\frac{2(\nu - 2)}{\nu} \right]^{m-1} \prod_{j=2}^m \frac{(N/2 + j - 1)\nu}{\nu - 2j}$$

from where the result follows directly.

(b) We can exploit the fact that if ε_t^* is distributed as a standardised Kotz, then ς_t is a gamma random variable with mean N and variance $N[(N+2)\kappa+2]$, whose higher order moments can be obtained from Lemma 1 in Amengual and Sentana (2011) with $b=0$.

(c) When ε_t^* is distributed as a DSMN, ς_t is a two-component scale mixture of $\chi_N^2 s$, so that conditioning on the mixing variate s ,

$$E[\varsigma_t^m | s = 1] = \left[\frac{1}{\alpha + (1-\alpha)\varkappa} \right]^m E(\zeta_t^m) \text{ and } E[\varsigma_t^m | s = 0] = \left[\frac{\varkappa}{\alpha + (1-\alpha)\varkappa} \right]^m E(\zeta_t^m)$$

where ζ_t is a χ_N^2 variate. Then, the required expression follows directly from the law of iterated expectations.

(d) Since $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t) | \mathbf{0}] = 0$ for $m < j$, we only need to compute $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t) | \mathbf{0}]$ for $m \geq j$, which can be written in terms of the higher order moments of the Gaussian distribution. For the 2^{nd} -order Laguerre polynomial we have

$$\begin{aligned} E[\varsigma_t^m p_{N/2-1,2}(\varsigma_t) | \mathbf{0}] &= \frac{1}{2} E[\varsigma_t^m | \mathbf{0}] - \frac{1}{N} E[\varsigma_t^{m+1} | \mathbf{0}] + \frac{1}{2N(N+2)} E[\varsigma_t^{m+2} | \mathbf{0}] \\ &= \left[\frac{1}{2} - \frac{2(N/2 + m + 1)}{N} + \frac{4(N/2 + m + 1)(N/2 + m + 2)}{2N(N+2)} \right] E[\varsigma_t^m | \mathbf{0}] \\ &= \frac{2m(m-1)}{N(N+2)} E[\varsigma_t^m | \mathbf{0}]. \end{aligned}$$

The same procedure applied to the 3^{rd} -order Laguerre polynomial yields the required result.

C.3 Moment generating functions

Not surprisingly, the moment generating function of a spherical random variable ε_t^* depends only on ς . Although it cannot be defined for the Student t distribution, it takes the following forms for the remaining distributions that we consider:

(a) Kotz distribution with excess kurtosis κ :

$$\Upsilon_k(t|\kappa) \equiv E[e^{t\varepsilon_t^*} | \kappa] = \{1 - [(N+2)\kappa+2]t\}^{-N/[(N+2)\kappa+2]}.$$

(b) Discrete scale mixture of normals distribution with mixing probability α and variance ratio \varkappa :

$$\Upsilon_{ds}(t|\alpha, \varkappa) \equiv E[e^{t\varepsilon_t^*} | (\alpha, \varkappa)'] = \alpha \left[1 - \frac{2t}{\alpha + (1-\alpha)\varkappa} \right]^{-N/2} + (1-\alpha) \left[1 - \frac{2\varkappa t}{\alpha + (1-\alpha)\varkappa} \right]^{-N/2}.$$

(c) 3^{rd} -order polynomial expansion with parameters c_2 and c_3 :

$$\Upsilon_{pe}^{J=3}(t|c_2, c_3) \equiv E[e^{t\varepsilon_t^*} | (c_2, c_3)'] = (1-2t)^{-N/2} \left[1 + \frac{2t^2}{(1-2t)^2} c_2 - \frac{4t^3}{(1-2t)^3} c_3 \right].$$

Derivation of the results:

(a) The result follows directly from the moment generating function of the Gamma distribution.

(b) Since ς_t is a two-component scale mixture of $\chi_N^2 s$, we can compute $E[e^{t\varsigma_t}|\alpha, \varkappa, s]$ for $s = 1$ and $s = 0$ by exploiting the fact that the relevant conditional distributions are Gamma with shape parameter $N/2$ and scale parameters

$$\frac{2}{\alpha + (1 - \alpha)\varkappa} \quad \text{and} \quad \frac{2\varkappa}{\alpha + (1 - \alpha)\varkappa}$$

respectively. Finally, the law of iterated expectation yields the desired result.

(c) The moment generating function of the polynomial expansion distribution can be easily obtained by applying Lemma 1 in Amengual and Sentana (2011). For the 2^{nd} -order Laguerre polynomial we have

$$\begin{aligned} E[e^{t\varsigma_t} p_{N/2-1,2}(\varsigma_t)|\mathbf{0}] &= \frac{1}{2}E[e^{t\varsigma_t}|\mathbf{0}] - \frac{1}{N}E[\varsigma_t e^{t\varsigma_t}|\mathbf{0}] + \frac{1}{2N(N+2)}E[\varsigma_t^2 e^{t\varsigma_t}|\mathbf{0}] \\ &= \frac{1}{2} \left(\frac{1}{1-2t} \right)^{N/2} - \left(\frac{1}{1-2t} \right)^{N/2+1} + \frac{1}{2} \left(\frac{1}{1-2t} \right)^{N/2+2} \\ &= (1-2t)^{-N/2} \left[\frac{(1-2t)^2 - 2(1-2t) + 1}{2(1-2t)^2} \right] \\ &= (1-2t)^{-N/2} \frac{2t^2}{(1-2t)^2}. \end{aligned}$$

The same procedure applied to the 3^{rd} -order Laguerre polynomial yields the required result.

C.4 Marginal and conditional distributions required for VaR and CoVaR calculations

Theorem 2.6 in Fang, Kotz and Ng (1990) characterises the marginal distribution of a partition of $\boldsymbol{\varepsilon}_t^*$ into n components. In particular, if we split $\boldsymbol{\varepsilon}_t^*$ into its first n elements, $\boldsymbol{\varepsilon}_{1t}^*$, and the remaining $N - n$ ones, $\boldsymbol{\varepsilon}_{2t}^*$ say, this theorem implies that

$$\begin{bmatrix} \boldsymbol{\varepsilon}_{1t}^* \\ \boldsymbol{\varepsilon}_{2t}^* \end{bmatrix} = \begin{bmatrix} e_t d_t \mathbf{u}_{1t} \\ e_t (1 - d_t) \mathbf{u}_{2t} \end{bmatrix},$$

where e_t is the generating variate, $d_t \sim \text{Beta}[n/2, (N-n)/2]$ and \mathbf{u}_{1t} and \mathbf{u}_{2t} are two independent vectors which are uniformly distributed on the unit sphere surface in \mathbb{R}^n and \mathbb{R}^{N-n} , respectively.

C.4.1 Marginal densities and CDFs of $z_{1t} = [\varepsilon_{1t}^*(\boldsymbol{\theta})]^2$

In the particular case of univariate marginals, it is easy to obtain the marginal probability density function of ε_{it}^{*2} (see Mood, Graybill, and Boes, 1974) by computing

$$h_1^s(z, \boldsymbol{\eta}) = \frac{\sqrt{\pi}\Gamma(N/2 - 1/2)}{\Gamma(N/2)} \int_0^\infty h^s\left(\frac{z}{y}, \boldsymbol{\eta}, N\right) \times \frac{(1-y)^{\frac{N}{2}-\frac{3}{2}}}{\sqrt{y}} dy.$$

From here, we can easily obtain $f_1^s(\varepsilon, \boldsymbol{\eta})$ using the change of variable formula as $h_1^s(\varepsilon^2, \boldsymbol{\eta}) \cdot |z|$.

Student t For the N -variate standardised Student t distribution with $\nu = 1/\eta$ degrees of freedom, the univariate marginal probability density function of $z = \varepsilon^2$ is

$$h_1^t(z, \eta) = \frac{\Gamma[1/2(1 + \eta^{-1})]}{\sqrt{\pi}\Gamma[(2\eta)^{-1}]} \frac{1}{\sqrt{z(z + \eta^{-1} - 2)}} \left[\frac{1 - 2\eta}{1 - 2\eta + \eta z} \right]^{\frac{1}{2\eta}},$$

while its cumulative distribution function is

$$H_1^t(z, \eta) = \frac{\Gamma[1/2(1 + \eta^{-1})]}{\sqrt{\pi}\Gamma[(2\eta)^{-1}]} \cdot i \cdot \text{Beta}\left(-\frac{\eta z}{1 - 2\eta}, \frac{1}{2}, \frac{\eta - 1}{2\eta}\right),$$

where $i = \sqrt{-1}$ and $\text{Beta}(z, a, b)$ is the incomplete beta function, defined by

$$\text{Beta}(z, a, b) = \int_0^z u^{a-1}(1-u)^{b-1} du.$$

Discrete scale mixture of normals For the N -variate standardised DSMN distribution with mixing probability α and variance ratio \varkappa :

$$h_1^{ds}(z, \alpha, \varkappa) = \frac{1}{\sqrt{z}\sqrt{2\pi}} \left\{ \alpha \sqrt{\frac{1}{\varpi}} \exp\left(-\frac{1}{2\varpi}z\right) + \left(\frac{1-\alpha}{\varkappa}\right) \sqrt{\frac{1}{\varpi\varkappa}} \exp\left(-\frac{1}{2\varpi\varkappa}z\right) \right\},$$

where $\varpi = [\alpha + (1 - \alpha)\varkappa]^{-1}$, while its cumulative distribution function is

$$H_1^{ds}(z, \alpha, \varkappa) = (1 - \alpha) \text{erf}\left(\frac{\sqrt{z}}{\sqrt{2\varpi}}\right) + \alpha \text{erf}\left(\frac{\sqrt{z}}{\sqrt{2\varpi\varkappa}}\right),$$

where $\text{erf}(x)$ is the standard ‘‘error function’’ defined by $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$.

3rd-order polynomial expansion For the N -variate standardised 3rd-order PE with parameters c_2 and c_3

$$h_1^{pe}(z, c_2, c_3) = \left\{ 1 + \frac{[z(z - 6) + 3]}{2N(N + 2)} c_2 - \frac{[z(45 + z(z - 15)) - 15]}{2N(N + 2)(N + 4)} c_3 \right\} \frac{1}{\sqrt{z}\sqrt{2\pi}} \exp\left(-\frac{z}{2}\right),$$

while its cumulative distribution function is

$$H_1^{pe}(z, c_2, c_3) = \frac{[15 + (z - 10)z] c_3 - (N + 4)(z - 3)c_2}{N(N + 2)(N + 4)\sqrt{2\pi}} \sqrt{z} \exp\left(-\frac{z}{2}\right) + \text{erf}\left(\frac{\sqrt{z}}{\sqrt{2}}\right).$$

C.4.2 Cumulative density functions of conditionals $(\varepsilon_{1t}^*(\boldsymbol{\theta})|\varepsilon_{2t}^*(\boldsymbol{\theta}))$

Using again Theorem 2.6 in Fang, Kotz and Ng (1990) we can obtain the marginal bivariate distribution $f_{1,2}^s(\varepsilon_1, \varepsilon_2, \boldsymbol{\eta})$, which together with $f_1^s(\varepsilon, \boldsymbol{\eta}) = h_1^s(\varepsilon^2, \boldsymbol{\eta}) \cdot |z|$, allow us to obtain the conditional pdfs. In this way,

Student t

$$F_{1|2}^t(\varepsilon_1, \varepsilon_2, \boldsymbol{\eta}) = \frac{1}{2} \left\{ 1 + \frac{\Gamma[1 + (2\eta)^{-1}]}{\sqrt{\pi}\Gamma[(1 + \eta)(2\eta)^{-1}]} \text{Beta} \left(-\frac{\eta\varepsilon_1^2}{1 + (\varepsilon_2^2 - 2)\eta}, \frac{1}{2}, -\frac{1}{2\eta} \right) \right\}.$$

Discrete scale mixture of normals

$$\begin{aligned} F_{1|2}^{ds}(\varepsilon_1, \varepsilon_2, \alpha, \varkappa) &= \left\{ (1 - \alpha) \exp \left(\frac{2\alpha + \varkappa}{2} \varepsilon_2^2 \right) \sqrt{\frac{1}{\varpi\varkappa}} + \alpha \exp \left(\frac{\alpha + \varkappa + \alpha\varkappa^2}{2\varkappa} \varepsilon_2^2 \right) \sqrt{\frac{1}{\varpi}} \right\}^{-1} \\ &\quad \frac{1}{2\sqrt{\varpi\varkappa}} \left\{ (1 - \alpha) \exp \left(\frac{2\alpha + \varkappa}{2} \varepsilon_2^2 \right) \left[1 + \text{erf} \left(\sqrt{\frac{1}{2\varpi\varkappa}} \varepsilon_1 \right) \right] \right. \\ &\quad \left. + \alpha \exp \left(\frac{\alpha + \varkappa + \alpha\varkappa^2}{2\varkappa} \varepsilon_2^2 \right) \sqrt{\varkappa} \left[1 + \text{erf} \left(\sqrt{\frac{1}{2\varpi}} \varepsilon_1 \right) \right] \right\}, \end{aligned}$$

where $\varpi = [\alpha + (1 - \alpha)\varkappa]^{-1}$.

3rd-order polynomial expansion

$$\begin{aligned} F_{1|2}^{pe}(\varepsilon_1, \varepsilon_2, c_2, c_3, N) &= \frac{1}{2} + \frac{1}{2\sqrt{\pi}} \left[2 + \frac{\varepsilon_2^4 - 6\varepsilon_2^2 + 3}{N(N+2)} c_2 - \frac{\varepsilon_2^6 - 15\varepsilon_2^4 + 45\varepsilon_2^2 - 15}{N(N+2)(N+4)} c_3 \right]^{-1} \\ &\quad \times \exp \left(-\frac{\varepsilon_1^2}{2} \right) \left\{ \exp \left(-\frac{\varepsilon_1^2}{2} \right) \sqrt{\pi} \text{erf} \left(\frac{\varepsilon_1}{\sqrt{2}} \right) \right. \\ &\quad \times \left(2 + \frac{\varepsilon_2^4 - 6\varepsilon_2^2 + 3}{N(N+2)} c_2 - \frac{\varepsilon_2^6 - 15\varepsilon_2^4 + 45\varepsilon_2^2 - 15}{N(N+2)(N+4)} c_3 \right) \\ &\quad \left. + \sqrt{2}\varepsilon_1 \left(\frac{(\varepsilon_1^4 - 13\varepsilon_1^2 + 3\varepsilon_2^4 + 3(\varepsilon_1^2 - 9)\varepsilon_2^2 + 33)}{N(N+2)(N+4)} c_3 - \frac{\varepsilon_1^2 + 2\varepsilon_2^2 - 5}{N(N+2)} c_2 \right) \right\}. \end{aligned}$$

C.5 Standard errors for parametric VaR and CoVaR

Given that $q_1(\lambda, \boldsymbol{\eta})$ satisfies

$$\lambda = F[q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}] = \int_0^{q_1(\lambda, \boldsymbol{\eta})} f_1(\varepsilon_{1t}^*; \boldsymbol{\eta}) d\varepsilon_{1t}^*,$$

if we differentiate this expression with respect to $\boldsymbol{\eta}$ we obtain

$$0 = f_1[q_1(\lambda, \boldsymbol{\eta}); \boldsymbol{\eta}] \frac{\partial q_1(\lambda, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} + \int_0^{q_1(\lambda, \boldsymbol{\eta})} \frac{\partial f_1(\varepsilon_{1t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{1t}^*,$$

whence

$$\frac{\partial q_1(\lambda, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = -\frac{1}{f_1[q_1(\lambda, \boldsymbol{\eta}); \boldsymbol{\eta}]} \int_0^{q_1(\lambda, \boldsymbol{\eta})} \frac{\partial f_1(\varepsilon_{1t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{1t}^*.$$

To relate this expression to the asymptotic variances of the non-parametric quantile estimators, it is convenient to write

$$\begin{aligned} \int_0^{q_1(\lambda, \boldsymbol{\eta})} \frac{\partial f_1(\varepsilon_{1t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{1t}^* &= \int_0^{q_1(\lambda, \boldsymbol{\eta})} \frac{\partial \ln f_1(\varepsilon_{1t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} f_1(\varepsilon_{1t}^*; \boldsymbol{\eta}) d\varepsilon_{1t}^* \\ &= \Pr[\varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta})] E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta})], \end{aligned}$$

where, importantly, the distribution used to compute the foregoing expectation is the same as the distribution used for estimation purposes. Hence, we will have that

$$V[q_1(\lambda, \hat{\boldsymbol{\eta}}_T)] = \frac{\lambda^2}{f_1^2[q_1(\lambda, \boldsymbol{\eta}); \boldsymbol{\eta}_0]} E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0] V[\hat{\boldsymbol{\eta}} | \boldsymbol{\eta}_0] E[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0].$$

Further, given that

$$0 = E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \boldsymbol{\eta}_0] = \lambda E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0] + (1 - \lambda) E[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \geq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0],$$

we can finally write

$$V[q_1(\lambda, \hat{\boldsymbol{\eta}}_T)] = \frac{\lambda(1 - \lambda)}{f_1^2[q_1(\lambda, \boldsymbol{\eta}); \boldsymbol{\eta}_0]} E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \leq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0] V[\hat{\boldsymbol{\eta}} | \boldsymbol{\eta}_0] E[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) | \varepsilon_{1t}^* \geq q_1(\lambda, \boldsymbol{\eta}), \boldsymbol{\eta}_0].$$

Let $f_{1,2}$ denote the joint bivariate distribution of ε_{1t}^* and ε_{2t}^* . By definition, we know that $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$ satisfies

$$\begin{aligned} \lambda_{2|1} &= \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} f_1(\varepsilon_{1t}^*; \boldsymbol{\eta}) \left(\int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} f_{2|1}(\varepsilon_{2t}^*, \varepsilon_{1t}^*; \boldsymbol{\eta}) d\varepsilon_{2t}^* \right) d\varepsilon_{1t}^* \\ &= \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} \int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} f_{1,2}(\varepsilon_{1t}^*, \varepsilon_{2t}^*; \boldsymbol{\eta}) d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\ &= \int_{-\infty}^0 \int_{-\infty}^0 f_{1,2}(\varepsilon_{1t}^* + q_1(\lambda_1, \boldsymbol{\eta}), \varepsilon_{2t}^* + q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}); \boldsymbol{\eta}) d\varepsilon_{2t}^* d\varepsilon_{1t}^*, \end{aligned}$$

where we have achieved constant limits of integration in the last expression by means of the change of variable

$$u(\varepsilon_{1t}^*, \varepsilon_{2t}^*) = \varepsilon_{1t}^* + q_1(\lambda_1, \boldsymbol{\eta}), \text{ and } v(\varepsilon_{1t}^*, \varepsilon_{2t}^*) = \varepsilon_{2t}^* + q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}),$$

whose Jacobian is 1. Differentiating the previous expression with respect to $\boldsymbol{\eta}$ yields

$$\begin{aligned} 0 &= \int_{-\infty}^0 \int_{-\infty}^0 \frac{f_{1,2}(\varepsilon_{1t}^* + q_1(\lambda_1, \boldsymbol{\eta}), \varepsilon_{2t}^* + q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}); \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\ &\quad + \frac{\partial q_1(\lambda_1, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \int_{-\infty}^0 \int_{-\infty}^0 \frac{\partial f_{1,2}(\varepsilon_{1t}^* + q_1(\lambda_1, \boldsymbol{\eta}), \varepsilon_{2t}^* + q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}); \boldsymbol{\eta})}{\partial \varepsilon_{1t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\ &\quad + \frac{\partial q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \int_{-\infty}^0 \int_{-\infty}^0 \frac{\partial f_{1,2}(\varepsilon_{1t}^* + q_1(\lambda_1, \boldsymbol{\eta}), \varepsilon_{2t}^* + q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta}); \boldsymbol{\eta})}{\partial \varepsilon_{2t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^*. \end{aligned}$$

Finally, undoing the change of variable we obtain

$$\begin{aligned} \frac{\partial q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= - \left(\int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} \int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} \frac{\partial f_{1,2}(\varepsilon_{1t}^*, \varepsilon_{2t}^*; \boldsymbol{\eta})}{\partial \varepsilon_{2t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \right)^{-1} \\ &\times \left\{ \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} \int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} \frac{\partial f_{1,2}(\varepsilon_{1t}^*, \varepsilon_{2t}^*; \boldsymbol{\eta})}{\partial \varepsilon_{1t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \right. \\ &\times \frac{1}{f_1[q_1(\lambda_1, \boldsymbol{\eta}); \boldsymbol{\eta}]} \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} \frac{\partial f_1(\varepsilon_{1t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{1t}^* \\ &\left. - \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} \int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} \frac{f_{1,2}(\varepsilon_{1t}^*, \varepsilon_{2t}^*; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \right\}. \end{aligned}$$

D Computational details

D.1 Scores and first order conditions

The damping factor (7) reduces to

$$\delta^t[\varsigma_t(\boldsymbol{\theta}), \eta] = (N\eta + 1)/[1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})]$$

for the Student t ,

$$\delta^k[\varsigma_t(\boldsymbol{\theta}), \kappa] = [N(N + 2)\kappa\varsigma_t^{-1}(\boldsymbol{\theta}) + 2]/[(N + 2)\kappa + 2]$$

for the Kotz,

$$\delta^{ds}[\varsigma_t(\boldsymbol{\theta}), \alpha, \varkappa] = [\alpha + (1 - \alpha)\varkappa] \cdot \frac{\alpha + (1 - \alpha)\varkappa^{-(N/2+1)} \exp\left[-\frac{[\alpha+(1-\alpha)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}{\alpha + (1 - \alpha)\varkappa^{-N/2} \exp\left[-\frac{[\alpha+(1-\alpha)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}$$

for the DSMN, and

$$\delta^{pe}[\varsigma_t(\boldsymbol{\theta}), c_2, c_3] = 1 - \frac{\sum_{j=1}^J c_j p_{N/2,j}[\varsigma_t(\boldsymbol{\theta})]}{1 + \sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\boldsymbol{\theta})]}$$

for the PE.

As for $\mathbf{e}_{rt}(\boldsymbol{\theta}, \boldsymbol{\eta})$, Fiorentini Sentana and Calzolari (2003) show that in the multivariate Student t case it becomes

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

In addition, Amengual and Sentana (2011) show that

$$s_{\kappa}^k(\boldsymbol{\theta}, \kappa) = \frac{N(N + 2)}{b^2(\kappa)} \left\{ \left(\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1 \right) - \left[\ln \varsigma_t(\boldsymbol{\theta}) - \psi\left(\frac{N}{b(\kappa)}\right) - \ln b(\kappa) \right] \right\},$$

in the case of the Kotz. For the multivariate discrete scale mixture of normals, we can use (4) to write the score with respect to the mixing parameter α as

$$s_{\alpha}^{ds}(\boldsymbol{\theta}, \alpha, \varkappa) = \frac{N}{2} \frac{1 - \varkappa}{[\alpha + \varkappa(1 - \alpha)]} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \alpha},$$

where

$$\begin{aligned} \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \alpha} &= \frac{1}{\exp(g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}])} \left\{ \exp\left(-\frac{1}{2\varpi} \varsigma_t(\boldsymbol{\theta})\right) - \varkappa^{-N/2} \exp\left(-\frac{1}{2\varpi \varkappa} \varsigma_t(\boldsymbol{\theta})\right) \right\} \\ &\quad - \frac{1}{\exp(g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}])} \frac{1 - \varkappa}{[\alpha + \varkappa(1 - \alpha)]^2} \frac{1}{2\varpi^2} \varsigma_t(\boldsymbol{\theta}) \\ &\quad \times \alpha \exp\left(-\frac{1}{2\varpi} \varsigma_t(\boldsymbol{\theta})\right) + (1 - \alpha) \varkappa^{-N/2-1} \exp\left(-\frac{1}{2\varpi \varkappa} \varsigma_t(\boldsymbol{\theta})\right), \end{aligned}$$

and the score with respect to the relative scale parameter \varkappa as

$$s_{\varkappa}^{ds}(\boldsymbol{\theta}, \alpha, \varkappa) = \frac{N}{2} \frac{1 - \alpha}{[\alpha + \varkappa(1 - \alpha)]} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varkappa},$$

where

$$\begin{aligned} \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varkappa} &= \frac{1}{\exp(g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}])} \left\{ -\frac{N}{2} (1 - \alpha) \varkappa^{-N/2-1} \exp\left(-\frac{1}{2\varpi \varkappa} \varsigma_t(\boldsymbol{\theta})\right) \right. \\ &\quad \left. - \varsigma_t(\boldsymbol{\theta}) \frac{1 - \alpha}{2} \alpha \exp\left(-\frac{1}{2\varpi} \varsigma_t(\boldsymbol{\theta})\right) \right. \\ &\quad \left. + (1 - \alpha) \alpha \frac{\varsigma_t(\boldsymbol{\theta})}{2\varkappa^2} \varkappa^{-N/2} \exp\left(-\frac{1}{2\varpi \varkappa} \varsigma_t(\boldsymbol{\theta})\right) \right\}. \end{aligned}$$

Finally, the scores of the 3^{rd} order PE distribution with respect to c_2 and c_3 will be

$$s_{c_2 t}^{pe}(\boldsymbol{\theta}, c_2, c_3) = \frac{p_{N/2,2}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}{1 + \sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}$$

and

$$s_{c_3 t}^{pe}(\tilde{\boldsymbol{\theta}}_T, c_2, c_3) = \frac{p_{N/2,3}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}{1 + \sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}.$$

We can then use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of $\boldsymbol{\eta}$ as $\tilde{\boldsymbol{\eta}}_T = \arg \max_{\boldsymbol{\eta}} L_T(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$, possibly subject to some inequality constraints on $\boldsymbol{\eta}$. For example, in the Student t case $\tilde{\boldsymbol{\eta}}_T$ will be characterised by the first-order Kuhn-Tucker (KT) conditions

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

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

Fiorentini, Sentana and Calzolari (2003) show that in the multivariate Student t case $s_{\eta t}(\boldsymbol{\theta}, 0)$ is proportional to the second generalised Laguerre polynomial. Similarly, Amengual and Sentana (2011) show that this is also the case for the score of the scale parameter of a DSMN. Therefore,

$$s_{\eta t}^t(\boldsymbol{\theta}, 0) = s_{\varkappa t}^{ds}(\boldsymbol{\theta}, 0) = p_{N/2-1,2}^g[\varsigma_t(\boldsymbol{\theta})].$$

Amengual and Sentana (2011) also provide the corresponding expressions for the α -component of $e_{rt}^{DSMN}(\boldsymbol{\theta}, 0)$ in the case of “outliers”, which is given by

$$\lim_{\alpha \rightarrow 0^+} s_{\alpha t}^{ds}(\boldsymbol{\theta}, \alpha, \varkappa) = \varkappa^{N/2} \exp\left(\frac{1-\varkappa}{2}\varsigma_t(\boldsymbol{\theta})\right) - 1 - \frac{1-\varkappa}{2\varkappa}(\varsigma_t(\boldsymbol{\theta})-N).$$

In contrast, in the case of “inliers” it will be given by

$$\lim_{\alpha \rightarrow 1^-} s_{\alpha t}^{ds}(\boldsymbol{\theta}, \alpha, \varkappa) = 1 - \varkappa^{-N/2} \exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right) - \frac{1-\varkappa}{2}(\varsigma_t(\boldsymbol{\theta})-N).$$

As for the polynomial expansion, we saw in Appendix C.1 that the shape parameters are also inequality constrained. Not surprisingly, Amengual and Sentana (2011) also show that $\mathbf{e}_{rt}^{pe}(\boldsymbol{\theta}, \mathbf{0}) = \{p_{N/2-1,2}^g[\varsigma_t(\boldsymbol{\theta})], p_{N/2-1,3}^g[\varsigma_t(\boldsymbol{\theta})]\}'$.

In contrast, the shape parameter κ in the Kotz distribution is unrestricted, and so the first order condition is standard.

D.2 Numerical issues

D.2.1 Random number generation

We sample Student t and DSMN exploiting the decomposition presented in section 2.1. Specifically, we simulate standardised versions of all these distributions by appropriately mixing a N -dimensional spherical normal vector with a univariate gamma random variable, and, in the case of DSMN, a draw from a scalar uniform, which we obtain from the NAG Fortran 77 Mark 19 library routines G05DDF, G05FFF and G05CAF, respectively (see Numerical Algorithm Group (2001) for details). To draw innovations from a PE, we use a modification of the inversion method. Specifically, we first compute the square Euclidean norm of the N -dimensional spherical normal vector, ζ say, which is distributed as a χ^2 with N degrees of freedom. We then use the G05NCF routine to find the solution to the equation $F(\varsigma, c_2, c_3, N) = F_{\chi_N^2}(\zeta)$, where

$$\begin{aligned} F(\varsigma, c_2, c_3, N) &= 1 - \frac{\Gamma(N/2, \varsigma/2)}{\Gamma(N/2)} - c_2 \times \frac{\varsigma^{N/2} e^{-\varsigma/2}}{2^{N/2+2} \Gamma(N/2+2)} (\varsigma - 2 - N) \\ &\quad + c_3 \times \frac{\varsigma^{N/2} e^{-\varsigma/2}}{2^{N/2+3} \Gamma(N/2+3)} [\varsigma^2 - 2d(N+4) + (N+2)(N+4)], \end{aligned}$$

with $\varsigma = \zeta$ as starting value. In this way, we make sure that the three distributions that we simulate share the random draws from the underlying $N \times 1$ uniform vector, which minimises Monte Carlo variability.

D.2.2 Estimation strategy

Our estimation procedure employs the following numerical strategy. First, we estimate the conditional mean and variance parameters $\boldsymbol{\theta}$ under normality with a scoring algorithm that combines the E04LBF routine with the analytical expressions for the score and the $\mathcal{A}(\phi_0)$ matrix in Proposition 2. Then, we compute consistent estimators of $\boldsymbol{\eta}$ using the expressions in Appendix D.2.3, which we use as initial values for the optimisation procedure that obtains the sequential ML estimator $\tilde{\eta}_T$ with the E04JYF routine. This estimator is then used as initial value for the efficient sequential MM estimator, which is obtained with the C05NCF routine. Since our model admits reparametrisation (1), we use expression (17) with M_{ss} and M_{sr} computed either analytically, or by Monte Carlo integration or quadrature. Finally, we use again $\tilde{\eta}_T$ as initial value for the sequential GMM estimators based on orthogonal polynomials using the E04JYF routine.

D.2.3 Initial consistent estimators of shape parameters

Student t The initial value of η is the moment estimator proposed by Fiorentini, Sentana and Calzolari (2003):

$$\eta_{init} = \max \left[\frac{\hat{\tau}_2^t}{(4\hat{\tau}_2^t + 2)}, 0 \right] \quad \text{where} \quad \hat{\tau}_2^t = \frac{\frac{1}{T} \sum_{t=1}^T \varsigma_t^2(\tilde{\boldsymbol{\theta}}_T)}{(N+2)N} - 1.$$

Discrete scale mixture of normals The initial values for α and \varkappa are obtained by running a standard EM algorithm that does not impose $E \left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) \right] = N$.

3rd-order polynomial expansion The initial values for c_2 and c_3 are moment estimators obtained as

$$c_{2:init} = \frac{\sum_{t=1}^T \varsigma_t^2(\tilde{\boldsymbol{\theta}}_T)}{4T} - \frac{N(N+2)}{4}$$

and

$$c_{3:init} = \frac{N+4}{2} c_{2:init} + \frac{N(N+2)(N+4)}{24} - \frac{\sum_{t=1}^T \varsigma_t^3(\tilde{\boldsymbol{\theta}}_T)}{24T}.$$

References

- Abramowitz, M. and Stegun I.A. (1964): *Handbook of mathematical functions*, AMS 55, National Bureau of Standards.
- Acharya, V.V., Lasse, H.P., Philippon, T. and Richardson, M. (2010): “Measuring systemic risk”, Federal Reserve Bank of Cleveland Working Paper 10-02.
- Adrian, T. and Brunnermeier, M. (2011): “CoVaR”, mimeo, Princeton.
- Amengual, D. and Sentana, E. (2010): “A comparison of mean-variance efficiency tests”, *Journal of Econometrics* 154, 16-34.
- Amengual, D. and Sentana, E. (2011): “Inference in multivariate dynamic models with elliptical innovations”, mimeo, CEMFI.
- Balestra, P. and Holly, A. (1990): “A general Kronecker formula for the moments of the multivariate normal distribution”, DEEP Cahier 9002, University of Lausanne.
- Berk, J. (1997): “Necessary conditions for the CAPM”, *Journal of Economic Theory* 73, 245-257.
- Berkane, M. and Bentler, P.M. (1986): “Moments of elliptically distributed random variates”, *Statistics and Probability Letters* 4, 333-335.
- Bickel, P.J. (1982): “On adaptive estimation”, *Annals of Statistics* 10, 647-671.
- Bollerslev, T. and Wooldridge, J. M. (1992): “Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances”, *Econometric Reviews* 11, 143-172.
- Bontemps, C. and Meddahi, N. (2011): “Testing distributional assumptions: A GMM approach”, forthcoming in the *Journal of Applied Econometrics*.
- Brownlees, C. and Engle, R.F. (2011): “Volatility, correlation and tails for systemic risk measurement”, mimeo NYU.
- Chamberlain, G. (1983): “A characterization of the distributions that imply mean-variance utility functions”, *Journal of Economic Theory* 29, 185-201.
- Crowder, M.J. (1976): “Maximum likelihood estimation for dependent observations”, *Journal of the Royal Statistical Society B*, 38, 45-53.
- Durbin, J. (1970): “Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables”, *Econometrica* 38, 410-421.
- Fang, K.T., Kotz, S. and Ng, K.W. (1990): *Symmetric multivariate and related distributions*, Chapman and Hall.

Fiorentini, G. and Sentana, E. (2010): “On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models”, mimeo, CEMFI.

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

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

Gouriéroux, C. and Jasiak, J. (2009): “Value at Risk”, in Y. Ait-Sahalia and L.P. Hansen (eds.) *Handbook of Financial Econometrics*, Elsevier.

Guidolin, M. and Timmermann, A. (2007): “Asset allocation under multivariate regime switching”, *Journal of Economic Dynamics and Control* 31, 3503-3544.

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

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

Hansen, L.P., Heaton, J. and Yaron, A. (1996): “Finite sample properties of some alternative GMM estimators”, *Journal of Business and Economic Statistics* 14, 262-280.

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

Jondeau, E. and Rockinger, M. (2003): “Conditional volatility, skewness and kurtosis: Existence, persistence and comovements”, *Journal of Economics Dynamics and Control* 27, 1699-1737.

Koenker, R. (2005): *Quantile regression*, Econometric Society Monograph, Cambridge.

Kotz, S. (1975): “Multivariate distributions at a cross-road”, in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical distributions in scientific work*, vol. I, 247-270, Reidel.

Ling, S. and McAleer, M. (2003): “Asymptotic theory for a vector ARMA-GARCH model”, *Econometric Theory* 19, 280-310.

Linton, O. (1993): “Adaptive estimation in ARCH models”, *Econometric Theory* 9, 539-569.

Longin, F. and Solnik, B. (2001): “Extreme correlation of international equity markets”,

Journal of Finance 56, 649-676.

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

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

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974): *Introduction to the theory of Statistics*, (3rd ed.), McGraw Hill.

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

Newey, W.K. (1984): "A method of moments interpretation of sequential estimators", *Economics Letters* 14, 201-206.

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047-1070.

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

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

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-599.

Owen, J. and Rabinovitch R. (1983): "On the class of elliptical distributions and their applications to the theory of portfolio choice", *Journal of Finance* 38, 745-752.

Pagan, A. (1986): "Two stage and related estimators and their applications", *Review of Economic Studies* 53, 517-538.

Pesaran, M.H., Schleicher, C. and Zaffaroni, P. (2009): "Model averaging in risk management with an application to futures markets", *Journal of Empirical Finance* 16(2), 280-305.

Pesaran B. and Pesaran, M.H. (2010): "Conditional volatility and correlations of weekly returns and the VaR analysis of the 2008 stock market crash", *Economic Modelling* 27, 1398-1416.

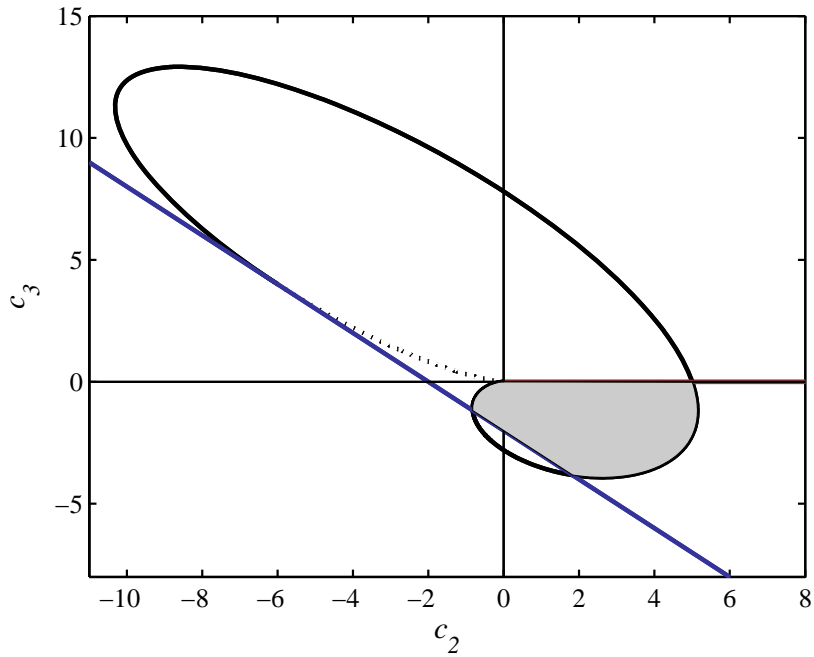
Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415-443.

Table 1: Finite sample properties of sequential estimators of the shape parameters

	True value	ESMM	SML	SMM
		Mean (Std.Dev)	Mean (Std.Dev)	Mean (Std.Dev)
<hr/> Student t <hr/>				
η	0.1	0.099 (0.007)	0.099 (0.007)	0.098 (0.011)
<hr/> DSMN <hr/>				
α	0.05	0.051 (0.009)	0.051 (0.010)	0.055 (0.014)
\varkappa	0.246	0.251 (0.023)	0.251 (0.023)	0.256 (0.030)
<hr/> PE <hr/>				
c_2	2.916	2.895 (0.128)	2.895 (0.128)	2.887 (0.171)
c_3	-1	-0.984 (0.186)	-0.984 (0.185)	-0.967 (0.399)

Notes: 1,600 replications, $T = 2,500$, $N = 5$. ESMM and SML refer to the efficient sequential MM and sequential ML estimators, respectively. The orthogonal polynomial MM estimator is labeled SMM. For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components; In turn, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2 - 1$ in the case of PE innovations. See Section 5.1 and Appendix D.2 for a detailed description of the Monte Carlo study.

Figure 1: Positivity region of a 3rd-order PE



Notes: The solid (dashed) black line represents the frontier defined by positive (negative) values of ς . Notice that if we imposed the above conditions for all $\varsigma \in \mathbb{R}$, then $c_3 = 0$ and $0 < c_2 < N$. Such a frontier, however, is overly restrictive because it does not take into account the non-negativity of ς . In this sense, the red line represents the tangent of $P_3(\varsigma)$ at $\varsigma = 0$ while the blue line is the tangent of $P_3(\varsigma)$ when $\varsigma \rightarrow +\infty$. The grey area, therefore, defines the admissible set in the (c_2, c_3) space. Focusing on $\varsigma \in \mathbb{R}_+$ only allows for a larger range of (c_2, c_3) with $c_3 < 0$, which is given by the difference between the dashed black line and the blue one.

Figure 2a: Standardised bivariate normal density

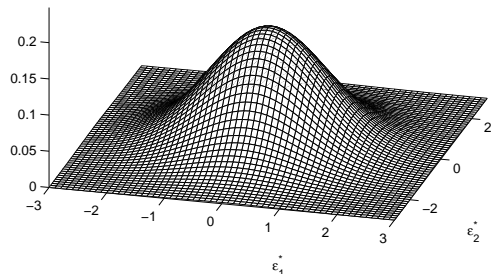


Figure 2b: Contours of a standardised bivariate normal density

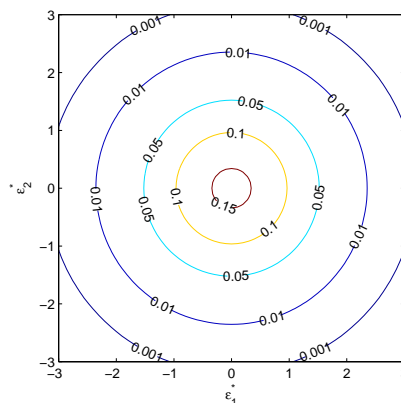


Figure 2c: Standardised bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

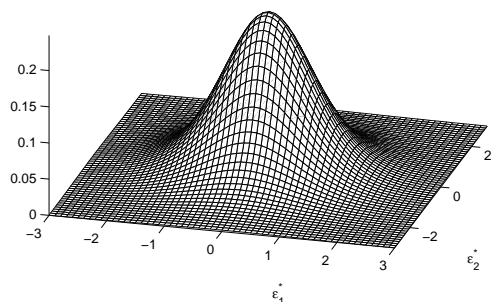


Figure 2d: Contours of a standardised bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

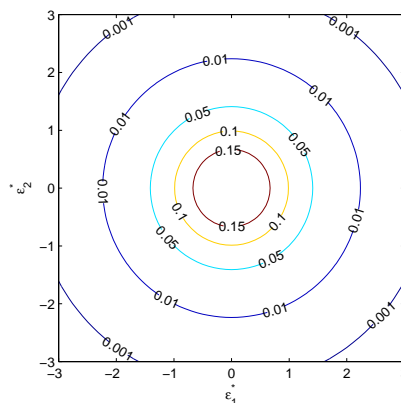


Figure 2e: Standardised bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

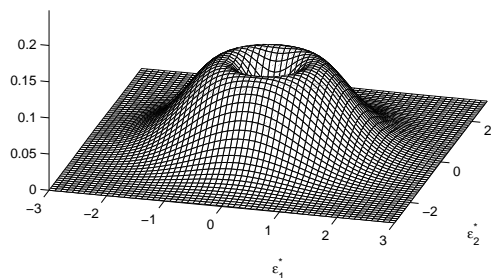


Figure 2f: Contours of a standardised bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

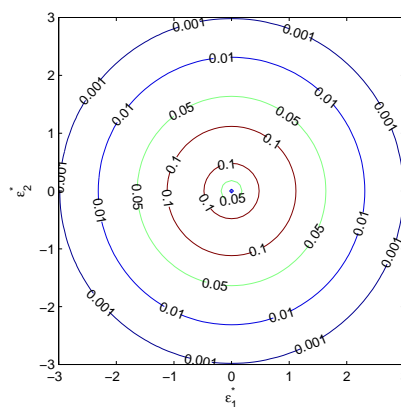


Figure 2g: Standardised bivariate DSMN density with multivariate excess kurtosis $\kappa = 0.125$ ($\alpha = 0.5$)

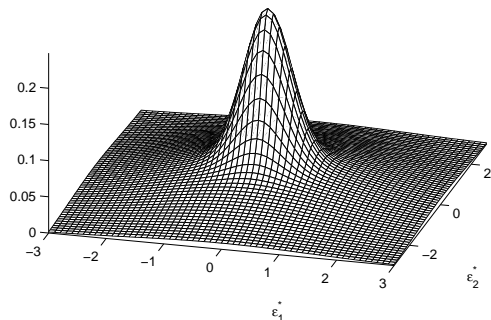


Figure 2h: Contours of a standardised bivariate DSMN density with multivariate excess kurtosis $\kappa = 0.125$ ($\alpha = 0.5$)

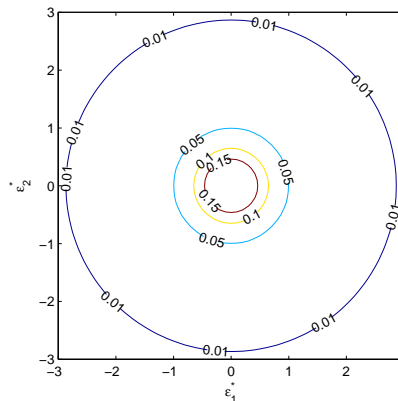


Figure 2i: Standardised bivariate 3rd-order PE with parameters $c_2 = 0$ and $c_3 = -0.2$

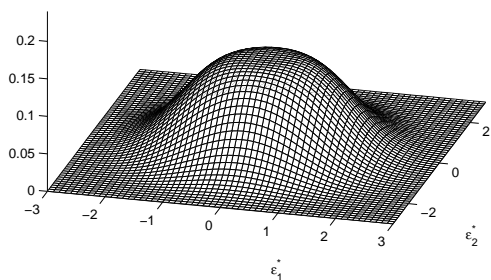


Figure 2j: Contours of a standardised 3rd-order PE with parameters $c_2 = 0$ and $c_3 = -0.2$

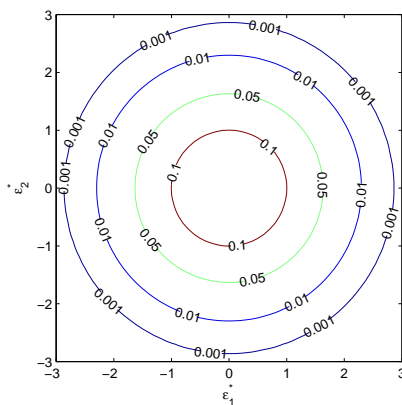
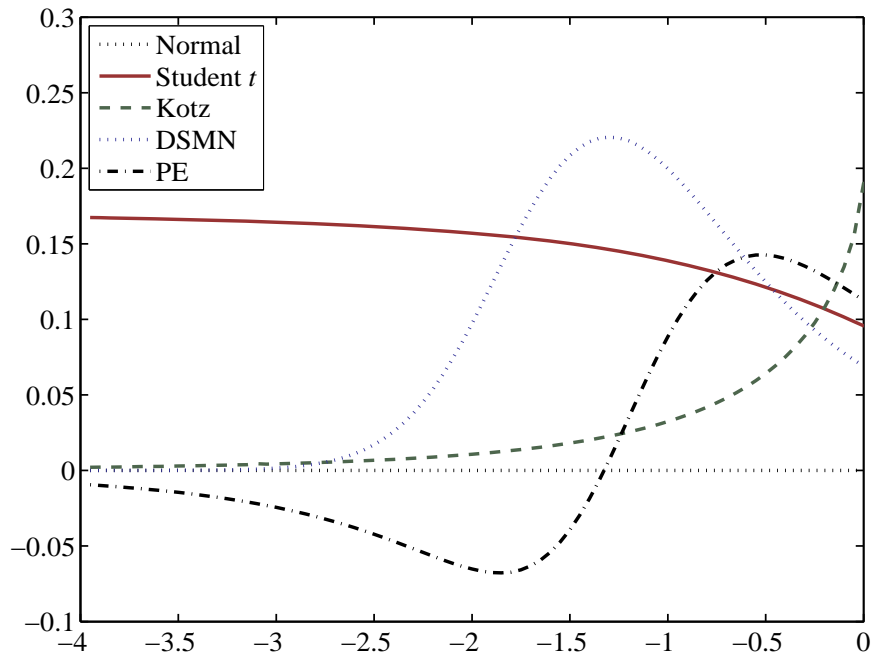


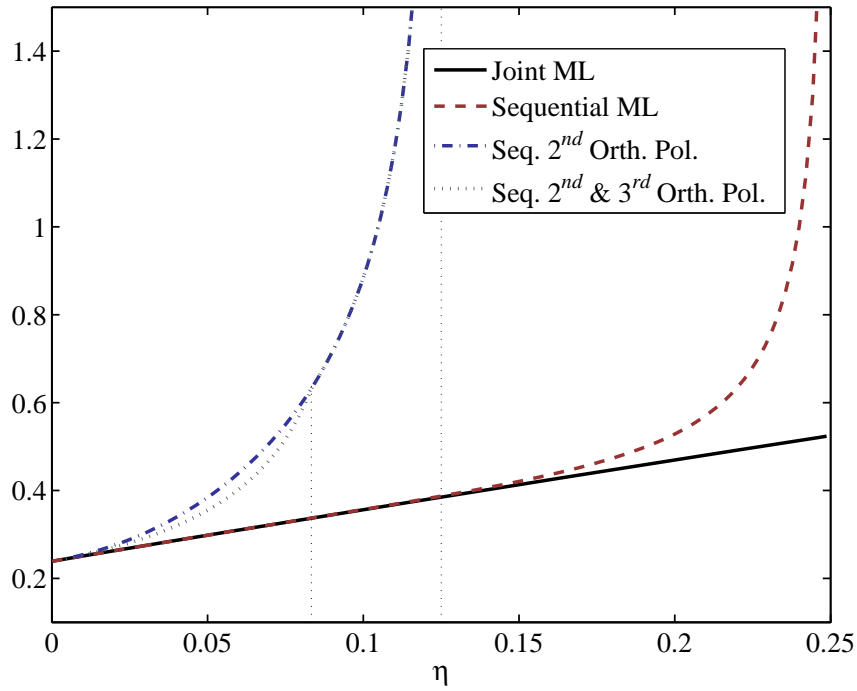
Figure 3: Exceedance correlation



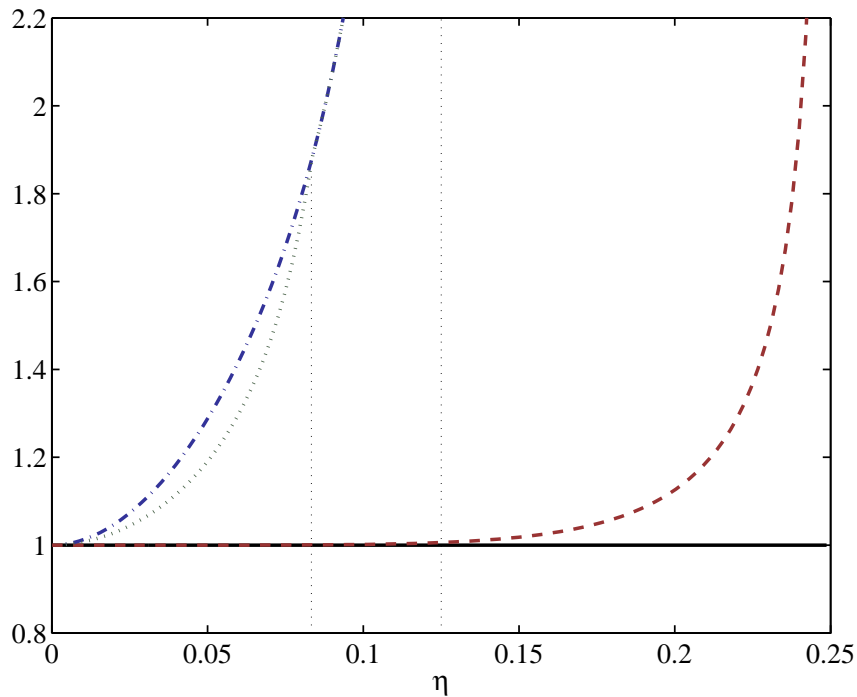
Notes: The exceedance correlation between two variables ε_1^* and ε_2^* is defined as $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \varrho, \varepsilon_2^* > \varrho)$ for positive ϱ and $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \varrho, \varepsilon_2^* < -\varrho)$ for negative ϱ (see Longin and Solnik, 2001). Horizontal axis in standard deviation units. Because all the distributions we consider are elliptical, we only report results for $\varrho < 0$. Student *t* distribution with 10 degrees of freedom, Kotz distribution with the same kurtosis, DSMN with parameters $\alpha = 0.05$ and the same kurtosis and 3rd-order PE with the same kurtosis and $c_3 = -1$.

Figure 4: Asymptotic efficiency of Student t estimators

Asymptotic standard errors of η estimators



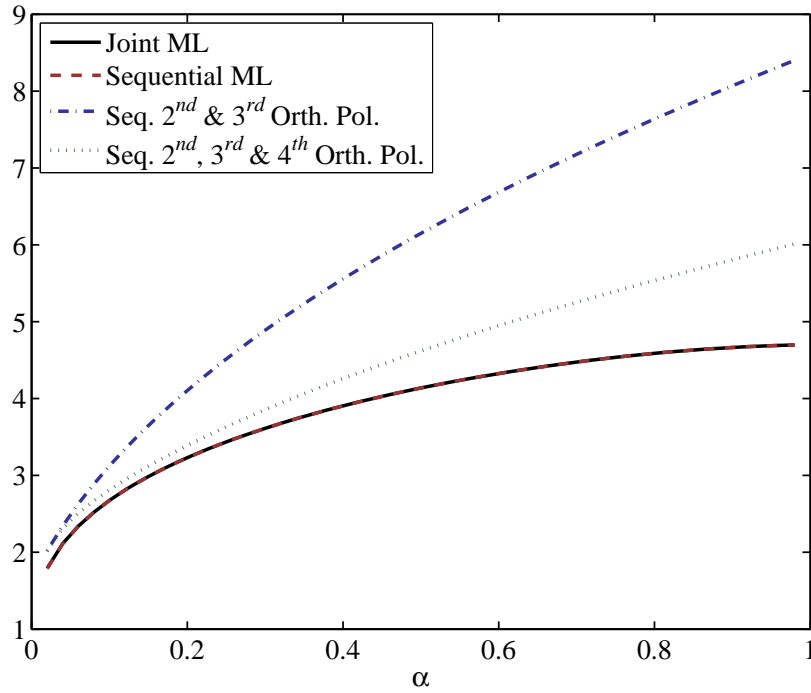
Relative efficiency of η estimators (with respect to Joint ML)



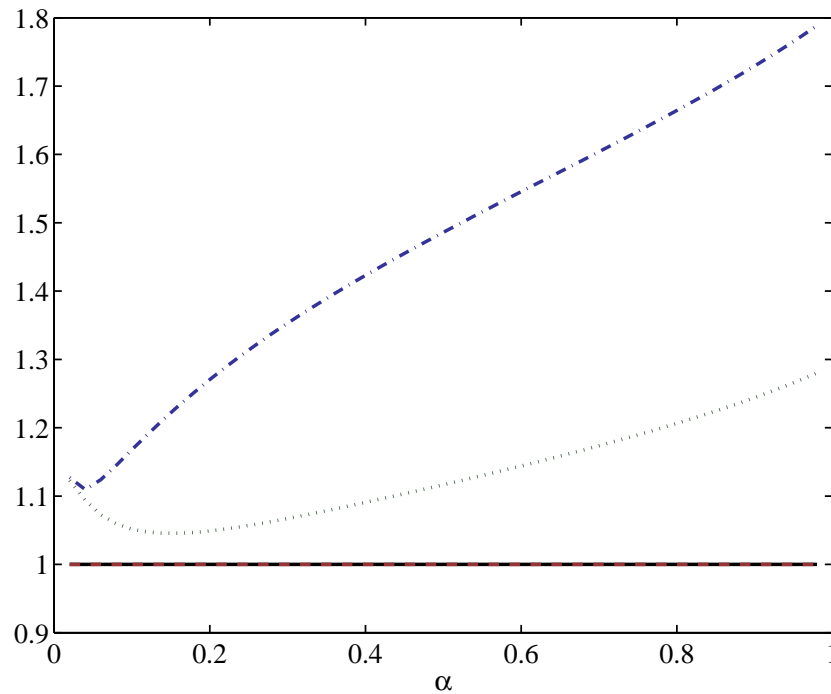
Notes: $N = 5$. For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 5a: Asymptotic efficiency of DSMN estimators ($\varkappa = 0.5$)

Asymptotic standard errors of α estimators



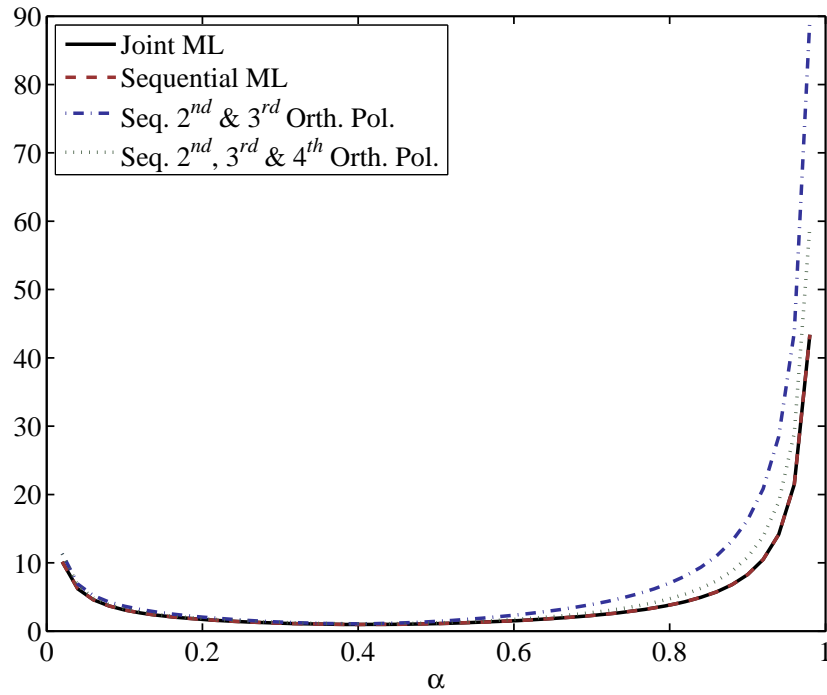
Relative efficiency of α estimators (with respect to Joint ML)



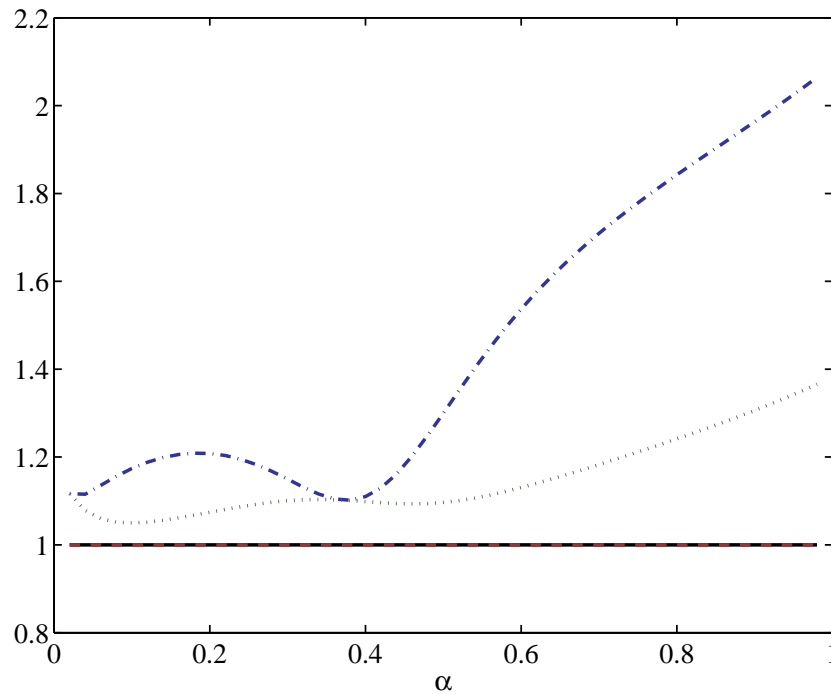
Notes: $N = 5$ and $\varkappa = 0.5$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 5b: Asymptotic efficiency of DSMN estimators ($\varkappa = 0.5$)

Asymptotic standard errors of \varkappa estimators



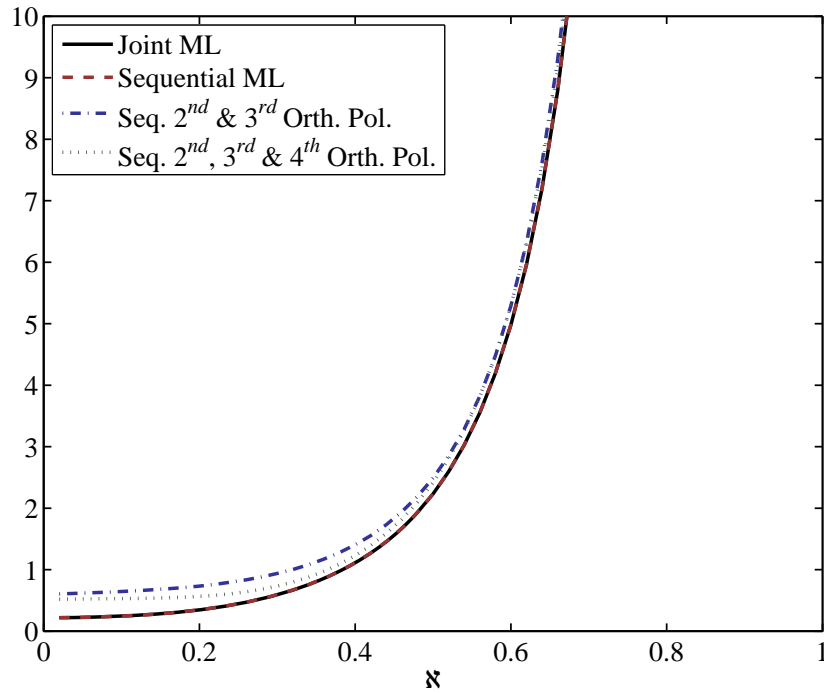
Relative efficiency of \varkappa estimators (with respect to Joint ML)



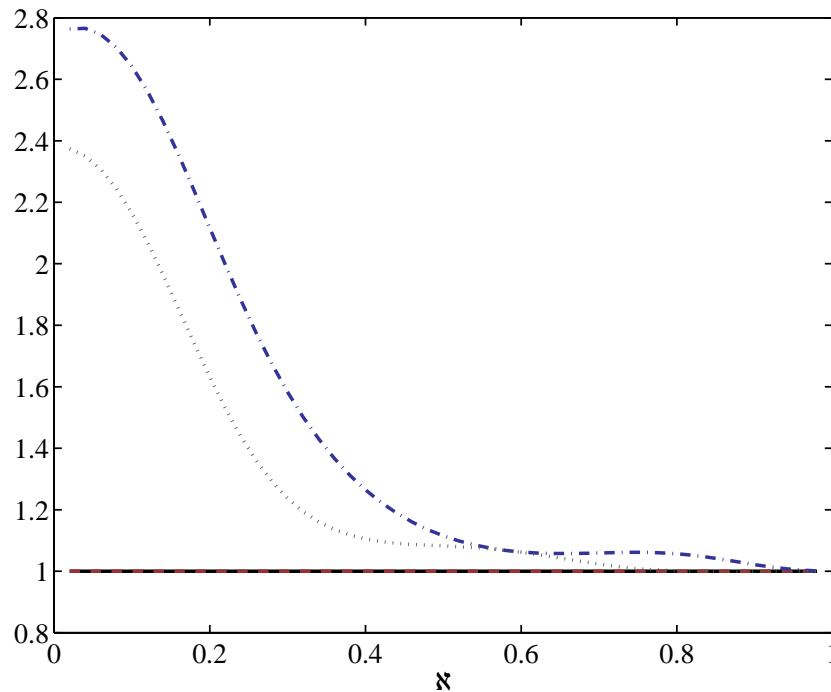
Notes: $N = 5$ and $\varkappa = 0.5$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 5c: Asymptotic efficiency of DSMN estimators ($\alpha = 0.05$)

Asymptotic standard errors of α estimators



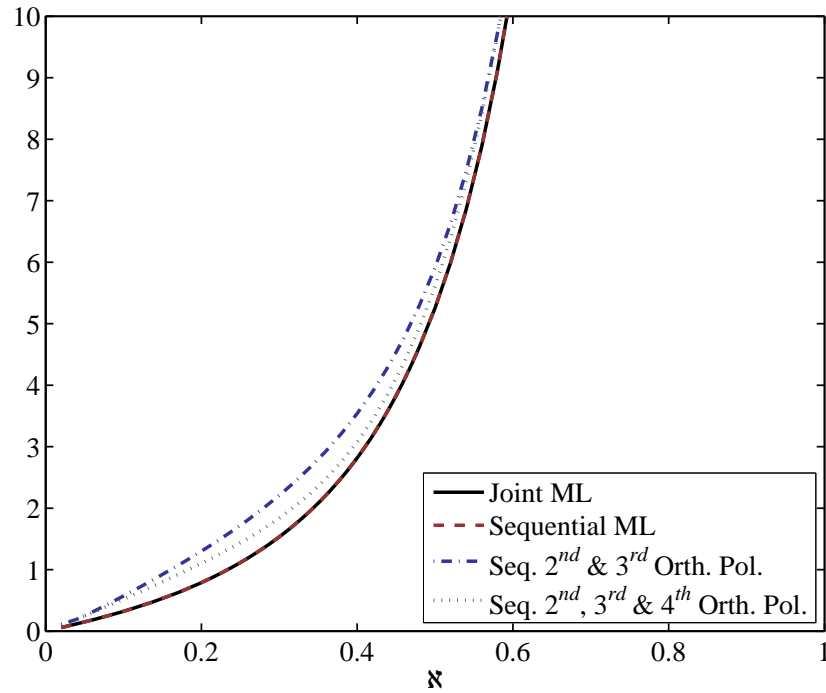
Relative efficiency of α estimators (with respect to Joint ML)



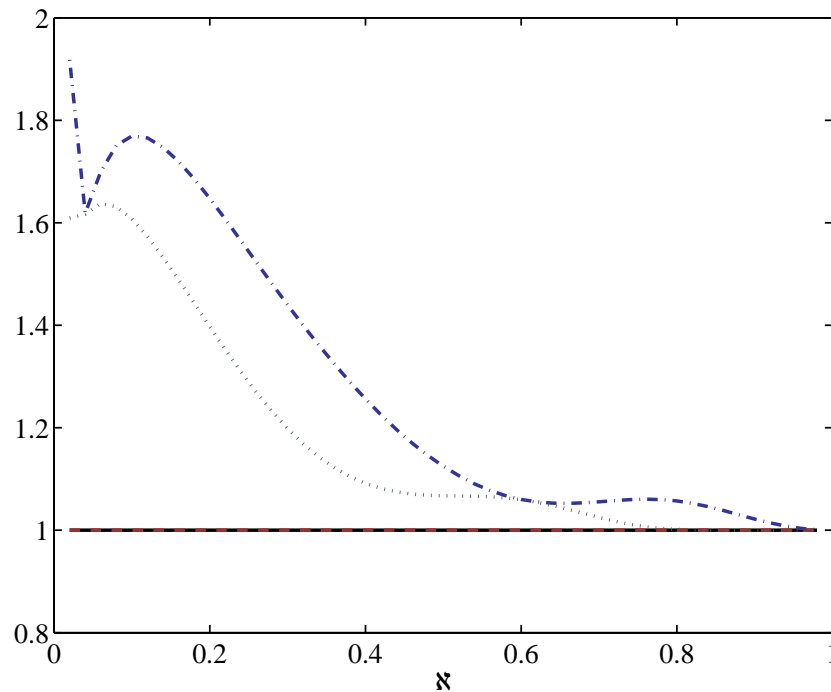
Notes: $N = 5$ and $\alpha = 0.05$. For DSMN innovations, α denotes the mixing probability and ζ is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 5d: Asymptotic efficiency of DSMN estimators ($\alpha = 0.05$)

Asymptotic standard errors of \varkappa estimators



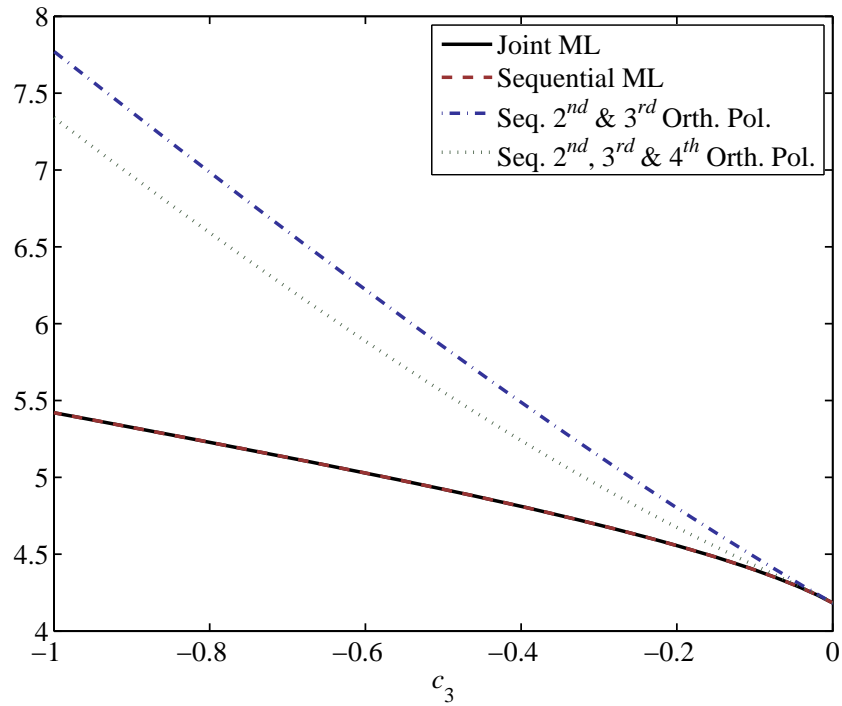
Relative efficiency of \varkappa estimators (with respect to Joint ML)



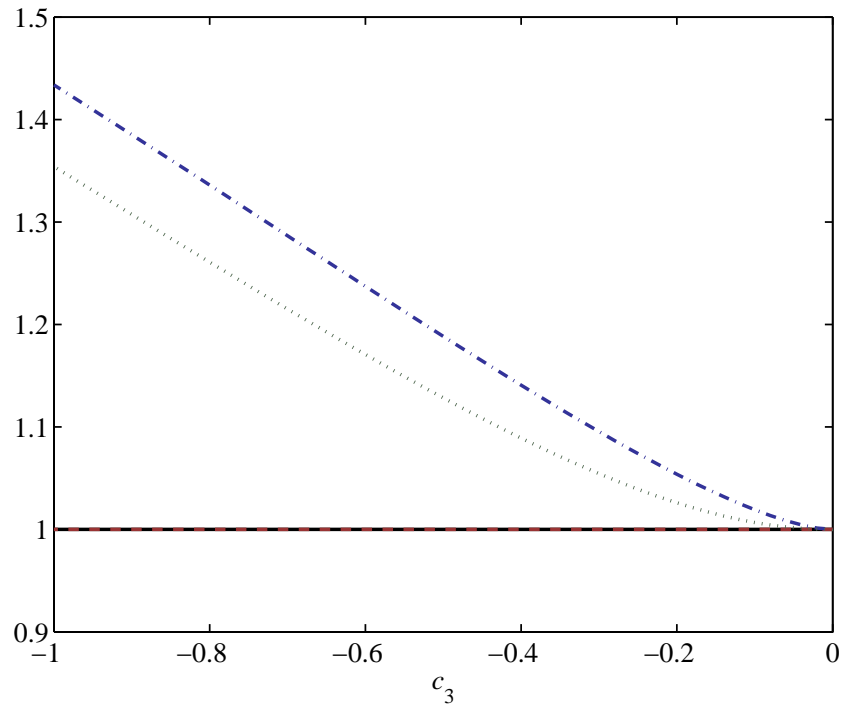
Notes: $N = 5$ and $\alpha = 0.05$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 6a: Asymptotic efficiency of PE estimators ($c_2 = 0$)

Asymptotic standard errors of c_2 estimators



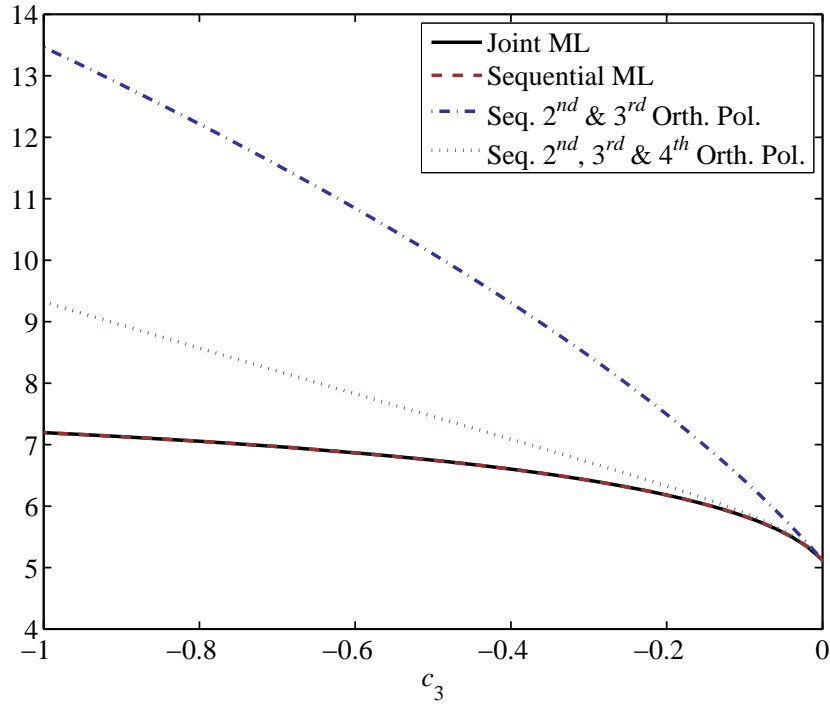
Relative efficiency of c_2 estimators (with respect to Joint ML)



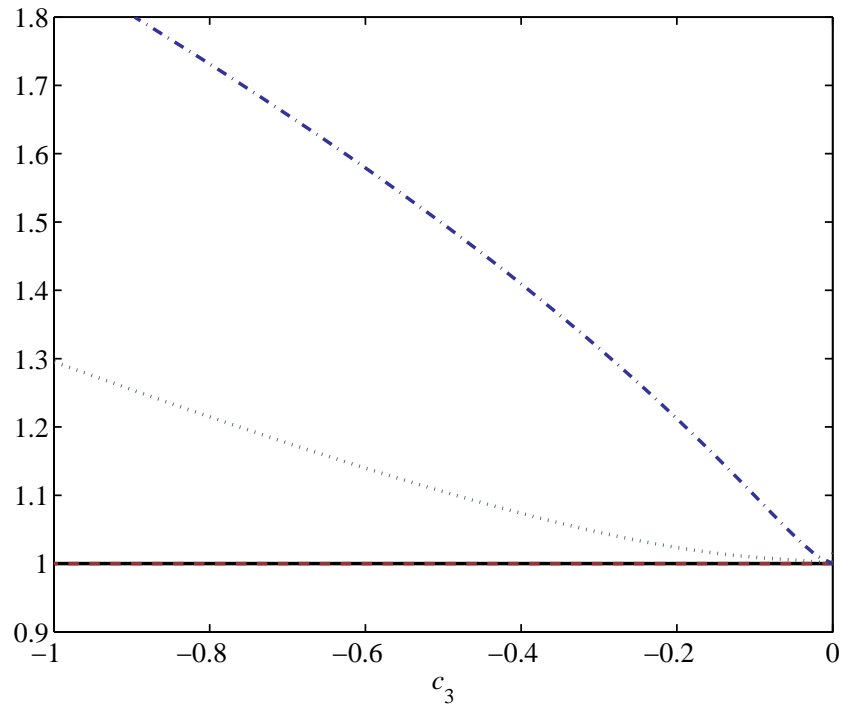
Notes: $N = 5$ and $c_2 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2nd and 3rd Laguerre polynomials with parameter $N/2 - 1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 6b: Asymptotic efficiency of PE estimators ($c_2 = 0$)

Asymptotic standard errors of c_3 estimators



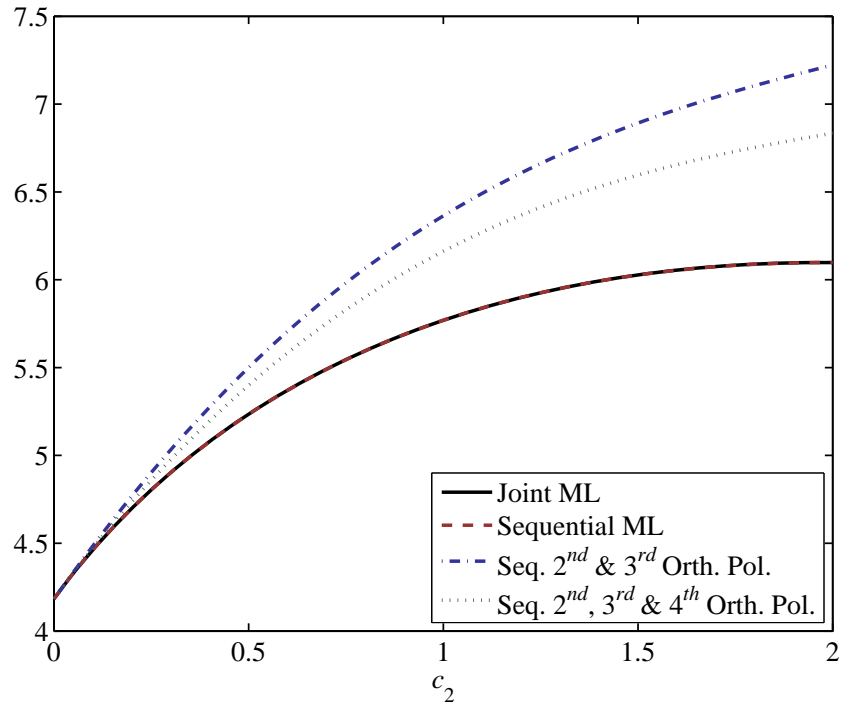
Relative efficiency of c_3 estimators (with respect to Joint ML)



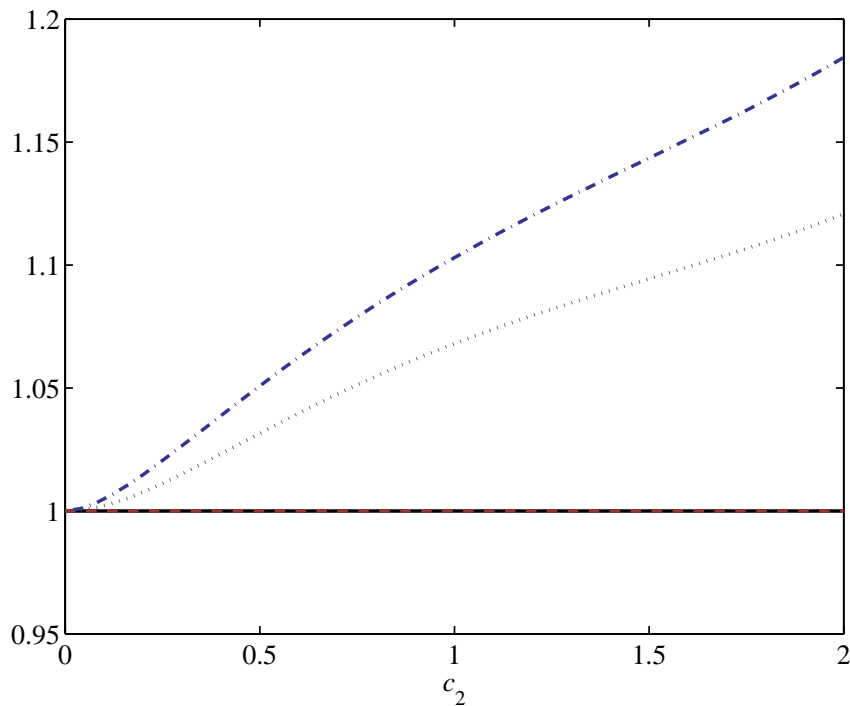
Notes: $N = 5$ and $c_2 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2nd and 3rd Laguerre polynomials with parameter $N/2 - 1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 6c: Asymptotic efficiency of PE estimators ($c_3 = 0$)

Asymptotic standard errors of c_2 estimators



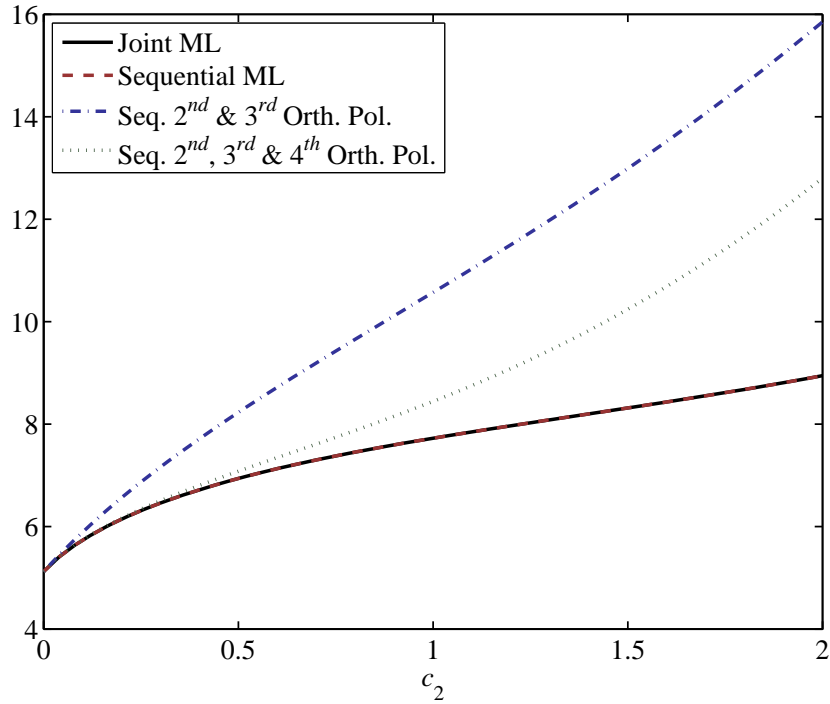
Relative efficiency of c_2 estimators (with respect to Joint ML)



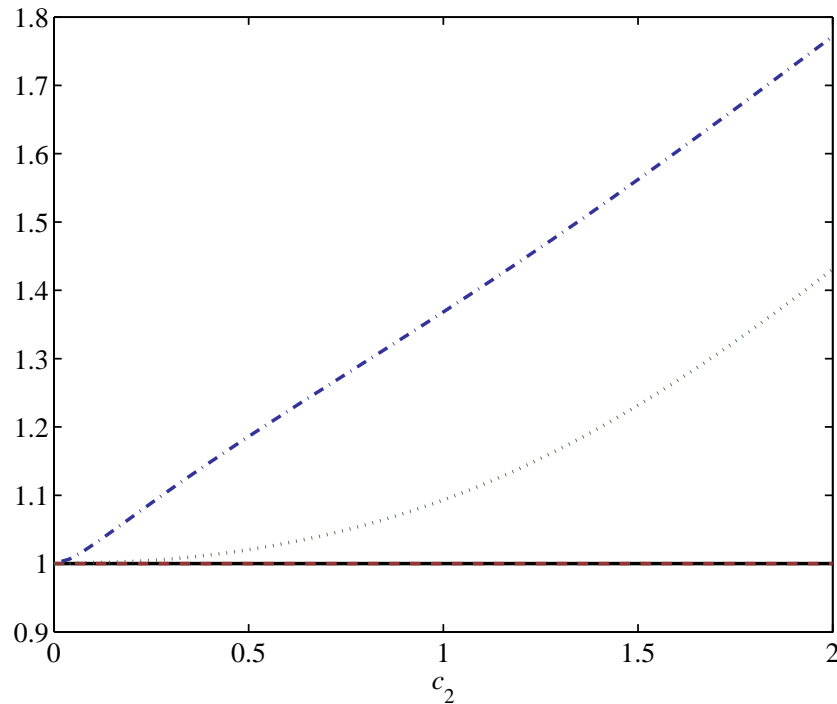
Notes: $N = 5$ and $c_3 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2nd and 3rd Laguerre polynomials with parameter $N/2 - 1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 6d: Asymptotic efficiency of PE estimators ($c_3 = 0$)

Asymptotic standard errors of c_3 estimators

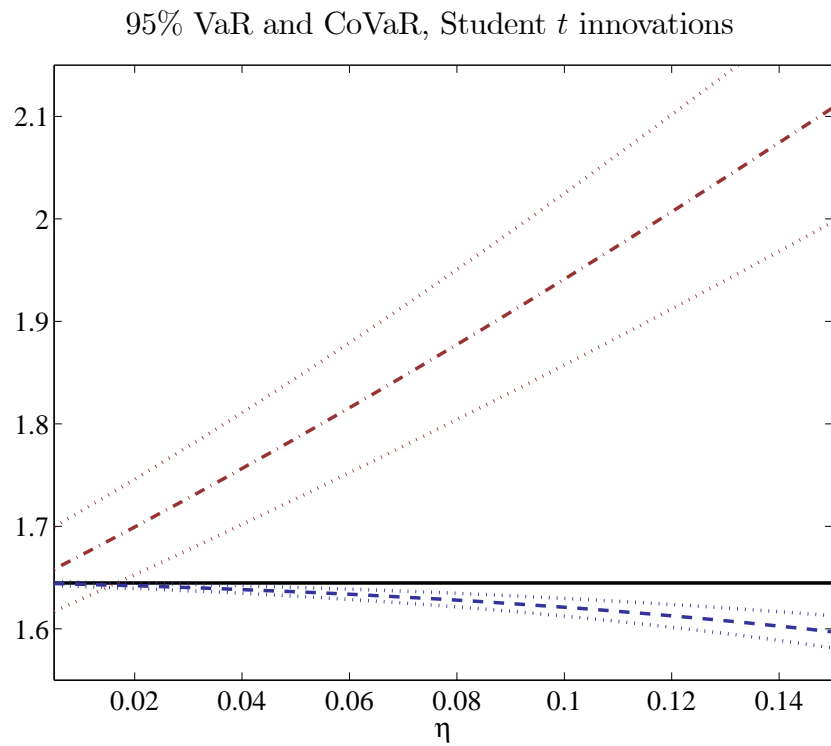
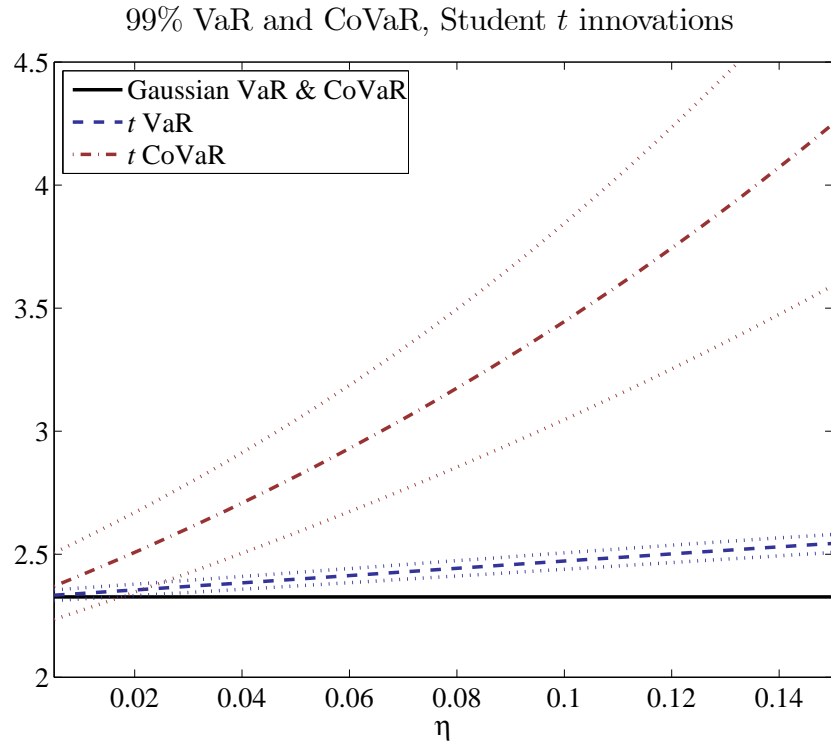


Relative efficiency of c_3 estimators (with respect to Joint ML)



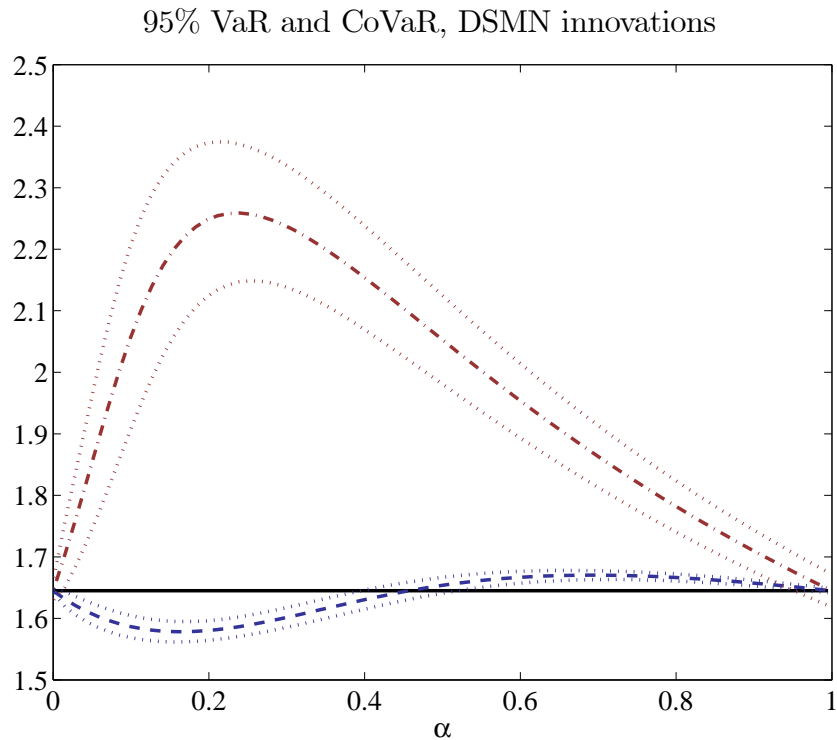
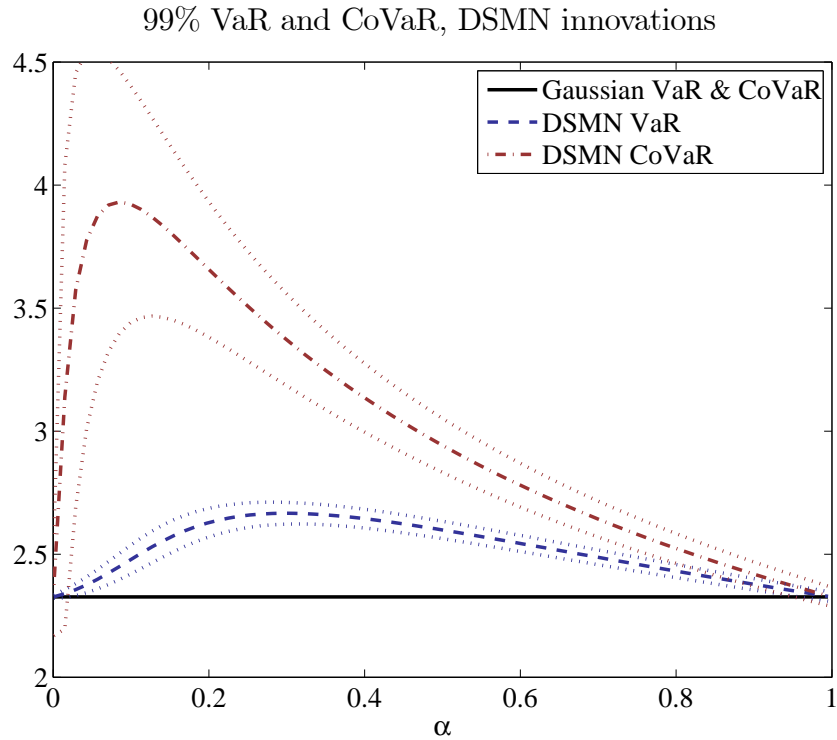
Notes: $N = 5$ and $c_3 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2nd and 3rd Laguerre polynomials with parameter $N/2 - 1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure 7a: VaR, CoVaR and their 95% confidence intervals



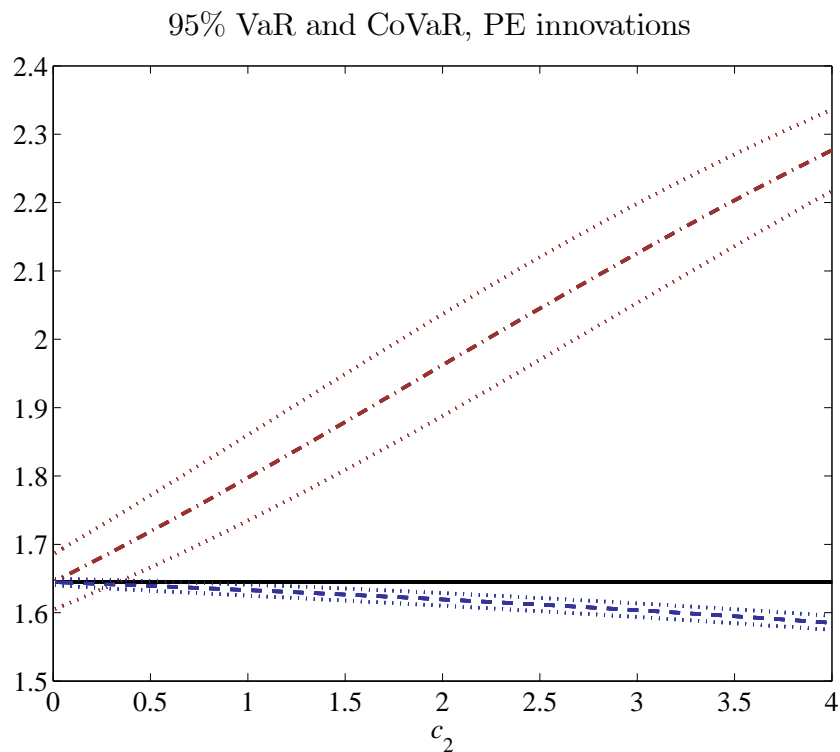
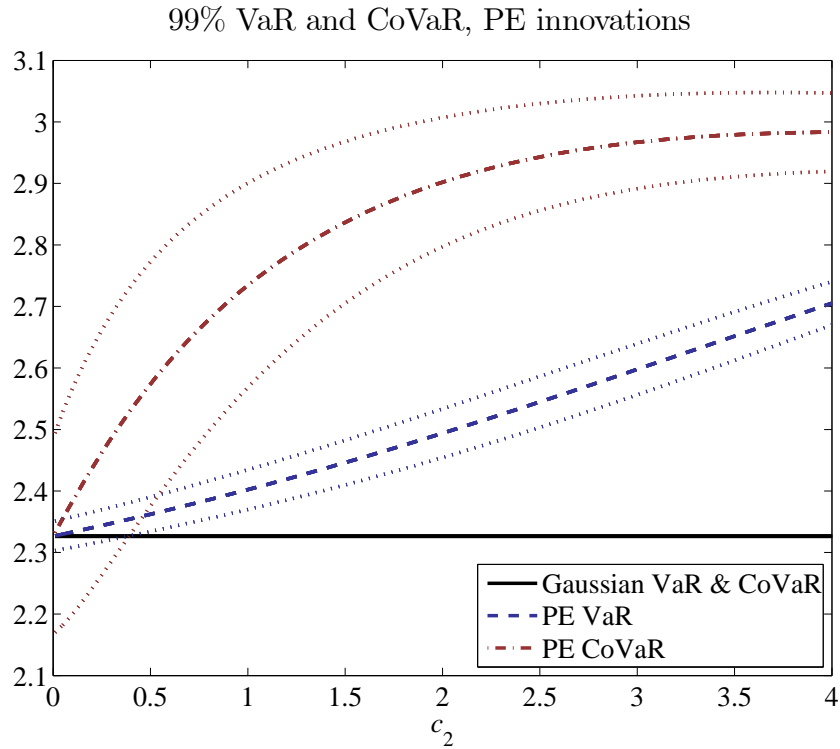
Notes: For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. Dotted lines represent the 95% confidence intervals based on the asymptotic variance of the sequential ML estimator for a hypothetical sample size of $T = 1,000$ and $N = 5$. The horizontal line represents the Gaussian VaR and CoVaR, which have zero standard errors.

Figure 7b: VaR, CoVaR and their 95% confidence intervals



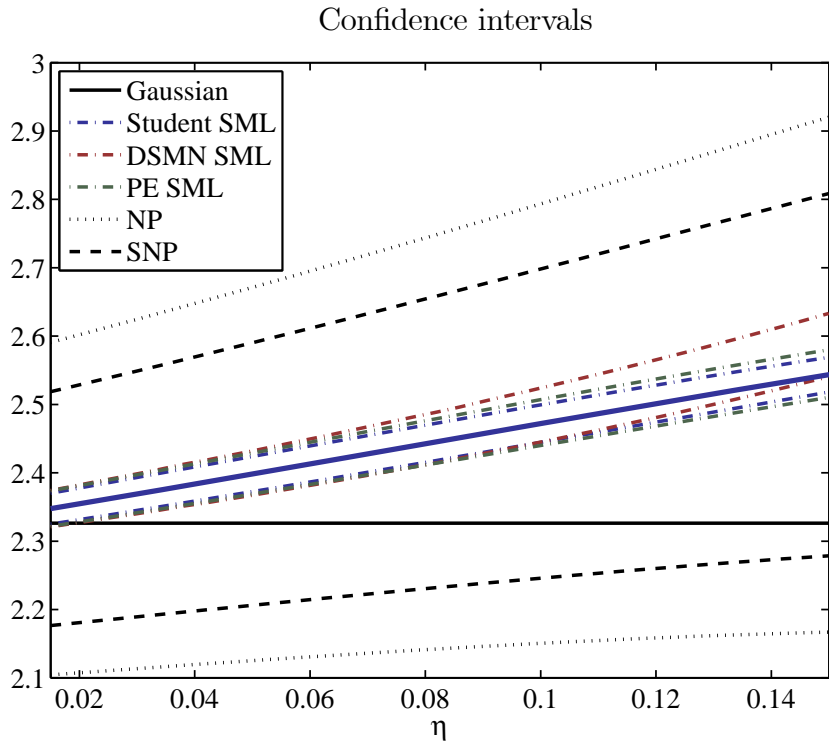
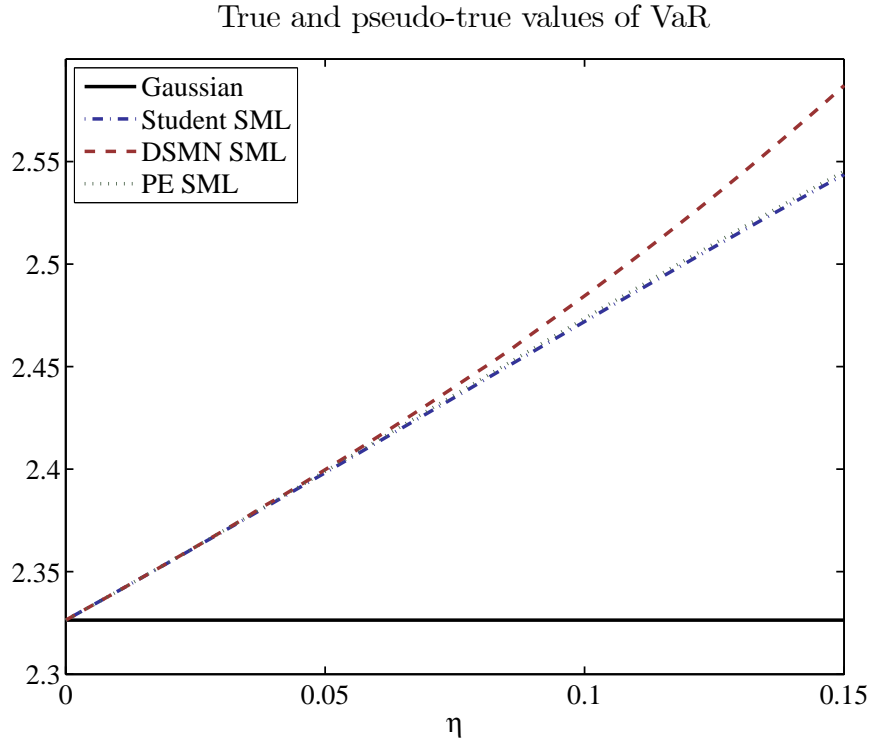
Notes: $\varkappa = 0.25$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components. Dotted lines represent the 95% confidence intervals based on the asymptotic variance of the sequential ML estimator for a hypothetical sample size of $T = 1,000$ and $N = 5$. The horizontal line represents the Gaussian VaR and CoVaR, which have zero standard errors.

Figure 7c: VaR, CoVaR and their 95% confidence intervals



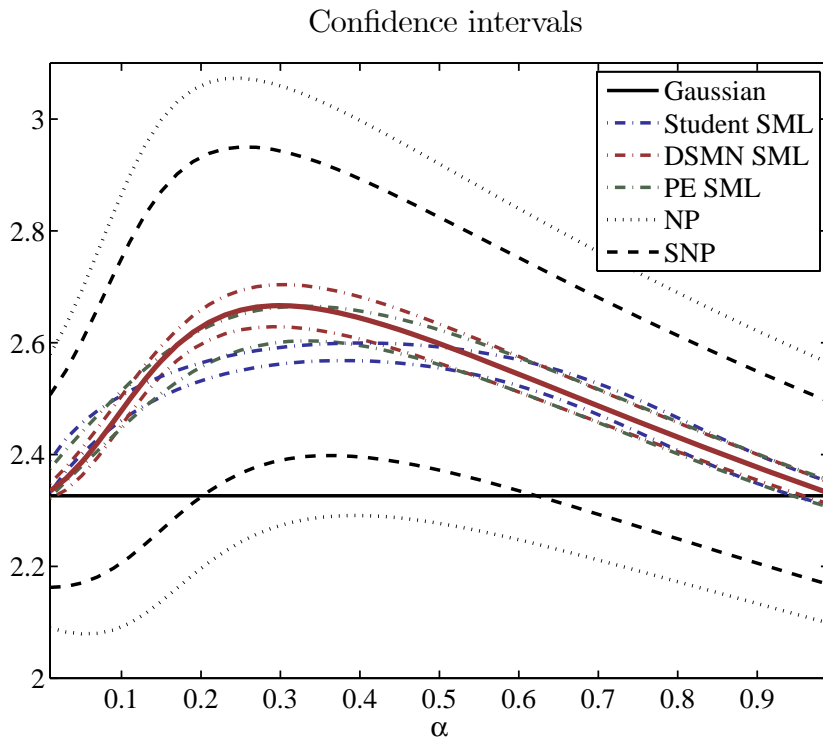
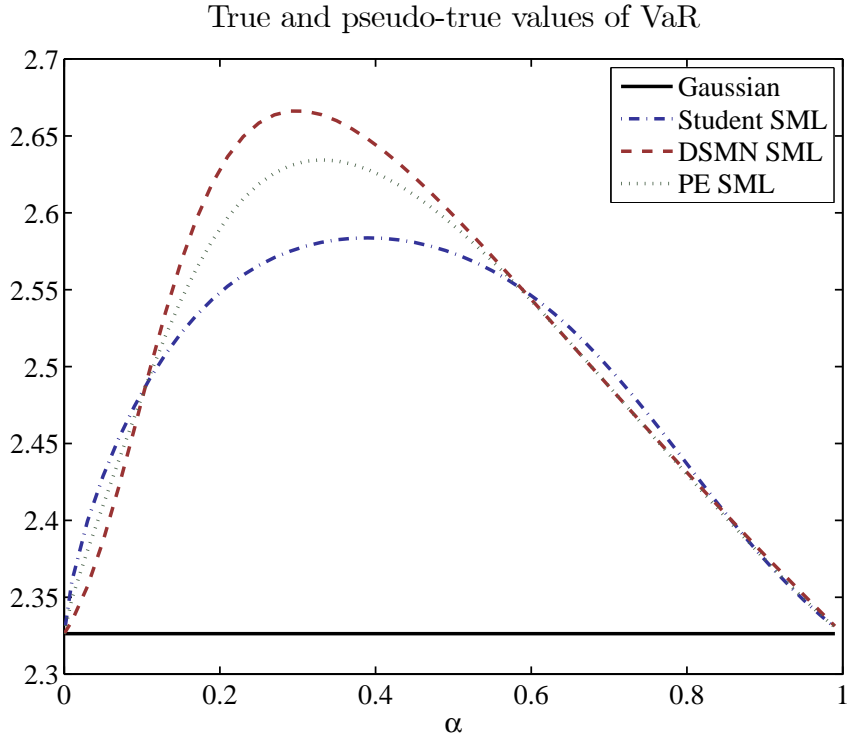
Notes: $c_3 = -c_2/3$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2 - 1$. Dotted lines represent the 95% confidence intervals based on the asymptotic variance of the sequential ML estimator for a hypothetical sample size of $T = 1,000$ and $N = 5$. The horizontal line represents the Gaussian VaR and CoVaR, which have zero standard errors.

Figure 8a: 99% VaR estimators, Student t innovations



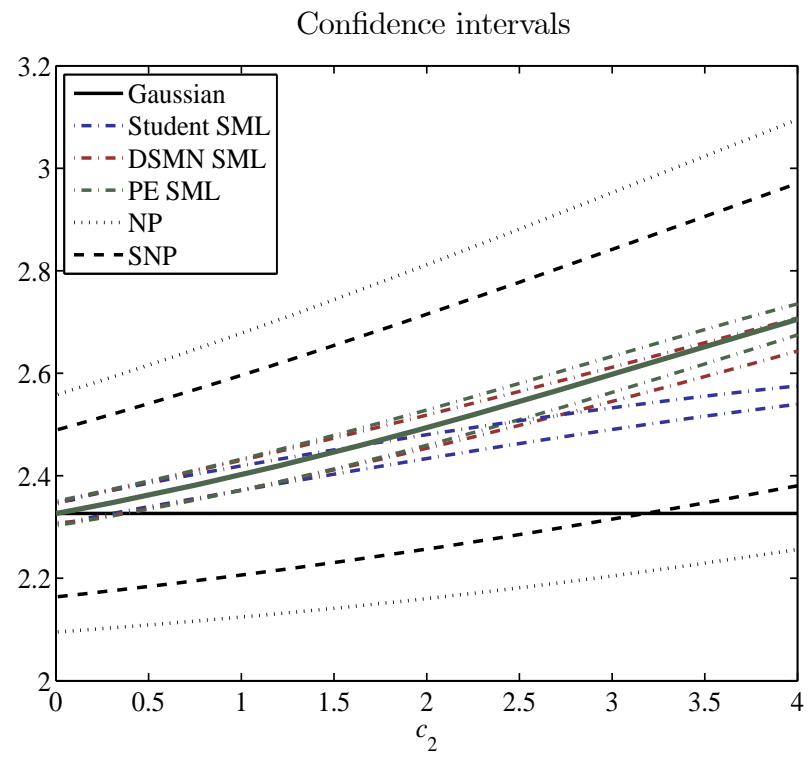
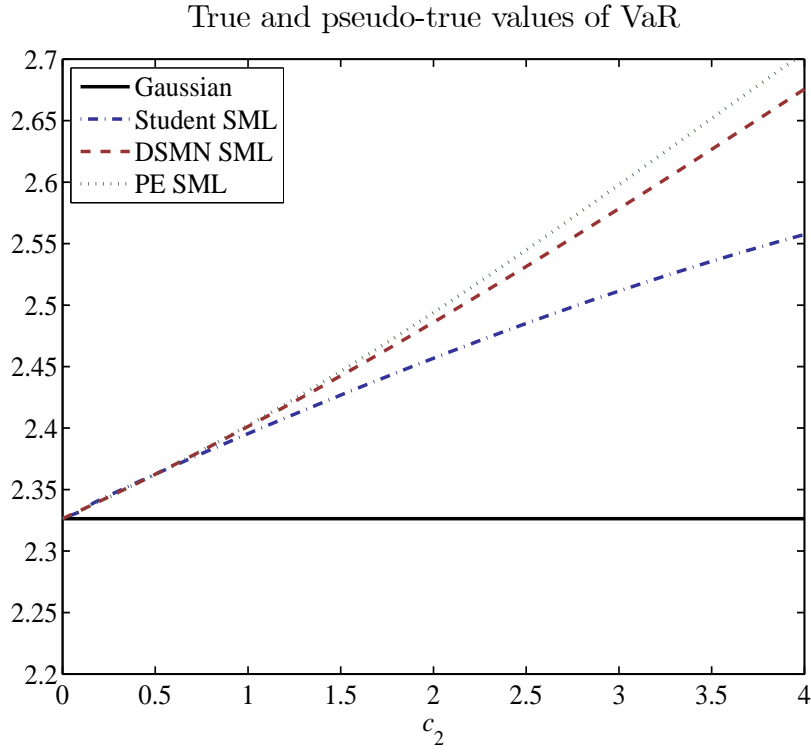
Notes: For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. Confidence intervals are computed using robust standard errors for a hypothetical sample size of $T = 1,000$ and $N = 5$. SML refers to sequential ML, NP refers to the fully nonparametric procedure based on the λ th empirical quantile of the standardised return distribution, while SNP denotes the nonparametric procedure that imposes symmetry of the return distribution (see Section 4.3 for details). The blue solid line is the true VaR.

Figure 8b: 99% VaR estimators, DSMN innovations



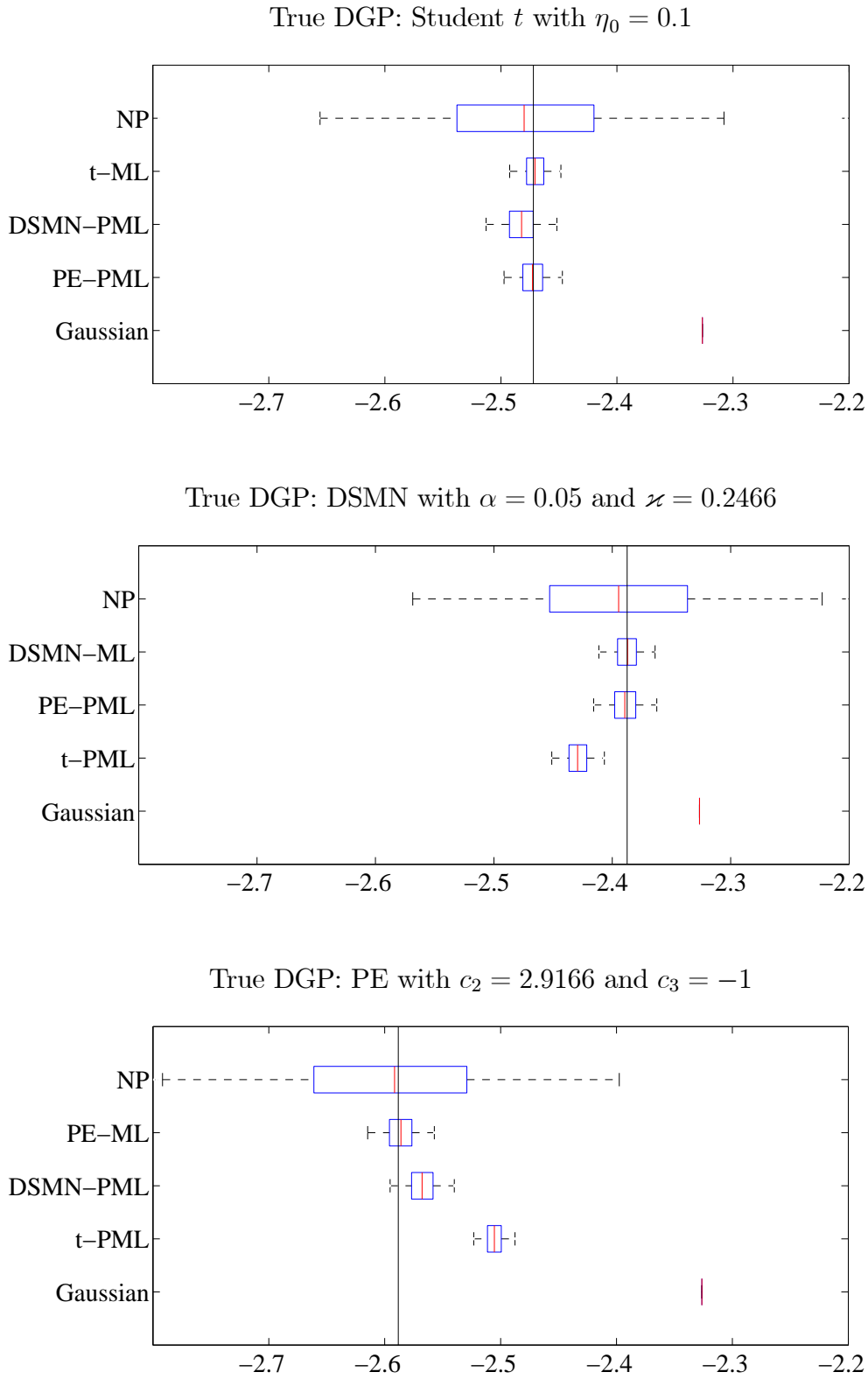
Notes: $\varkappa = 0.25$. For DSMN innovations, α denotes the mixing probability and \varkappa is the variance ratio of the two components. Confidence intervals are computed using robust standard errors for a hypothetical sample size of $T = 1,000$ and $N = 5$. SML refers to sequential ML, NP refers to the fully nonparametric procedure based on the λ th empirical quantile of the standardised return distribution, while SNP denotes the nonparametric procedure that imposes symmetry of the return distribution (see Section 4.3 for details). The red solid line is the true VaR.

Figure 8c: 99% VaR estimators, PE innovations



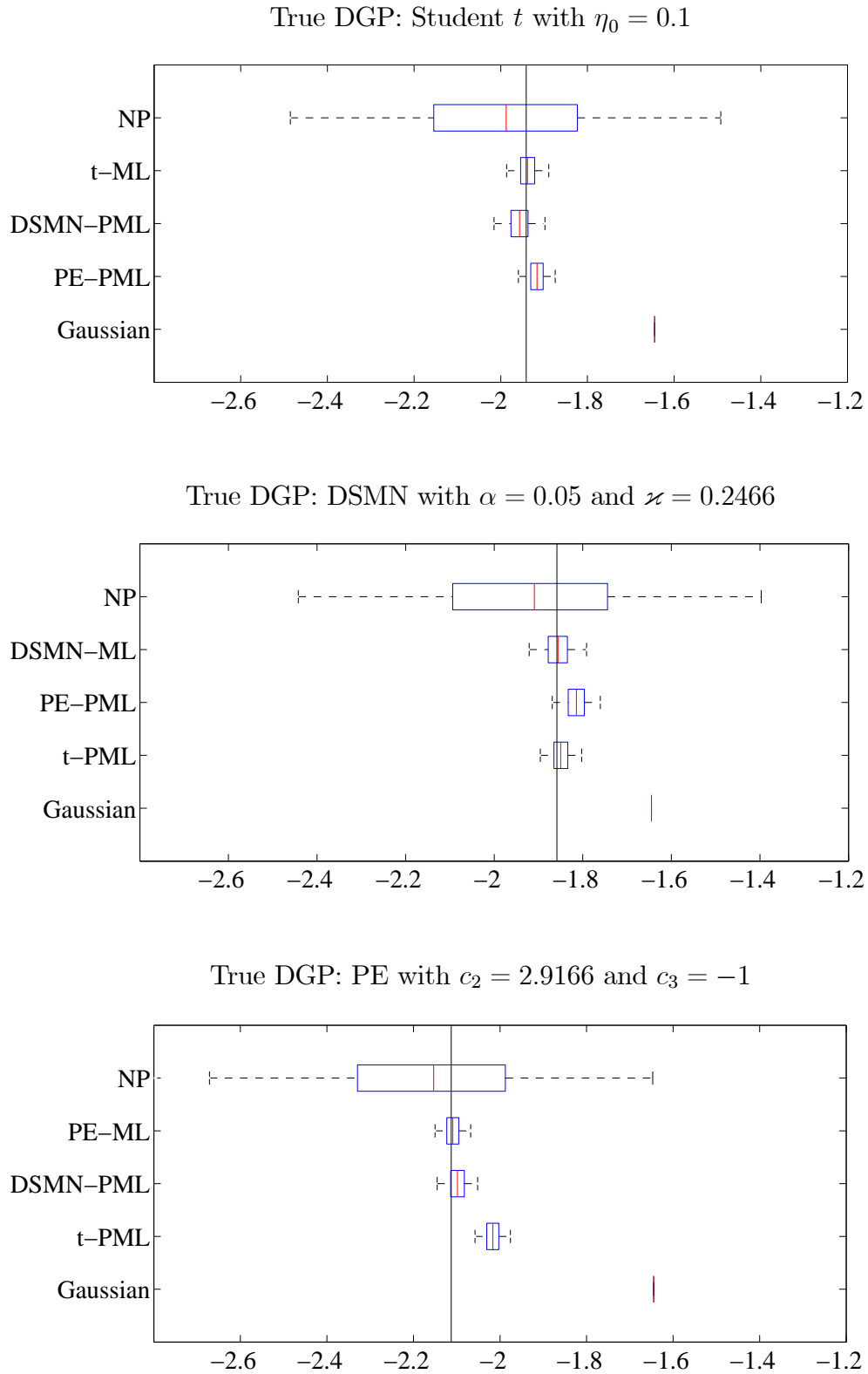
Notes: $c_3 = -c_2/3$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2nd and 3rd Laguerre polynomials with parameter $N/2 - 1$. Confidence intervals are computed using robust standard errors for a hypothetical sample size of $T = 1,000$ and $N = 5$. SML refers to sequential ML, NP refers to the fully nonparametric procedure based on the λ th empirical quantile of the standardised return distribution, while SNP denotes the nonparametric procedure that imposes symmetry of the return distribution (see Section 4.3 for details). The green solid line is the true VaR.

Figure 9a: Monte Carlo distributions of 99% VaR estimators



Notes: 1,600 replications, $T = 2,500$, $N = 5$. The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. ML (PML) means (pseudo) maximum likelihood estimator, NP nonparametric estimator. Vertical lines represent the true values. See Section 5.1 and Appendix D.2 for a detailed description of the Monte Carlo study.

Figure 9b: Monte Carlo distributions of 95% CoVaR estimators



Notes: 1,600 replications, $T = 2,500$, $N = 5$. The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. ML (PML) means (pseudo) maximum likelihood estimator, NP nonparametric estimator. Vertical lines represent the true values. See Section 5.1 and Appendix D.2 for a detailed description of the Monte Carlo study.

Copyright © 2012
D. Amengual, G. Fiorentini,
E. Sentana