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
Revised: December 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 their efficiency, and obtain moment conditions leading to sequential estimators as efficient as their joint ML counterparts. We also obtain standard errors for VaR and CoVaR, and analyse the effects on these measures of distributional misspecification. Finally, we illustrate the small sample performance of these procedures through simulations and apply them to analyse the risk of large eurozone banks.

Keywords: Confidence Intervals, Elliptical Distributions, Efficient Estimation, Global Systematically Important Banks, Systemic risk, Risk Management.

JEL: C13, C32, G01, G11

*We would like to thank Manuel Arellano, Christian Bontemps, Antonio Díez de los Ríos, Olivier Faugeras, Javier Mencía, Francisco Peñaranda, Marcos Sanso, David Veredas and audiences at the Bank of Canada, CEMFI, Chicago Booth, CREST, ECARES ULB, Koç, Princeton, Rimini, Toulouse, the Finance Forum (Granada, 2011), the Symposium of the Spanish Economic Association (Málaga, 2011) and the Conference in honour of M. Hashem Pesaran (Cambridge, 2011) for useful comments and suggestions. We also thank the editors and two anonymous referees for valuable feedback. Luca Repetto provided able research assistance for the empirical application. 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

Both 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) or Owen and Rabinovitch (1983)). Moreover, they generalise the multivariate normal distribution but retain its analytical tractability irrespective of the number of assets. However, non-Gaussian ML estimators often achieve efficiency gains under correct specification at the risk of returning inconsistent parameter estimators under distributional misspecification (see 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. 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 Pseudo ML estimators of the mean and variance parameters as first step estimators offer an attractive compromise because they preserve the consistency of the fist two conditional moments under distributional misspecification as long as those moments are correctly specified and the fourth moments are bounded (see Bollerslev and Wooldridge (1992)), while allowing for more realistic conditional distributions. From a more practical point of view, they also simplify the computations by reducing the dimensionality of the optimisation problem at each stage, thereby increasing the researcher's confidence that she has not found a local minimum. In this regard, it is worth bearing in mind that most commercially available econometric packages have been fine tuned to the Gaussian case, which even leads to closed-form estimators in commonly used models.

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.

To keep the exposition simple we focus on elliptical distributions in the text, and relegate more general cases to the supplemental appendix. We illustrate our results with several examples that nest the normal, including the Student t and 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 asymptotically valid standard errors of sequential estimators, assess their efficiency, and obtain the optimal moment conditions that lead to sequential MM estimators as efficient as their joint ML counterparts. Although we consider multivariate conditionally heteroskedastic dynamic regression models, our results 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 sequential ML estimators of flexible parametric families of distributions offer substantial efficiency gains, while incurring in small biases.

Finally, we illustrate our results with data for four Global Systematically Important Banks from the eurozone. As expected, we find that their stock returns display considerable non-normality even after controlling for time-varying volatilities and correlations, which in turn gives rise to the type of non-linear dependence that is relevant for systemic risk measurement.

The rest of the paper is as follows. In section 2, we describe the model, present the elliptical distributions we use as examples and introduce a convenient reparametrisation satisfied by most static and dynamic models. Then, in section 3 we discuss the sequential ML and GMM estimators, and compare their efficiency. In section 4, we study the effect 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, and the empirical application in section 6. Finally, we present our conclusions in section 7. 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 a parametric dynamic regression model with time-varying variances and covariances. Typically, the N dependent variables, $\mathbf{y}_t$, are assumed to be generated as:

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

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

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

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 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 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 $\mathbf{u}_t$. 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 use $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$, 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

$$\kappa = E(\varsigma_t^2)/[N(N+2)] - 1 \tag{2}$$

will also be bounded. The most prominent examples are the standardised multivariate Student t, in which ς_t is proportional to an F random variable with N and ν degrees of freedom, and the limiting Gaussian case, when ς_t becomes a χ^2_N. Since this involves no additional parameters, we identify the normal distribution with $\boldsymbol{\eta}_0 = \mathbf{0}$, while for the Student t we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions. Normality is thus achieved as $\eta \to 0$ (see Fiorentini, Sentana and Calzolari (2003)). Other more flexible families of spherical distributions that we will also use to illustrate our general results are:

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

$$\varsigma_t = [s_t + (1-s_t)\varkappa]/[\alpha + (1-\alpha)\varkappa] \cdot \zeta_t$$

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, ς_t will be a two-component scale mixture of $\chi^{2\prime}_N s$, with shape parameters α and $\varkappa$. Like all scale mixture of normals (including the Student t), this distribution is necessarily leptokurtic but approaches the multivariate normal when $\varkappa \to 1$, $\alpha \to 1$ or $\alpha \to 0$, although near those limits the distributions can be rather different (see Amengual and Sentana (2011) for further details).[1]

Polynomial expansion: $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ 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 $h_o(\varsigma_t)$ denotes the density function of a χ^2 with N degrees of freedom, and

$$P_J(\varsigma_t) = 1 + \sum\nolimits_{j=2}^{J} c_j p^g_{N/2-1,j}(\varsigma_t)$$

is a J^{th} order polynomial written in terms of the generalised Laguerre polynomial of order j and parameter $N/2-1$, $p^g_{N/2-1,j}(.)$ (see Appendix C for some detailed expressions). As a result, the $J-1$ shape parameters will be given by $c_2, c_3, \ldots, c_J$. 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

[1] Multiple component discrete scale mixtures of normals would be tedious but straightforward to deal with. As is well known, they 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.

become negative. For that reason, in Appendix D.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 $\varsigma \geq 0$. PE reduce to the normal when $c_j = 0$ for all j, and 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.

In Figure F1 in the supplemental appendix we plot the densities of a normal, a Student t, 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 2 plots the so-called exceedance correlation (see Longin and Solnik, 2001) for those uncorrelated marginal components. 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.

2.3 A convenient reparametrisation

Throughout this paper we assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied because we want to leave unspecified the conditional mean vector and covariance matrix to maintain full generality.[2] But for the sake of brevity in the main text we focus in the class of models for which the following reparametrisation is admissible:

Reparametrisation 1 *A homeomorphic transformation $\mathbf{r}(.) = [\mathbf{r}_1'(.), r_2'(.)]'$ 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 rank$[\partial \mathbf{r}'(\boldsymbol{\theta})/\partial \boldsymbol{\theta}] = p$ in a neighbourhood of $\boldsymbol{\theta}_0$, such that*

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_1), \quad \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_2 \boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1) \quad \forall t, \tag{3}$$

with

$$E[\ln |\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)| | \boldsymbol{\phi}_0] = k \quad \forall \boldsymbol{\vartheta}_1. \tag{4}$$

Expression (3) simply requires that one can construct pseudo-standardised residuals

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

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. The 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

[2]Primitive conditions for specific multivariate models can be found for instance in Ling and McAleer (2003).

case, the information matrix will be block diagonal between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, while in the second case the general expressions we provide in Appendix B apply.

Given that we can multiply ϑ_2 by some scalar positive smooth function of $\boldsymbol{\vartheta}_1$, $k(\boldsymbol{\vartheta}_1)$ say, and divide $\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_1)$ by the same function without violating (3), condition (4) simply provides a particularly convenient normalisation.

As we shall see, it turns out that under reparametrisation 1 the asymptotic dependence between estimators of the conditional mean and variance parameters and estimators of the shape parameters is generally driven by a scalar parameter. As a result, the asymptotic variances of the estimators of $\boldsymbol{\eta}$ we consider next will not depend on the functional form of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ or $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$.[3]

3 Sequential estimators of the shape parameters

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

Let $L_T(\boldsymbol{\phi})$ denote the sample log-likelihood function of a sample of size T, so that $\hat{\boldsymbol{\phi}}_T = \arg\max_{\boldsymbol{\phi}} L_T(\boldsymbol{\phi})$ is the joint ML estimator of $\boldsymbol{\phi}' = (\boldsymbol{\theta}', \boldsymbol{\eta}')$ and $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\eta} L_T(\tilde{\boldsymbol{\theta}}_T, \mathbf{0})$ the Gaussian pseudo MLE of $\boldsymbol{\theta}$. We can use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of $\boldsymbol{\eta}$ as $\tilde{\boldsymbol{\eta}}_T = \arg\max_{\eta} L_T(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$.[4] Interestingly, these sequential ML estimators 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

$$h(\varsigma_t; \boldsymbol{\eta}) = \pi^{N/2}/\Gamma(N/2) \cdot \varsigma_t^{N/2-1} \exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})], \tag{5}$$

where $g(\varsigma_t, \boldsymbol{\eta})$ is the kernel and $c(\boldsymbol{\eta})$ the constant of integration of the (log) density of $\boldsymbol{\varepsilon}_t^*$ (see expression (2.21) in Fang, Kotz and Ng (1990)). Thus, we could obtain the infeasible ML estimator of $\boldsymbol{\eta}$ by maximising the log-likelihood function of the observed $\varsigma_t(\boldsymbol{\theta}_0)'s$, $\sum_{t=1}^T \ln h\left[\varsigma_t(\boldsymbol{\theta}_0); \boldsymbol{\eta}\right]$. Although in practice the standardised residuals are usually unobservable, it is easy to prove from (5) 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 1 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$ and reparametrisation (1) is admissible, then the asymptotic variance of the sequential ML estimator of $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\eta}}_T$, is*

$$\mathcal{F}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta}_0) + \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}'(\boldsymbol{\eta}_0)\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta}_0) \cdot [N/(2\vartheta_{20})]^2 \mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_0), \tag{6}$$

[3] Bickel (1982) exploited parametrisation (1) 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 (3) 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 misspecification of the elliptical distribution.

[4] Often there will be inequality constraints on $\boldsymbol{\eta}$, but we postpone the details to Appendix D.1.

where $\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ denotes the information matrix, $\mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ the asymptotic variance of the PML estimator of ϑ_2 given in (A4), $\mathrm{M}_{sr}(\boldsymbol{\eta}) = -E\left\{N^{-1}\varsigma_t(\boldsymbol{\theta})\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\boldsymbol{\eta}'\middle|\,\boldsymbol{\phi}\right\}$ and $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma$, while the asymptotic variance of the feasible ML estimator of $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}_T$, is

$$\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) = \mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\eta}_0) + \mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\eta}_0)\mathrm{M}'_{sr}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\eta}_0)\cdot[N/(2\vartheta_{20})]^2\mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0), \tag{7}$$

where $\mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0)$ is the asymptotic variance of the feasible ML estimator of ϑ_2 given in (A5).

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 (6) nor (7) effectively depend on ϑ_2, which drops out from those formulas.

It is easy to see from (6) and (7) that $\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \leq \mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \leq \mathcal{F}(\boldsymbol{\phi}_0)$ regardless of the distribution, with equality between $\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ and $\mathcal{F}(\boldsymbol{\phi}_0)$ if and only if $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = \mathbf{0}$, 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.

A more interesting question in practice is the relationship between $\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ and $\mathcal{F}(\boldsymbol{\phi}_0)$. The following result gives us the answer by exploiting Theorem 5 in Pagan (1986):

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 reparametrisation (1) is admissible, then $\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \leq \mathcal{F}(\boldsymbol{\phi}_0)$, with equality if and only if*

$$\mathrm{M}'_{sr}(\boldsymbol{\eta}_0)\left[\mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0) - \mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi})\right]\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = 0.$$

Hence, the scalar nature of ϑ_2 implies that the only case in which $\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) = \mathcal{F}(\boldsymbol{\phi}_0)$ with $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) \neq \mathbf{0}$ will arise when the Gaussian PMLE of ϑ_2 is as efficient as the joint ML.[5]

Finally, note that since the asymptotic variance of the Gaussian PML estimator of $\boldsymbol{\theta}$ will become unbounded as $\kappa_0 \to \infty$, if $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) \neq \mathbf{0}$ the asymptotic distribution of $\tilde{\boldsymbol{\eta}}_T$ will also be non-standard in that case, unlike that of the joint ML estimator $\hat{\boldsymbol{\eta}}_T$.

3.2 Sequential GMM estimators of $\boldsymbol{\eta}$

If we can compute the expectations of $L \geq q$ functions of ς_t, $\boldsymbol{\upsilon}(.)$ say, then we can also compute a sequential GMM estimator of $\boldsymbol{\eta}$ by minimising the quadratic form $\bar{\mathbf{n}}'_T(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})\boldsymbol{\Omega}\bar{\mathbf{n}}_T(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$, where $\boldsymbol{\Omega}$ is a positive definite weighting matrix, and $\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}) = \boldsymbol{\upsilon}[\varsigma_t(\boldsymbol{\theta})] - E\{\boldsymbol{\upsilon}[\varsigma_t(\boldsymbol{\theta})]|\boldsymbol{\phi}\}$. 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

[5] The original Kotz (1975) distribution provides an example in which $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = 0$ *and* $\mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0) = \mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0)$.

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 $\boldsymbol{\phi}$. 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, 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, $\tilde{\boldsymbol{\theta}}_T$, using a standard first-order expansion. In those cases in which reparametrisation (1) is admissible, a much simpler equivalent procedure is as follows:[6]

Proposition 3 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$ and reparametrisation (1) is admissible, then the optimal sequential GMM estimator of $\boldsymbol{\eta}$ based on $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$ will be asymptotically equivalent to the optimal sequential GMM estimator based on $\mathbf{n}_t^{\circ}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$, where*

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

with

$$\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi}) = cov\left[\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}), \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N | \, \boldsymbol{\phi}\right],$$

are the residuals from the theoretical IV regression of $\mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta})$ on $\varsigma_t(\boldsymbol{\theta})/N - 1$ using as instrument $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N - 1$.

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.

3.2.1 Higher order moments and orthogonal polynomials

It seems natural to use powers of ς_t to estimate $\boldsymbol{\eta}$. Specifically, we can consider:

$$\ell_{mt}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \varsigma_t^m(\boldsymbol{\theta}) / \left[2^m \prod\nolimits_{j=1}^{m}(N/2 + j - 1)\right] - [1 + \tau_m(\boldsymbol{\eta})], \tag{8}$$

where $\tau_m(\boldsymbol{\eta})$ are the higher order moment parameter of spherical random variables introduced by Berkane and Bentler (1986) (see also Maruyama and Seo (2003)).[7] But given that for $m = 1$, expression (8) reduces to $\ell_{1t}(\boldsymbol{\theta}) = \varsigma_t(\boldsymbol{\theta})/N - 1$ irrespective of $\boldsymbol{\eta}$, we have to start with $m \geq 2$.

[6] See Bontemps and Meddahi (2012) for alternative approaches in moment-based specification testing.

[7] We derive expressions for $\tau_m(\boldsymbol{\eta})$ for our examples of elliptical distributions in Appendix D.2. A noteworthy property of those examples is that their moments are always bounded, with the exception of the Student t. Appendix D.3 contains the moment generating functions for the DSMN and the 3^{rd}-order PE.

An alternative is to consider influence functions defined by the relevant m^{th} order orthogonal polynomial $p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \sum_{h=0}^{m} a_h(\boldsymbol{\eta})\varsigma_t^h(\boldsymbol{\theta})$.[8] Again, we have to consider $m \geq 2$ because the first two non-normalised polynomials are always $p_0(\varsigma_t) = 1$ and $p_1(\varsigma_t) = \ell_{1t}(\boldsymbol{\theta})$ for all $\boldsymbol{\eta}$.

Given that $\{p_1[\varsigma_t(\boldsymbol{\theta})], p_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], ..., p_M[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}$ is a full-rank linear transformation of $[\ell_{1t}(\boldsymbol{\theta}), \ell_{2t}(\boldsymbol{\theta},\boldsymbol{\eta}), ..., \ell_{Mt}(\boldsymbol{\theta},\boldsymbol{\eta})]$, 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 4 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with $E[\varsigma_t^{2M}|\boldsymbol{\eta}_0] < \infty$ and reparametrisation (1) is admissible, then the optimal sequential estimator of $\boldsymbol{\eta}$ based on $\mathbf{p}'[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \{p_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], ..., p_M[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}$ and $\ell_t'(\boldsymbol{\theta},\boldsymbol{\eta}) = [\ell_{2t}(\boldsymbol{\theta},\boldsymbol{\eta}), ..., \ell_{Mt}(\boldsymbol{\theta},\boldsymbol{\eta})]$ are asymptotically equivalent, with an asymptotic variance that reflects the sample uncertainty in $\tilde{\boldsymbol{\theta}}_T$ given by*

$$\mathcal{J}_M(\boldsymbol{\phi}_0) = \left(\mathcal{H}_{\mathbf{p}}'(\boldsymbol{\phi}_0)\left[\mathcal{G}_{\mathbf{p}}(\boldsymbol{\phi}_0) + \{(N/2) + [N(N+2)\kappa_0/4]\}\Bbbk_{\mathbf{p}}(\boldsymbol{\phi}_0)\Bbbk_{\mathbf{p}}(\boldsymbol{\phi}_0)'\right]^{-1}\mathcal{H}_{\mathbf{p}}(\boldsymbol{\phi}_0)\right)^{-1},$$

where $\mathcal{H}_{\mathbf{p}}(\boldsymbol{\phi})$ is an $(M-1)\times q$ matrix with representative row $E\{p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]s_{\boldsymbol{\eta}t}'(\boldsymbol{\phi})]|\boldsymbol{\phi}\}$ and $\mathcal{G}_{\mathbf{p}}(\boldsymbol{\phi})$ is a diagonal matrix of order $M-1$ with representative element $V\{p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}$.

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, although the precision with which we can estimate $\tau_m(\boldsymbol{\eta})$ rapidly decreases with m.

3.2.2 Efficient sequential GMM estimators of $\boldsymbol{\eta}$

Our previous GMM optimality discussion applies to a fixed set of moments involving powers of ς_t. But there are many other alternative estimating functions that one could use, including the rational functions advocated by Bontemps and Meddahi (2012) for testing the univariate Student t or (smoothed versions of) the check functions used in quantile estimation (see Koenker (2005)), which are well defined even if the higher order moments are unbounded (see Dominicy and Veredas (2010) for a closely related approach). Therefore, it seems relevant to ask which estimating functions would lead to the most efficient sequential estimators of $\boldsymbol{\eta}$ taking into account the sampling variability in $\tilde{\boldsymbol{\theta}}_T$. The following result answers this question by exploiting the characterisation of efficient sequential estimators in Newey and Powell (1998):

Proposition 5 *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 reparametrisation (1) is admissible, then the efficient influence function is given by the efficient parametric score of $\boldsymbol{\eta}$:*

$$\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) - [(1+2/N)\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]^{-1}\mathrm{M}_{sr}'(\boldsymbol{\eta})\left[\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N - 1\right], \quad (9)$$

[8] Appendix C contains the expressions for the coefficients of the second and third order orthogonal polynomials of the different examples we consider.

which is the residual from the theoretical regression of $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$ *on* $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N-1$.

Importantly, the resulting sequential MM estimator of $\boldsymbol{\eta}$ will achieve the efficiency of the feasible ML estimator, which is the largest possible, because (i) the variance of the efficient parametric score $\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ in (9) coincides with $\mathcal{I}^{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ in (7); and (ii) $\mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)$ is also the expected value of the Jacobian matrix of (9) with respect to $\boldsymbol{\eta}$.

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 with the multivariate Student t. Given that when reparametrisation (1) is admissible Proposition 4 implies the asymptotic equivalence between the sequential MM estimators of η based on the fourth moment and the second order polynomial, the following proposition compares the efficiency of these estimators to the sequential ML estimator of η:

Proposition 6 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ *is i.i.d.* $t(\mathbf{0},\mathbf{I}_N,\nu_0)$ *with* $\nu_0>8$*, then* $\mathcal{F}(\boldsymbol{\phi}_0)\leq\mathcal{J}_2(\boldsymbol{\phi}_0)$.

This proposition shows that sequential ML is always more efficient than sequential MM based on the second order polynomial. Nevertheless, Proposition 5 implies that there is a sequential MM procedure that is more efficient than sequential ML.

Given that $\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)=\mathbf{0}$ under normality from Proposition E1, it is clear that, asymptotically, $\tilde{\eta}_T$ will be as 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 distribution of the sequential MM estimator $\mathring{\eta}_T$ will also tend to be half normal as the sample size increases when $\eta_0=0$, since $\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)$ is root-T consistent for κ, which is 0 in the Gaussian case. In fact, $\mathring{\eta}_T$ will be as efficient as $\hat{\eta}_T$ under normality because $p_2[\varsigma_t(\boldsymbol{\theta}),\eta]$ is proportional to $s_{\eta t}(\boldsymbol{\theta}_0,0)$. In contrast, $\mathring{\eta}_T$ will not be root-T consistent when $4\leq\nu_0\leq 8$ because $\mathcal{J}_2(\boldsymbol{\phi}_0)$ will diverge to infinity as ν_0 converges to 8 from above. Moreover, since κ is infinite for $2<\nu_0\leq 4$, $\mathring{\eta}_T$ will not even be consistent in the interior of this range.

3.3.2 Asymptotic standard errors and relative efficiency

Under the maintained assumption that reparametrisation (1) is admissible, which covers most static and dynamic models, we have used the results in Propositions 1 and 4 to compute the asymptotic standard deviations and relative efficiency of the joint MLE and efficient sequential MM estimator, the sequential MLE, and finally the sequential GMM estimators based on orthogonal polynomials.

In the case of the Student t distribution, all estimators behave similarly for slight departures from normality ($\eta < .02$ or $\nu > 50$). As η increases, the GMM estimators become relatively less efficient, with the exactly identified GMM estimator being the least efficient, as expected from Proposition 6. 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.

Since the DSMN distribution has two shape parameters, we consider the two following exercises: first, we maintain the scale ratio parameter $\varkappa$ equal to .5 and report the asymptotic efficiency as a function of the mixing probability parameter α; secondly, we look at the asymptotic efficiency of the different estimators fixing the mixing probability at $\alpha = .05$. 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. 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.

The same general pattern emerges in the case of the PE distribution for which we also consider two situations, maintaining one of the parameters fixed to 0 while reporting the asymptotic efficiency as a function of the remaining parameter. Again sequential MLE shows 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 c_3 when $c_2 = 0$.

For more detailed results, see Figures F2 to F4 in the supplemental appendix, which display the asymptotic standard deviation (top panels) and the relative efficiency (bottom panels).

3.4 Misspecification analysis

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 a situation 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 7 *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 $\boldsymbol{\vartheta}$ 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*

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

where $\boldsymbol{\phi}_\infty = (\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty)$, $\boldsymbol{\eta}_\infty$ solves $E[\,\mathbf{e}_{rt}(\boldsymbol{\vartheta}_0, \boldsymbol{\eta}_\infty)|\,\boldsymbol{\varphi}_0] = \mathbf{0}$, $\mathcal{H}_{rr}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E[\,\partial \mathbf{e}_{rt}(\boldsymbol{\phi})/\partial \boldsymbol{\eta}'|\,\boldsymbol{\varphi}]$,

$$\mathcal{E}_r(\boldsymbol{\phi}; \boldsymbol{\varphi}) = \mathcal{O}_{rr}^{-1}(\boldsymbol{\phi}; \boldsymbol{\varphi}) + (N/4)[2(\kappa+1) + N\kappa]\mathcal{O}_{rr}^{-1}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathrm{M}_{sr}^{O\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathrm{M}_{sr}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathcal{O}_{rr}^{-1}(\boldsymbol{\phi}; \boldsymbol{\varphi}),$$

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

In section 4.3 we will use this result to obtain robust standard errors.

4 Application to risk measures

Most institutional investors 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.[9]

[9] Acharya et al. (2010) and Brownlees and Engle (2011) consider instead the Marginal Expected Shortfall, defined 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.

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

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

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \left[\begin{pmatrix} \mu_{Mt} \\ \mathbf{a}_t(\boldsymbol{\theta}) + \mathbf{b}_t(\boldsymbol{\theta})\mu_{Mt} \end{pmatrix}\right], \ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2_{Mt} & \sigma_{Mt}\mathbf{b}'_t(\boldsymbol{\theta}) \\ \sigma_{Mt}\mathbf{b}_t(\boldsymbol{\theta}) & \sigma^2_{Mt}\mathbf{b}_t(\boldsymbol{\theta})\mathbf{b}'_t(\boldsymbol{\theta}) + \boldsymbol{\Omega}_t(\boldsymbol{\theta}) \end{bmatrix} \quad (10)$$

and $\sigma^2_{Mt} = \sigma^2_M + \gamma(\varepsilon^2_{Mt-1} - \sigma^2_M) + \beta(\sigma^2_{Mt-1} - \sigma^2_M)$. In this model, μ_{Mt} and σ^2_{Mt} denote the conditional mean and variance of r_{Mt}, while $\mathbf{a}_t(\boldsymbol{\theta})$ and $\mathbf{b}_t(\boldsymbol{\theta})$ are respectively the alpha and beta of the other $N-1$ assets with respect to the market portfolio and $\boldsymbol{\Omega}_t(\boldsymbol{\theta})$ 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.[10]

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 $\mathbf{w}_t = (w_{Mt}, w_{2t}, ...w_{Nt})'$ 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

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

where $\mu_{wt} = \mathbf{w}'_t\boldsymbol{\mu}_t$ and $\sigma^2_{wt} = \mathbf{w}'_t\boldsymbol{\Sigma}_t\mathbf{w}_t$ are the expected excess return and variance of r_{wt}, and $F(.)$ is the cumulative distribution function of a zero mean - unit variance random variable within the appropriate elliptical class.[11]

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

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

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

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

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 $r_{ht} = r_{wt} - [cov_{t-1}(r_{wt}, r_{Mt})V_{t-1}^{-1}(r_{Mt})]r_{Mt}$ 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.[12] More formally,

$$\begin{aligned}
\lambda_2 &= \Pr\left[\varepsilon_{ht}^* \leq q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})\,|\varepsilon_{Mt}^* \leq q_1(\lambda_1, \boldsymbol{\eta})\right] \\
&= \int_{-\infty}^{q_1(\lambda_1, \boldsymbol{\eta})} f_1(\varepsilon_{1t}^*, \boldsymbol{\eta})\left[\int_{-\infty}^{q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})} f_{2|1}(\varepsilon_{2t}^*, \varepsilon_{1t}^*; \boldsymbol{\eta}) d\varepsilon_{2t}^*\right] d\varepsilon_{1t}^*, \\
q_1(\lambda, \boldsymbol{\eta}) &= \frac{1}{\sigma_{Mt}}\left[1 - R_0 - \mu_{Mt} - \frac{V_t}{w_M W_{t-1}}\right].
\end{aligned}$$

In Appendix D.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 means, standard deviations, correlations and quantiles. Given that our main interest lies in the sequential estimators of the shape parameters, in the rest of this section we shall focus on the sampling variability in estimating $q_1(\lambda, \boldsymbol{\eta})$ and $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$.

[12] Adrian and Brunnermeier (2011) condition instead on the market return r_{Mt} being *at* its λ_1^{th} quantile.

In parametric models, these quantiles would be known with certainty for all values of λ 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 D.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 3 displays confidence bands for parametric VaR and CoVaR computed with the Student t (3a-b), DSMN (3c-d) and PE (3e-f) distributions. To save space, we only look at the 1% and 5% significance levels for the case in which λ_1=λ_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 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

$$\lambda(1-\lambda)/f^2\left[q_1(\lambda)\right], \tag{11}$$

where $f(.)$ 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. 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. It is easy to prove that the asymptotic variance of this quantile estimator will be

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

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

$$\lambda(1-\lambda)/f^2\left[q_1\left(\lambda,\boldsymbol{\eta}\right);\boldsymbol{\eta}\right]E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\varepsilon_{1t}^*\leq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}\right]V(\tilde{\boldsymbol{\eta}}_T|\boldsymbol{\eta})E\left[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\varepsilon_{1t}^*\geq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}\right] \quad (12)$$

which coincides with (11) 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 $V\left[\tilde{\boldsymbol{\eta}}_T|\boldsymbol{\varphi}_0\right]$. Specifically, 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 $\boldsymbol{\theta}$ 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 (12), but with $\boldsymbol{\eta}_0$ replaced by the pseudo-true value of $\boldsymbol{\eta}$ defined in Proposition 7, $\boldsymbol{\eta}_\infty$.

The left panels of Figure 4 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 right 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 (10) calibrated to the empirical application in section 6. Specifically, we simulate and estimate a model in which $N=5$, $\mu_M=0.07/52$, $\sigma_M=.24/\sqrt{52}$, $\mathbf{a}=\mathbf{0}$, $\mathbf{b}=(1.2,1.2,1,1)$, $vecd(\boldsymbol{\Omega})=(6,12,24,48)$, $\gamma=0.1$ 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 a 3^{rd}-order PE also with the same kurtosis and $c_3=-1$. Finally, we also simulate data from a spherical distribution whose generating variable e_t is independently drawn from the empirical distribution function of $\varsigma_t(\boldsymbol{\theta})$ evaluated at the Gaussian PML estimates obtained from the eurozone bank data described in section 6. The computational advantages of the sequential estimators are particularly noticeable for model (10), which under normality can be estimated by means of four linear regressions and a single univariate GARCH model. Although we have considered other sample sizes, for the sake of brevity we only report the results for $T = 1,000$ observations (plus another 100 for initialisation) based on 1,600 Monte Carlo replications. This sample size corresponds roughly to 4 years of daily data or 20 years of weekly data. The numerical strategy employed by our estimation procedure is described in Appendix E.3. 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 four different estimators of the shape parameters under correct specification, as well as (the square root of) the mean across simulations of the estimates of their asymptotic variances. Specifically, we consider joint ML (ML), sequential ML (SML), efficient sequential MM (ESMM), and orthogonal polynomial-based MM (SMM) estimators that use the 2^{nd} polynomial in the case of the Student t, and the 2^{nd} and 3^{rd} 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 3^{rd}-order PE, respectively.

The behavior of the different estimators is in line with the results in Section 3.4. The standard deviations of ESMM and SML essentially coincide, as expected from Figures F2-F4. 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 F2-F4. 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 worth noting is that the estimators of the DSMN parameters α and $\varkappa$ seem to be slightly upward biased, and that the bias increases when 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.

Finally, the sample analogues of the asymptotic variance covariance matrices are in general reliable, which probably reflects the fact that we use the theoretical expressions in section 3.

Specifically, the mean across simulations of the asymptotic variance estimates are very close to the Monte Carlo variances of the estimators, with the exception of the SMM estimator, for which they tend to overestimate the sampling variablility of the shape parameters.

5.3 Sampling distribution of VaR and CoVaR measures

We used the ML and SML estimators of the shape parameters to compute parametric VaR and CoVaR measures using the conditional and marginal CDFs in Appendix D.4. 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.

The left panels of Figure 5 summarise the sampling distribution of the different estimates of $q_1(\lambda_1, \boldsymbol{\eta})$ for $\lambda_1 = .99$ by means of box-plots for the different DGPs. 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 seven rows with the true joint ML and three SML-based measures, as well as the two non-parametric ones (denoted by NP and SNP) 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 quantile is pronounced.

The same general pattern emerges in the right panels of Figure 5, which compares the different estimates of $q_{2|1}(\lambda_2, \lambda_1, \boldsymbol{\eta})$ for $\lambda_2 = \lambda_1 = .95$. For the distributions we use as examples, the effects of distributional misspecification seem to be 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.

Finally, the results in Figures 5g-h, which are based on data generated from the empirical distribution of the eurozone banks in section 6, indicate that the parametric procedures based on the Student t distribution and the DSMN provide rather accurate estimates of the "true"

[13] Alternatively, 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)).

VaR and CoVaR, which we compute by using a single path simulation of size 5 million.

6 Empirical application to G-SIBs eurozone banks

The Financial Stability Board (FSB) has recently updated its list of globally systematically important banks (G-SIBs), allocating them to four buckets corresponding to their required level of common equity as a percentage of risk-weighted assets on top of the 7% baseline in the Basel III Accord.[14] Despite the lack of a formal definition, G-SIBS are deemed fundamental players in any future global financial crisis. Given that the ongoing negative feedback loop between banks and weak sovereigns in several peripheral euro area countries might end up triggering such a crisis, it seems particularly relevant to illustrate our procedures with some eurozone G-SIBS. Specifically, we look at the flagship commercial banks from Germany (Deutsche Bank), France (BNP Paribas), Spain (Banco Santander) and Italy (Unicredit Group). Interestingly, the FSB has classified Deutsche and BNP Paribas in the fourth and third buckets (2.5% and 2% capital surcharges, respectively), but Santander and Unicredit in the first one (1% surcharge), in marked contrast with the credit ratings of the sovereign debt of their countries of origin.

We use a capitalisation weighted total return index of the 80 most important commercial banks domiciled in the eurozone as representative of the banking sector in the European Monetary Union. We also adopt the perspective of a German investor, and convert all the different stock indices to D-Marks prior to January 1st, 1999, when the euro became the official numeraire.[15] Figure 6a shows the recent evolution of the total return indices for each of the four aforementioned banks and the whole sector normalised to 100 at the end of 2006 to facilitate comparisons. The temporal pattern of these price series through the different phases of the 2007-09 global credit crisis is fairly homogeneous, and the same is by and large true during the European sovereign debt crisis that started in 2010 when investors shifted their attention to the size of the fiscal imbalances in Greece. As we shall see below, though, there are important

[14] The new regulation has introduced a 2.5% mandatory capital conservation buffer in addition to a minimum common equity requirement of 4.5%; see Basle Committe on Banking Supervision (2011) for further details. There will also be a countercyclical buffer imposed within a range of 0-2.5%.

[15] The Datastream codes of the total return indices used are D:DBKX(RI) (Deutsche Bank), F:BNP(RI) (BNP Paribas), E:SCH(RI) (Banco Santander), I:UCG(RI) (Unicredit) and finally BANKSEM(RI) for the EMU commercial bank index. The first four series are reported in local currency, while the last one is denominated in US \$. We then convert them to DM/Euro by crossing the relevant exchange rates against the British pound (DMARKER, FRENFRA, ITALIRE, SPANPES and USDOLLR). We define weekly returns as Wednesday to Wednesday log index changes in order to minimise the incidence of filled forward prices due to public holidays and other gaps. Finally, we work with excess returns by subtracting the continuously compounded rate of return on the one-week Eurocurrency rate in DM/Euros (ECWGM1W). Our final balanced panel includes 984 observations from the second half of October 1993 to the end of August 2012.

differences across institutions from a risk perspective.

But first, in Figure 6b we compare the one-week ahead 99% Value at Risk estimates (in percentage terms) for the eurozone banking portfolio that the different estimation procedures previously discussed generate. Despite the massive rejection of the multivariate normality assumption using the LM test based on the second order Laguerre polynomial put forward by Fiorentini, Sentana and Calzolari (1983), the effect of using a non-normal distribution seems relatively minor, although the Gaussian values are systematically lower than the rest (see Tables F1 and F2 in the supplemental appendix for parameter estimates and the quantiles that they imply). The only other difference worth mentioning is the fact that the non-Gaussian MLEs of the Arch (Garch) parameter γ (β) tend to be lower (higher) than the corresponding Gaussian PMLEs. As a result, the VaR spikes that the joint estimators generate are somewhat lower but last a bit longer than the ones obtained with the sequential estimators. In order to increase the realism of our model, we have considered a generalised version of (10) in which we allow both systematic and idiosyncratic variances to evolve over time as Gqarch(1,1) processes (see Sentana (1995)), and do not impose the CAPM restrictions on the intercepts.

Figures 6c-6f depict the different estimates of the one-week ahead specific exposure CoVaR (in percentage terms) at the 5% level of each of the four banks when the fall in the euro area bank index *exceeds* its 5^{th} percentile. Not surprisingly, the Gaussian CoVaR estimates are significantly lower than the rest. As in the case of the VaR figures, the differences between the non-Gaussian and Gaussian estimates of the Garch parameters are once again noticeable. But the most striking feature of those pictures is the marked heterogeneity across banks, which is patently visible regardless of distributional assumptions. Although all four institutions were affected in varying degrees by the turmoil in financial markets after the Lehman Brothers collapse, the effects of the European sovereign debt crisis is far more heterogeneous. While so far Deutsche Bank and Banco Santander have suffered relatively minor contagion effects from increases in the riskiness of the eurozone banking sector, BNP Paribas and especially Unicredit have been substantially more sensitive. This is particularly true in the second half of 2011, as the international alarm over the eurozone crisis grew, and the Spanish and Italian governments' borrowing costs rocketed. Although there is no reliable weekly data on the banks balance sheet structure, many commentators have attributed such differences to the extent financial institutions were stricken with sovereign debt from peripheral countries.

7 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 innovations distribution, which can be easily obtained from the standardised innovations evaluated at the Gaussian PML estimators. We consider both sequential ML estimators and 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. 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 asymptotically valid standard errors of sequential estimators, assess their efficiency and obtain the optimal moment conditions that lead to sequential MM estimators as efficient as their joint ML counterparts. 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 the required quantiles. Not surprisingly, our results indicate that the standard errors are larger for CoVaR than for VaR. 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.

Given that Gaussian PMLEs are sensitive to outliers, it seems relevant to explore other consistent but more "robust" estimators of the conditional mean and variance parameters. For example, when reparametrisation 1 is admissible, Fiorentini and Sentana (2010) suggest combining a likelihood-based estimator of $\boldsymbol{\vartheta}_1$, which remains consistent when the elliptical distribution has been misspecified, with a consistent closed-form estimator of the overall scale parameter ϑ_2.

Similarly, the sequential estimation approach that we have studied could be applied to models with non-spherical innovations, which would be particularly relevant from an empirical perspective given that tail dependence seems to be stronger for falls in prices than for increases. In principle, most of the theoretical results in sections 3 and 4 will survive (see e.g. Propositions B1, B2, B3 or B5), but in practice it might be necessary to focus on parsimonious multivariate distributions, such as the location-scale mixtures of normals in Mencía and Sentana (2009).

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). It would also be worth extending the tools used to evaluate value at risk models (see e.g. Lopez (1999) and the references therein) to cover systemic risk measures such as CoVar accounting for sampling variability in the estimation of both the conditioning set and the quantile of the relevant conditional distribution. All these topics constitute interesting avenues for future research.

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. (2011): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Basel Committee on Banking Supervision (2011): "Basel III: A global regulatory framework for more resilient banks and banking systems", mimeo, Bank of International Settlements.

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. (2012): "Testing distributional assumptions: a GMM approach", *Journal of Applied Econometrics* 27, 978-1012.

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.

Dominicy, Y. and Veredas, D. (2010): "The method of simulated quantiles", forthcoming in the *Journal of Econometrics*

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.

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.

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.

Lopez, J.A. (1999): "Methods for evaluating value-at-risk estimates", Federal Reserve Bank of San Francisco Economic Review 2, 3-17.

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

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

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

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 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.

Sentana, E. (1995): "Quadratic ARCH Models", *Review of Economic Studies* 62, 639-661.

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

Appendix

A Proofs

A.1 Preliminary results for reparameterisation 1

Given our assumptions on $\mathbf{r}(.)$, 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 the score vector for $\boldsymbol{\vartheta}$ will be

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ \mathbf{Z}_{\vartheta_2 s}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_1 lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\vartheta_2 s}(\boldsymbol{\vartheta}) \end{bmatrix} = \left\{ \begin{matrix} \vartheta_2^{-1/2}\left[\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_1)/\partial\boldsymbol{\vartheta}_1\right]\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1) \\ 0 \end{matrix}\right.$$
$$\left.\begin{matrix} \frac{1}{2}\{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_1)]/\partial\boldsymbol{\vartheta}_1\}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_1)] \\ \frac{1}{2}\vartheta_2^{-1}vec'(\mathbf{I}_N) \end{matrix}\right\}, \tag{A1}$$

with $\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ given in (E12) and (E13), respectively. As a result,

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

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is

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

where $\mathrm{M}_{ss}(\boldsymbol{\eta}) = E\left\{2[N(N+2)]^{-1}\varsigma_t^2(\boldsymbol{\theta})\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma\middle|\boldsymbol{\phi}\right\}$ and $\mathbf{Z}_{\boldsymbol{\vartheta}_1 s}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = E[\mathbf{Z}_{\boldsymbol{\vartheta}_1 st}(\boldsymbol{\vartheta})|\boldsymbol{\vartheta},\boldsymbol{\eta}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$, as well as the law of iterated expectations, together with the results in Proposition E1. In this context, condition (4) implies that $\mathbf{W}_{\boldsymbol{\vartheta}_1}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be $\mathbf{0}$, so that (B8) reduces to $\mathbf{W}_s(\boldsymbol{\phi}_0) = \begin{bmatrix} 0 & \cdots & 0 & N/(2\vartheta_2) \end{bmatrix}'$.

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 $(\vartheta_2,\boldsymbol{\eta})$. As for the unconditional variance of $s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$, it will be given by

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

while its covariance with $s_{\boldsymbol{\eta}t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be $\mathrm{M}_{sr}(\boldsymbol{\eta}_0)N/(2\vartheta_2)$.

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 (4) holds. In particular, we can show that

$$\mathcal{A}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = \frac{N^2}{4\vartheta_2^2} E\left\{ \left[\delta(\varsigma_t,\boldsymbol{\eta}])\frac{\varsigma_t}{N}-1\right]\left(\frac{\varsigma_t}{N}-1\right)\Big|\,\boldsymbol{\theta},\boldsymbol{\eta}\right\} = \frac{N}{2\vartheta_2^2} \tag{A3}$$

and

$$\mathcal{C}_{\vartheta_2\vartheta_2}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = \frac{\{2(\kappa+1)+N\kappa\}}{4}\frac{4\vartheta_2^2}{N}. \tag{A4}$$

Finally, given a vector $\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta})$ of influence functions that only depend on $\boldsymbol{\vartheta}$ through $\varsigma_t(\boldsymbol{\vartheta})$:

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

But since $\mathbf{W}_{\boldsymbol{\vartheta}_1}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is $\mathbf{0}$, then $cov\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}),\mathbf{s}_{\boldsymbol{\vartheta}_1 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\vartheta},\boldsymbol{\eta}] = \mathbf{0}$ for the same reasons as before.

Proposition 1

Proposition B1 together with the results in Section A.1 imply that the asymptotic variance of the sequential ML estimator will be given by (6). As for $\hat{\boldsymbol{\eta}}_T$, the results in Appendix E combined with the partioned inverse formula imply that $\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)$ can be written as either

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

or (7), with

$$\mathcal{I}^{\vartheta_2\vartheta_2}(\boldsymbol{\phi}_0) = \frac{1}{2\mathrm{M}_{ss}(\boldsymbol{\eta}_0)+N[\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1-\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}'_{sr}(\boldsymbol{\eta}_0)]}\frac{4\vartheta_2^2}{N}. \qquad \square \tag{A5}$$

Proposition 2

It follows directly by combining Proposition B2 with the results in Section A.1. $\square$

Proposition 3

If we combine Proposition B3 with the results in Section A.1, we can prove that

$$cov\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}), s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\phi}] = \frac{N}{2\vartheta_2}cov\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}),\delta(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\Bigg|\,\boldsymbol{\phi}\right] = \frac{N}{2\vartheta_2}\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})$$

in view of (A2). Further, using (A3) it immediately follows that

$$\begin{aligned} \mathbf{n}_t^{\perp}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}) - cov\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}),\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\phi}]\,cov^{-1}\left[\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta},\mathbf{0}),\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\phi}]\,\mathbf{s}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta},\mathbf{0}) \\ &= \mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}) - \frac{cov\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}),s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\phi}]}{cov\left[s_{\vartheta_2 t}(\boldsymbol{\vartheta},\mathbf{0}),s_{\vartheta_2 t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\right|\boldsymbol{\phi}]}s_{\vartheta_2 t}(\boldsymbol{\vartheta},\mathbf{0}) = \mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta}) - \frac{N}{2}\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})\left[\frac{\varsigma_t(\boldsymbol{\theta})}{N}-1\right] \end{aligned}$$

regardless of the original model. But this expression coincides with

$$\mathbf{n}_t^\circ(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}) - \frac{cov\left[\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}),\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t/N|\,\boldsymbol{\phi}\right]}{cov\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t/N,\varsigma_t/N|\,\boldsymbol{\phi}\right]}\left[\frac{\varsigma_t(\boldsymbol{\theta})}{N}-1\right]$$

by definition of $\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})$ since $E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t/N-1\right](\varsigma_t/N-1)|\,\boldsymbol{\phi}\right\} = 2/N$ (see Fiorentini and Sentana (2010) for a proof). On this basis, we can use Proposition B3 to show that the asymptotic variance of the sample average of $\mathbf{n}_t^\circ(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$ will be

$$\begin{aligned}
V\left[\mathbf{n}_t^\circ(\boldsymbol{\vartheta},\boldsymbol{\eta})|\,\boldsymbol{\phi}\right] &= V\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta})|\,\boldsymbol{\phi}\right] - \frac{N}{2}\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})cov'\left[\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}),\frac{\varsigma_t}{N}\right|\boldsymbol{\phi}\Bigg] \\
&-\frac{N}{2}cov\left[\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}),\frac{\varsigma_t}{N}\right|\boldsymbol{\phi}\Bigg]\mathbb{k}'_{\mathbf{n}}(\boldsymbol{\phi}) + \frac{N^2}{4}\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})'V\left(\frac{\varsigma_t}{N}-1\right|\boldsymbol{\phi}\Bigg) \\
= V\left[\mathbf{n}_t(\boldsymbol{\vartheta},\boldsymbol{\eta})|\,\boldsymbol{\phi}\right] &- \frac{N}{2}\left[\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})\ddot{\mathbb{k}}'_{\mathbf{n}}(\boldsymbol{\phi}) + \ddot{\mathbb{k}}_{\mathbf{n}}(\boldsymbol{\phi})\mathbb{k}'_{\mathbf{n}}(\boldsymbol{\phi})\right] + \left(\frac{N}{2}+\frac{N(N+2)\kappa}{4}\right)\mathbb{k}_{\mathbf{n}}(\boldsymbol{\phi})\mathbb{k}'_{\mathbf{n}}(\boldsymbol{\phi}),
\end{aligned}$$

with $\ddot{\mathbb{k}}_{\mathbf{n}}(\boldsymbol{\phi}) = cov\left[\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta}),\varsigma_t/N|\,\boldsymbol{\phi}\right]$, where we have used the fact that $V(\varsigma_t/N) = (N+2)\kappa/N + 2/N$, which follows from (2). Finally, given that $\partial\mathbf{n}_t^\circ(\boldsymbol{\theta},\boldsymbol{\eta})/\partial\boldsymbol{\eta}' = \partial\mathbf{n}_t^\perp(\boldsymbol{\theta},\boldsymbol{\eta})/\partial\boldsymbol{\eta}' = \partial\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta})/\partial\boldsymbol{\eta}'$, the optimal sequential GMM estimators based on $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$ and $\mathbf{n}_t^\circ(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$ will be asymptotically equivalent. □

Proposition 4

In view of Proposition 3, we can easily create moments that are invariant to the sampling uncertainty surrounding $\tilde{\boldsymbol{\theta}}_T$. Specifically, for $m \geq 1$ we get

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

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 (C9) implies that

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

As a result, we can write $\{p_2^\circ[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],...,p_M^\circ[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}$ as a full-rank linear transformation of $[\ell_{2t}^\circ(\boldsymbol{\theta},\boldsymbol{\eta}),...,\ell_{Mt}^\circ(\boldsymbol{\theta},\boldsymbol{\eta})]$, which confirms the asymptotic equivalence in the case of two-step GMM procedures, and the numerical equivalence for single-step ones. In addition, the expression for $V\{p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}$ follows directly from the expression for the polynomials. Specifically, $\mathcal{G}_{\mathbf{p}}$ is a diagonal matrix of order $M-1$ with representative element

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

Similarly, the orthogonality of the polynomials implies that $\Bbbk_{\mathbf{p}}(\mathbf{0}) = cov\{p_m[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \varsigma_t(\boldsymbol{\theta})/N|\boldsymbol{\phi}\} = \mathbf{0}$. Finally, in order to derive expressions for $Cov\{p_m[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}], \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta})/N|\boldsymbol{\phi}\}$, we can use Lemma 1 in Fiorentini and Sentana (2010) to show that

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

Hence, we obtain $\Bbbk_{\mathbf{p}}(\boldsymbol{\eta})$, an $M-1$ vector, with representative element

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

Proposition 5

It follows by combining Proposition B5 and the results in Section A.1. $\square$

Proposition 6

Both sides of the 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 $\mathcal{I}_{\eta\eta}^{-1}(\boldsymbol{\phi}_0) \leq \mathcal{G}_p(\boldsymbol{\phi}_0)/\mathcal{H}_p^2(\boldsymbol{\phi}_0)$, where $\mathcal{H}_p(\boldsymbol{\phi}) = -E[\partial p_2[\varsigma_t(\boldsymbol{\theta}), \eta]/\partial \eta|\boldsymbol{\phi}]$. As for the second component, we must compare $\mathcal{I}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_0)/\mathcal{I}^2_{\eta\eta}(\boldsymbol{\phi}_0)$ with $\mathcal{N}'_p(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_p(\boldsymbol{\phi}_0)/\mathcal{H}^2(\boldsymbol{\phi}_0)$, where $\mathcal{N}_p(\boldsymbol{\phi}_0) = -E[\partial p_2[\varsigma_t(\boldsymbol{\theta}), \eta]/\partial \boldsymbol{\theta}'|\boldsymbol{\phi}]$. Using the results in Proposition B6, it is easy to see that the second expression will be larger than the first one if and only if

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

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

Proposition 7

It follows directly from Proposition B7 and the fact that under reparametrisation (1),

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

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

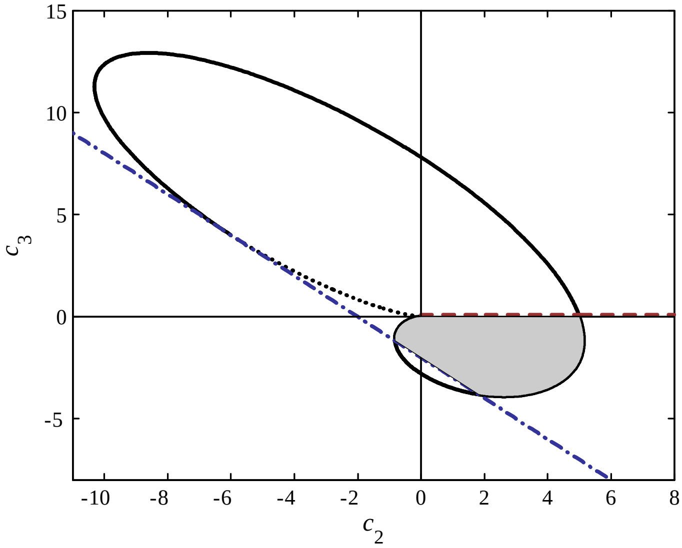
Table 1: Finite sample properties of sequential estimators of shape parameters

		ML	ESMM	SML	SMM
Student t ($\eta_0 = 0.1$)					
	Mean	0.0992	0.0982	0.0981	0.0954
η	MC Std.Dev.	0.0114	0.0113	0.0113	0.0196
	MC Av. Std.Err.	0.0112	0.0112	0.0112	0.0300*
DSMN ($\alpha_0 = 0.05$, $\varkappa_0 = 0.246$)					
	Mean	0.0526	0.0537	0.0537	0.0620
α	MC Std. Dev.	0.0148	0.0152	0.0152	0.0213
	MC Av. Std. Err.	0.0154	0.0160	0.0161	0.0318
	Mean	0.2518	0.2574	0.2574	0.2697
$\varkappa$	MC Std. Dev.	0.0349	0.0352	0.0352	0.0438
	MC Av. Std. Err.	0.0362	0.0370	0.0372	0.0557
PE ($c_{20} = 2.916$, $c_{30} = -1$)					
	Mean	2.9137	2.8663	2.8673	2.8508
c_2	MC Std. Dev.	0.2019	0.1995	0.1997	0.2630
	MC Av. Std. Err.	0.1953	0.1956	0.1963	0.2718
	Mean	-1.0013	-0.9605	-0.9588	-0.9153
c_3	MC Std. Dev.	0.3037	0.2963	0.2972	0.5942
	MC Av. Std. Err.	0.2957	0.2951	0.2960	0.7859

Notes: 1,600 replications, $T = 1,000$, $N = 5$. ML is the joint ML estimator while ESMM and SML refer to the efficient sequential MM and sequential ML estimators, respectively. The orthogonal polynomial MM estimator is labeled SMM. MC Std. Dev. refers to the standard deviation of estimated shape parameters across replications. MC Av. Std. Err is the square root of the mean across simulated samples of the estimated variances of the shape parameters. 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 F for a detailed description of the Monte Carlo study.

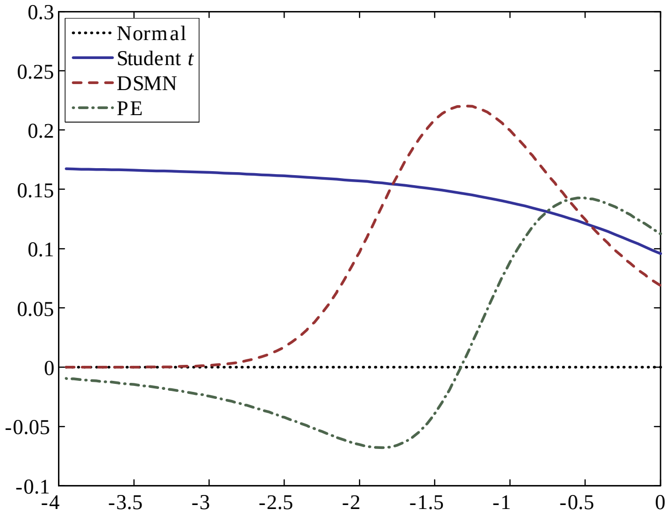
*This excludes 63 samples whose parameter estimates were below 8 degrees of freedom.

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



Notes: The solid (dotted) black line represents the frontier defined by positive (negative) values of ς. The blue (dotted-dashed) line represents the tangent of $P_3(\varsigma)$ at $\varsigma = 0$ while the red (dashed) line is the tangent of $P_3(\varsigma)$ when $\varsigma \to +\infty$. The grey area defines the admissible set in (c_2, c_3) space.

Figure 2: Exceedance correlation

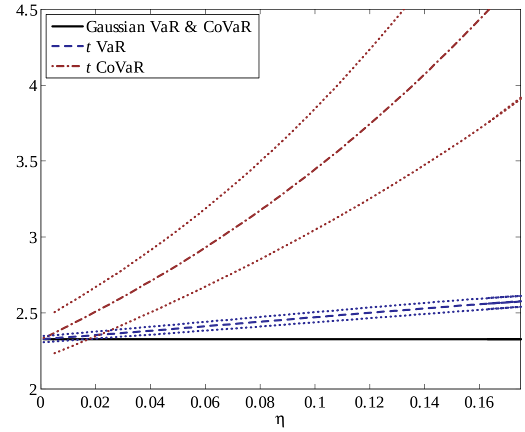


Notes: The exceedance correlation between two variables ε_1^* and ε_2^* is defined as $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* > \varrho, \varepsilon_2^* > \varrho)$ for positive ϱ and $corr(\varepsilon_1^*, \varepsilon_2^* | \varepsilon_1^* < \varrho, \varepsilon_2^* < \varrho)$ for 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 3^{rd}-order PE with the same kurtosis and $c_3 = -1$.

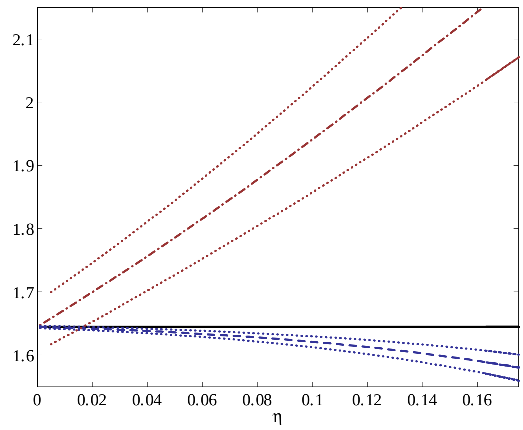
Figure 3: VaR, CoVaR and their 95% confidence intervals

Student t innovations

(a) 99% VaR and CoVaR

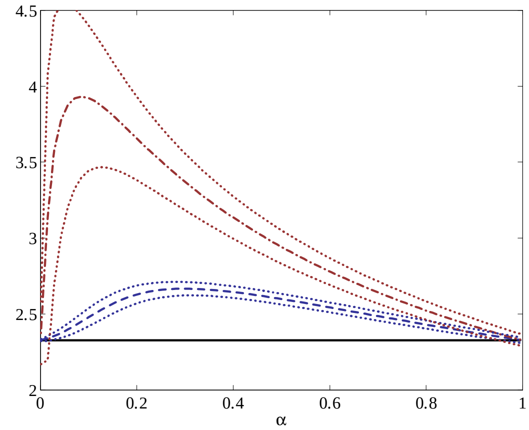


(b) 95% VaR and CoVaR

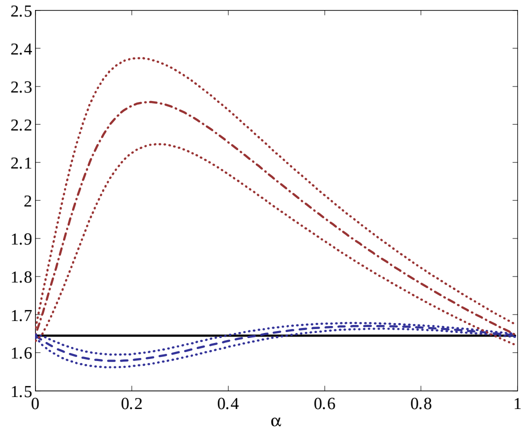


DSMN innovations

(c) 99% VaR and CoVaR

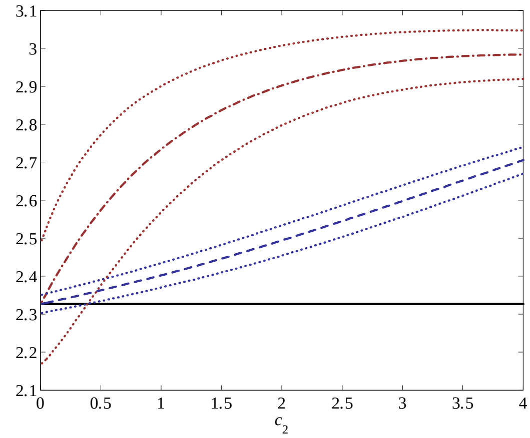


(d) 95% VaR and CoVaR

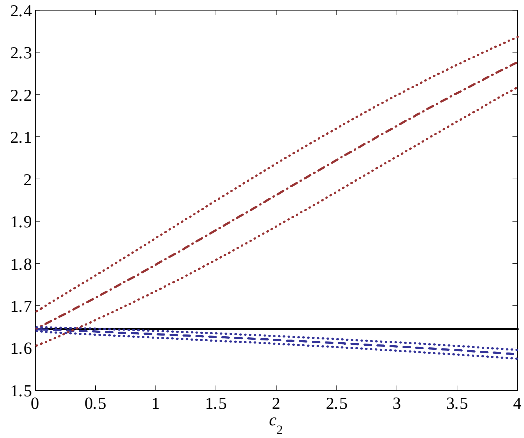


PE innovations

(e) 99% VaR and CoVaR

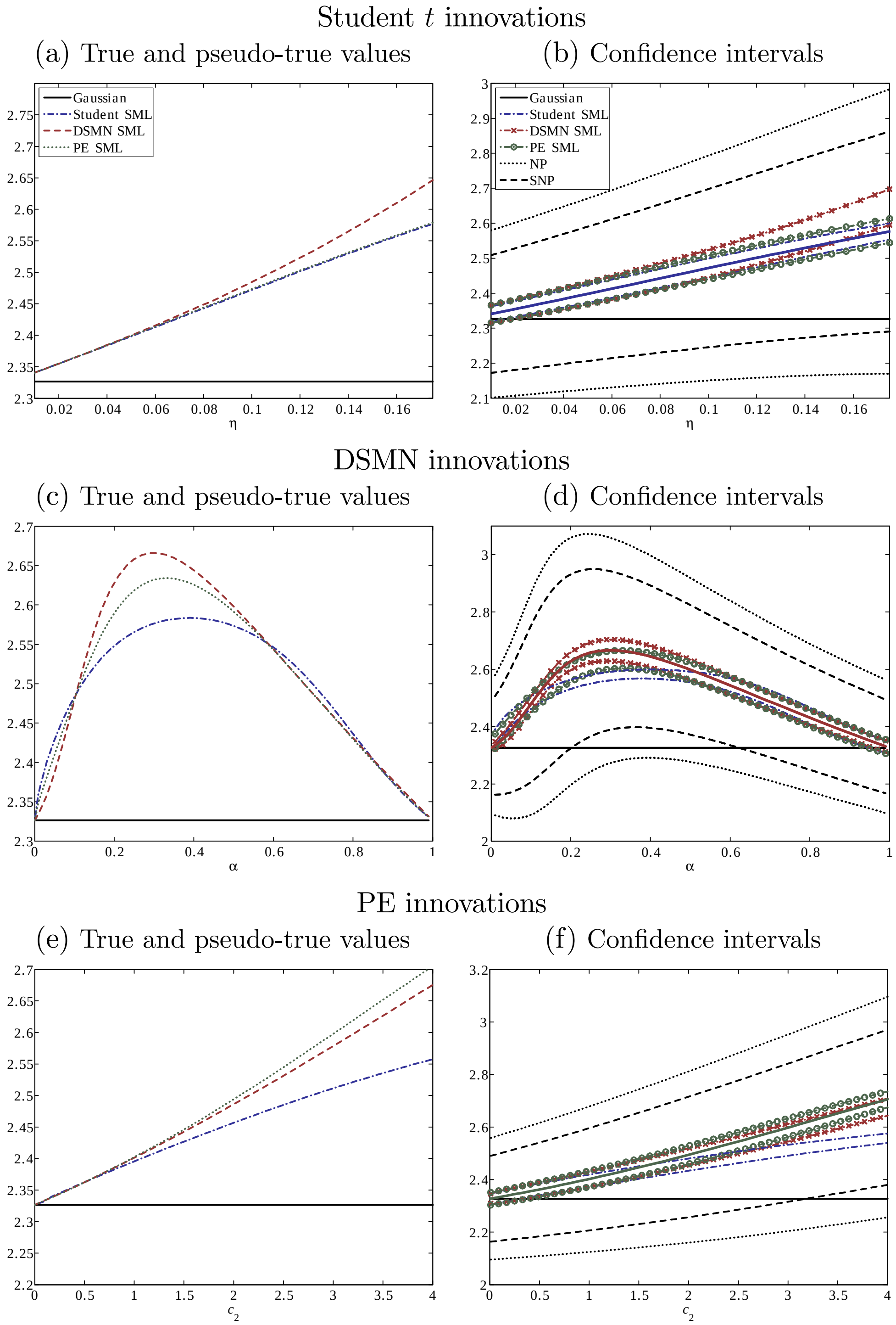


(f) 95% VaR and CoVaR



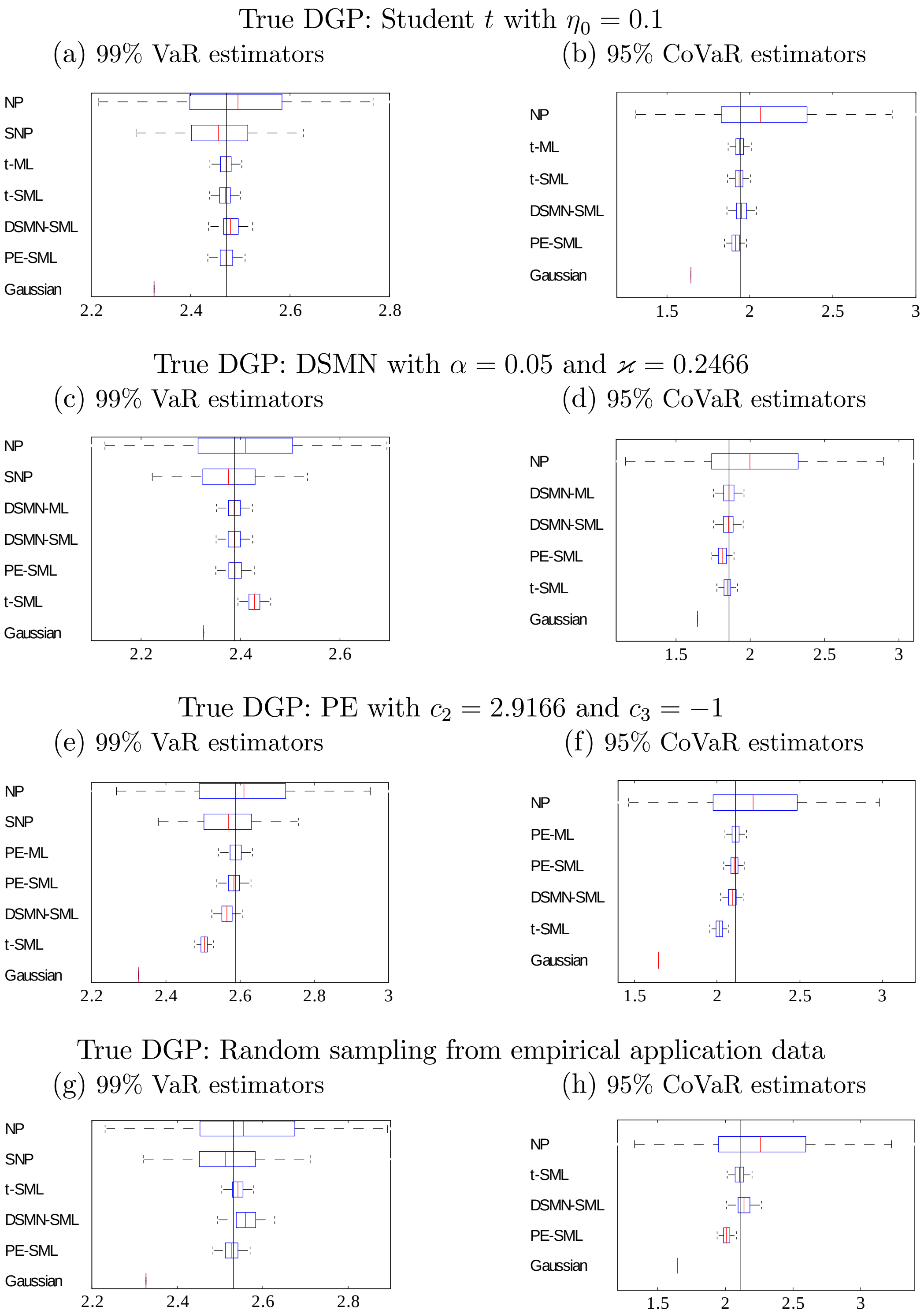
Notes: For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. For DSMN innovations, α denotes the mixing probability, while the variance ratio of the two components $\varkappa$ remains fixed at 0.25. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, with $c_3 = -c_2/3$. 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 4: VaR (99%) estimators and confidence intervals



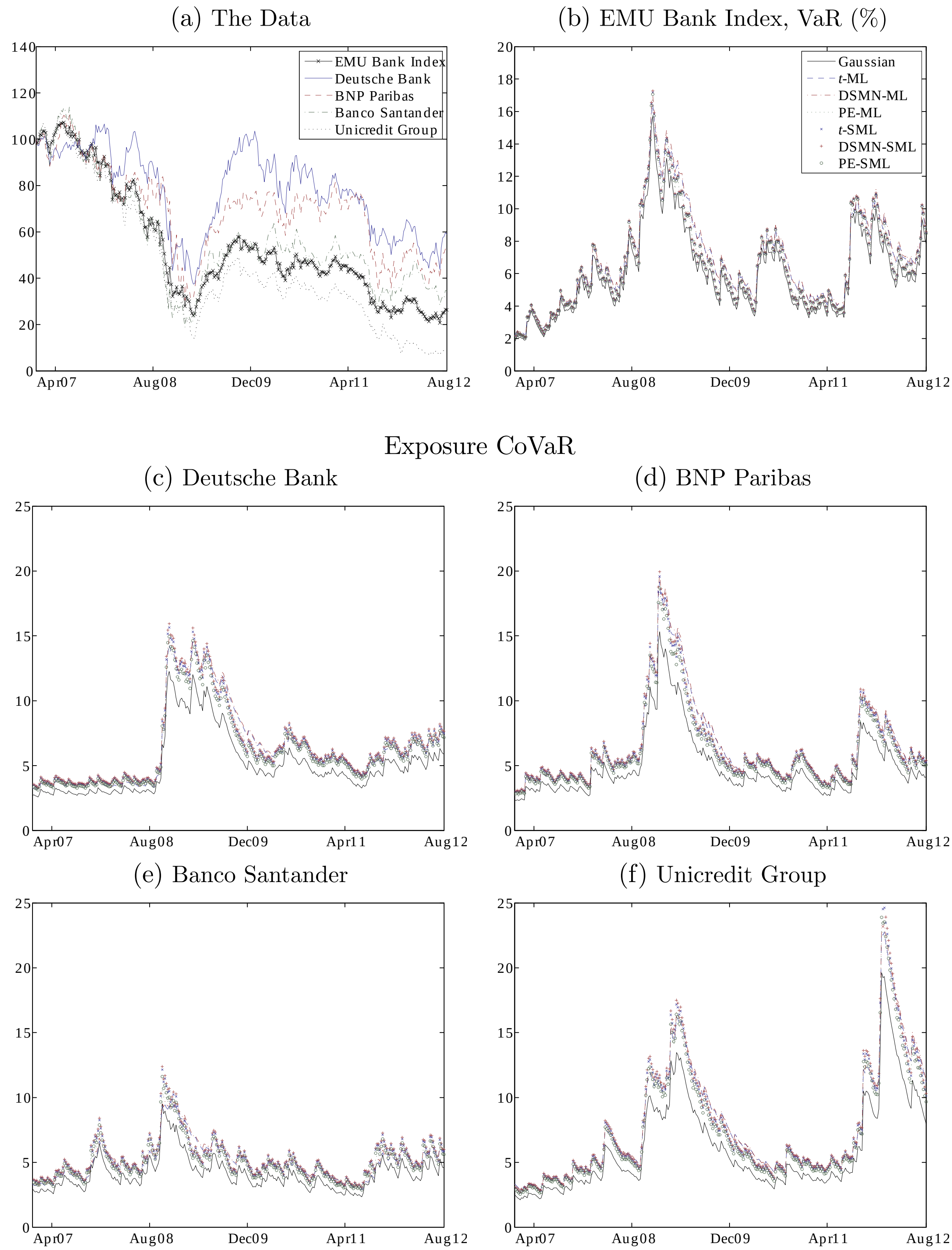
Notes: For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. For DSMN innovations, α denotes the mixing probability, while the variance ratio of the two components $\varkappa$ remains fixed at 0.25. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, with $c_3 = -c_2/3$. 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 1% 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 5: Monte Carlo distributions of VaR and CoVaR estimators



Notes: 1,600 replications, $T = 1,000$, $N = 5$. The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The length of the whiskers is one interquartile range. For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. For DSMN innovations, α and $\varkappa$ denote the mixing probability and the variance ratio of the two components, respectively. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2 - 1$. ML and SML denote joint and sequential maximum likelihood estimator, respectively, while NP and SNP refers to the nonparametric estimators. Vertical lines represent the true values. See Section 5.1 and Appendix E.2 for a detailed description of the Monte Carlo study.

Figure 6: Application to G-SIBS Euro zone banks



Notes: Sample: October 27, 1993 – August 29, 2012. For model specification see Section 6. Excess returns are computed by subtracting the continuously compounded rate of return on the one-week Eurocurrency rate in DM/Euros applicable over the relevant week. Exposure CoVaR figures (in percentage terms) are at the 5% level when the fall in the euro area bank index exceeds its 5^{th} percentile. ML and SML denote joint and sequential maximum likelihood estimates, respectively.