A COMPARISON OF MEAN-VARIANCE EFFICIENCY TESTS

Dante Amengual and Enrique Sentana

CEMFI Working Paper No. 0806

April 2008

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551 Fax (34) 914 291 056
Internet: www.cemfi.es

We would like to thank Manuel Arellano, Christian Bontemps, Marcelo Fernandes, Gabriele Fiorentini, Javier Mencía, Nour Meddahi, Francisco Peñaranda and Kevin Sheppard, as well as audiences at Queen Mary, the Imperial College Financial Econometrics Conference (London, 2007), the XV Finance Forum (Majorca), and the XXXII Symposium on Economic Analysis (Granada) for useful comments and suggestions. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Education and Science through grant SEJ 2005-08880 (Sentana) is gratefully acknowledged.

CEMFI Working Paper 0806
April 2008

A COMPARISON OF MEAN-VARIANCE EFFICIENCY TESTS

Abstract

We analyse the asymptotic properties of mean-variance efficiency tests based on generalised methods of moments, and parametric and semiparametric likelihood procedures that assume elliptical innovations. We study the trade-off between efficiency and robustness, and prove that the fully parametric estimators provide asymptotically valid inferences when the conditional distribution of the innovations is elliptical but possibly misspecificed and heteroskedastic. We compare the small sample performance of the alternative tests in a Monte Carlo study, and find some discrepancies with their asymptotic properties. Finally, we present an empirical application to US stock returns, which rejects the mean-variance efficiency of the market portfolio.

JEL Codes: C12, C13, C14, C16, G11, G12.
Keywords: Adaptivity, Elliptical Distributions, Financial Returns, Portfolio choice, Semiparametric Estimators.

Dante Amengual
Princeton University
amengual@princeton.edu

Enrique Sentana
CEMFI
sentana@cemfi.es

1 Introduction

Mean-variance analysis is widely regarded as the cornerstone of modern investment theory. Despite its simplicity, and the fact that more than five and a half decades have elapsed since Markowitz published his seminal work on the theory of portfolio allocation under uncertainty (Markowitz (1952)), it remains the most widely used asset allocation method. A portfolio with excess returns r_{Mt} is mean-variance efficient with respect to a given set of N assets with excess returns $\mathbf{r}_t$ if it is not possible to form another portfolio of those assets and r_{Mt} with the same expected return as r_{Mt} but a lower variance. Despite the simplicity of this definition, testing for mean-variance efficiency is of paramount importance in many practical situations, such as mutual fund performance evaluation (see De Roon and Nijman (2001) for a recent survey), gains from portfolio diversification (Errunza, Hogan and Hung (1999)), or tests of linear factor asset pricing models, including the CAPM and APT, as well as other empirically oriented asset pricing models (see e.g. Campbell, Lo and MacKinlay (1996) or Cochrane (2001) for advanced textbook treatments).

As is well known, r_{Mt} will be mean-variance efficient with respect to $\mathbf{r}_t$ in the presence of a riskless asset if and only if the intercepts in the theoretical least squares projection of $\mathbf{r}_t$ on a constant and r_{Mt} are all 0 (see Jobson and Korkie (1982), Gibbons, Ross and Shanken (1989) and Huberman and Kandel (1987)). Therefore, it is not surprising that this early literature resorted to ordinary least squares (OLS) to test those theoretical restrictions empirically. If the distribution of $\mathbf{r}_t$ conditional on r_{Mt} (and their past) were multivariate normal, with a linear mean $\mathbf{a} + \mathbf{b}r_{Mt}$ and a constant covariance matrix $\boldsymbol{\Omega}$, then OLS would produce efficient estimators of the regression intercepts $\mathbf{a}$, and consequently, optimal tests of the mean-variance efficiency restrictions $H_0 : \mathbf{a} = \mathbf{0}$. In addition, it is possible to derive an F version of the test statistic whose sampling distribution in finite samples is known under exactly the same restrictive distributional assumptions (see Gibbons, Ross and Shanken (1989)). In this sense, this F-test generalises the t-test proposed by Black, Jensen and Scholes (1972) in univariate contexts.

However, many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic. For that reason, MacKinlay and Richardson (1991) proposed alternative tests based on the generalised method of moments (GMM) that are robust to non-normality, unlike traditional OLS test statistics.

More recently, Hodgson, Linton, and Vorkink (2002; hereinafter HLV) developed a semiparametric estimation and testing methodology that enabled them to obtain optimal mean-variance efficiency tests under the assumption that the distribution of $\mathbf{r}_t$ conditional on r_{Mt} (and their past) is elliptically symmetric. Specifically, HLV showed that their proposed estimators of $\mathbf{a}$ and $\mathbf{b}$ are adaptive under the aforementioned assumptions of linear conditional mean and constant conditional variance, which means that they are as efficient as infeasible maximum likelihood estimators that use the correct parametric elliptical density with full knowledge of its shape parameters. Elliptical distributions are very attractive in this context because they guarantee that mean-variance analysis is fully compatible with expected utility maximisation regardless of investors' preferences (see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997)). Moreover, they generalise the multivariate normal distribution, but at the same time they retain its analytical tractability irrespective of the number of assets.

Nevertheless, the finite sample performance of such semiparametric inference procedures may not be well approximated by the first-order asymptotic theory that justifies them. For that reason, an alternative approach worth considering is a feasible maximum likelihood estimator based on the correct elliptical distribution, but which includes the unknown shape parameters as additional arguments in the maximisation algorithm (see e.g. Kan and Zhou (2006)). However, unless we are careful, this last approach may provide misleading inference if the relevant conditional distribution does not coincide with the assumed one, even if both are elliptical. Similarly, the HLV approach may also lead to erroneous inferences if the true conditional distribution is either heteroskedastic or asymmetric.

The purpose of our paper is to shed some light on these efficiency-consistency trade-offs in the context of mean-variance efficiency tests. To do so, we derive the asymptotic properties of tests based on GMM, HLV and elliptically-based parametric ML estimators under correct specification and under several potentially relevant forms of misspecification. In particular, we study those situations in which the distribution of the innovations is (i) *i.i.d.* elliptical but different from the one assumed for estimation purposes, (ii) *i.i.d.* but asymmetric, and (iii) elliptical but conditionally heteroskedastic. In addition, given that it is far from trivial to obtain exact finite sample distributions once we abandon the Gaussianity assumption, we also analyse the reliability of the usual asymptotic approximations by Monte Carlo methods.[1]

[1] See Beaulieu, Dufour and Khalaf (2007) for a method to obtain the exact distribution of the Gibbons, Ross and Shanken (1989) F-statistic conditional on the full sample path of r_{Mt} when the innovations are *i.i.d.*

Finally, we apply those different procedures to test the mean-variance efficiency of the US aggregate stock market portfolio with respect to industry portfolios, and the book-to-market sorted portfolios popularised by Fama and French (1993). We do so using monthly data over the period July 1962 to June 2007. Importantly, we also compute specification tests to assess the adequacy of our parametric distributional assumptions.

The rest of the paper is organised as follows. In section 2, we introduce the model and the three aforementioned estimation procedures, and obtain their asymptotic distributions under the assumption that the innovations are *i.i.d.* elliptical. Then in section 3 we derive the asymptotic properties of mean-variance efficiency tests in alternative misspecified contexts. An extensive Monte Carlo evaluation of the different parameter estimators and testing procedures can be found in Section 4, while Section 5 reports our empirical results. Finally, we present our conclusions and suggestions for future work in section 6. Proofs are gathered in the Appendix.

2 Econometric methods

2.1 Model description

Consider the following multivariate, conditionally homoskedastic, linear regression model

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \mathbf{u}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*, \tag{1}$$

where $\boldsymbol{\Omega}^{1/2}$ is an $N \times N$ "square root" matrix such that $\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2} = \boldsymbol{\Omega}$, $\boldsymbol{\varepsilon}_t^*$ is a standardised vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\omega}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\omega}_0) = \mathbf{I}_N$, $\boldsymbol{\gamma}' = (\mathbf{a}', \mathbf{b}')$, $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$, the subscript 0 refers to the true values of the parameters, and I_{t-1} denotes the information set available at $t-1$, which contains at least past values of r_{Mt} and $\mathbf{r}_t$. To complete the conditional model, we need to specify the distribution of $\boldsymbol{\varepsilon}_t^*$. We shall initially assume that conditional on r_{Mt} and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as some particular member of the elliptical family with a well defined density, or $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\omega}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ for short, where $\boldsymbol{\eta}$ are some q additional parameters that determine the shape of the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$.[2] The most prominent example is the spherical normal distribution, which we denote by $\boldsymbol{\eta} = \mathbf{0}$. Another popular example is a standardised multivariate t with ν_0 degrees of freedom, or *i.i.d.*

[2] If $\boldsymbol{\varepsilon}_t^*$ is distributed as a spherically symmetric multivariate random vector, then we can write $\boldsymbol{\varepsilon}_t^* = e_t\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and $e_t = \sqrt{\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*}$ is a nonnegative random variable that is independent of $\mathbf{u}_t$. Assuming that $E\left[e_t^2\right] < \infty$, then $\boldsymbol{\varepsilon}_t^*$ can be standardised by setting $E\left[e_t^2\right] = N$, so that $E\left[\boldsymbol{\varepsilon}_t^*\right] = \mathbf{0}$ and $V\left[\boldsymbol{\varepsilon}_t^*\right] = \mathbf{I}_N$.

$t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ for short. As is well known, the multivariate student t approaches the multivariate normal as $\nu_0 \to \infty$, but has generally fatter tails. For that reason, we define η as $1/\nu$, which will always remain in the finite range [0,1/2) under our assumptions. As in Zhou (1993) we also consider two other illustrative examples: a Kotz distribution and a discrete scale mixture of normals.

The original Kotz distribution (see Kotz (1975)) is such that ς_t is a gamma random variable with mean N and variance $N[(N+2)\kappa_0 + 2]$, where

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

is the coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$ (see Mardia (1970)). The Kotz distribution nests the multivariate normal distribution for $\kappa = 0$, but it can also be either platykurtic ($\kappa < 0$) or leptokurtic ($\kappa > 0$). However, the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view.

For that reason, we also consider a standardised version of a two-component scale mixture of multivariate normals, which can be generated as

$$\boldsymbol{\varepsilon}_t^* = \frac{s_t + (1 - s_t)\sqrt{\varkappa}}{\sqrt{\pi + (1-\pi)\varkappa}} \cdot \boldsymbol{\varepsilon}_t^\circ, \tag{2}$$

where $\boldsymbol{\varepsilon}_t^\circ$ is a spherical multivariate normal, s_t is an independent Bernoulli variate with $P(s_t = 1) = \pi$ and $\varkappa$ is the variance ratio of the two components. Not surprisingly, ς_t will be a two-component scale mixture of $\chi_N^2\prime s$. As all scale mixtures of normals, the distribution of $\boldsymbol{\varepsilon}_t^*$ is leptokurtic, so that

$$\kappa = \frac{\pi(1-\pi)(1-\varkappa^2)}{[\pi + (1-\pi)\varkappa]^2} \geq 0,$$

with equality if and only if either $\varkappa = 1$, $\pi = 1$ or $\pi = 0$, when it reduces to the spherical normal. In general, though, we require at least sixth moments to globally identify $\boldsymbol{\eta} = (\pi, \varkappa)'$.[3] In this sense, a noteworthy property of all discrete mixtures of normals is that their density and moments are always bounded.

Figure 1 plots the densities of a normal, a student t, a platykurtic Kotz distribution and a discrete scale mixture of normals in the bivariate case, as well as their contours. Although they all have 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

[3]Since the labels of the components are arbitrary, we also need to impose either $0 \leq \varkappa \leq 1$ or $\pi \geq \frac{1}{2}$.

dependence" between the components, the normal being the only example in which lack of correlation is equivalent to stochastic independence. Allowing for dependence beyond correlation is particularly important in the context of multiple financial assets, in which the probability of the joint occurrence of several extreme events is regularly underestimated by the multivariate normal distribution.

2.2 Parameter estimation

2.2.1 Maximum likelihood estimators

Let $\boldsymbol{\phi} = (\boldsymbol{\gamma}', \boldsymbol{\omega}', \boldsymbol{\eta})' \equiv (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $2N + N(N+1)/2 + q$ parameters of interest, which we assume variation free. The log-likelihood function of a sample of size T based on a particular parametric spherical assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{\Omega}|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \mathbf{a} - \mathbf{b}r_{Mt}$.[4]

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into three blocks, $\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi})$, $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\phi})$, and $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\gamma}$, $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$, respectively. A straightforward application of expression (2) in Fiorentini and Sentana (2007) implies that

$$\mathbf{s}_{\gamma t}(\boldsymbol{\phi}) = \begin{pmatrix} 1 \\ r_{Mt} \end{pmatrix} \otimes \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \tag{3}$$

$$\mathbf{s}_{\omega t}(\boldsymbol{\phi}) = \frac{1}{2}\mathbf{D}_N'\left[\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right] vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta}) - \boldsymbol{\Omega}\right\}, \tag{4}$$

where $\mathbf{D}_N$ is the duplication matrix of order N such that $vec(\boldsymbol{\Omega}) = \mathbf{D}_N vech(\boldsymbol{\Omega})$ (see Magnus and Neudecker (1989)), while the scalar

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

reduces to

$$(N\eta + 1)/[1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})]$$

in the student t case, to

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

[4] Fiorentini, Sentana and Calzolari (2003) provide expressions for $c(\eta)$ and $g_t\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]$ in the multivariate student case, which under normality collapse to $-(N/2)\log\pi$ and $-\frac{1}{2}\varsigma_t(\boldsymbol{\theta})$, respectively.

in the case of the Kotz distribution, to

$$[\pi+(1-\pi)\varkappa]\cdot\frac{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{[\pi+(1-\pi)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{[\pi+(1-\pi)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]} \tag{5}$$

for the two-component mixture, and to 1 under Gaussianity.[5]

Given correct specification, the results in Crowder (1976) imply that the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ evaluated at the true parameter values has the martingale difference property. His results also imply that, under suitable regularity conditions, which typically require that both r_{Mt} and r_{Mt}^2 are strictly stationary process with absolutely summable autocovariances, the asymptotic distribution of the feasible ML estimator will be given by the following expression

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

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

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

and $\mathbf{h}_t(\boldsymbol{\phi})$ denotes the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}'=\partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. These expressions adopt particularly simple forms for our model of interest:

Proposition 1 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt},I_{t-1};\boldsymbol{\phi}$ in (1) is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ with density $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$, then the only non-zero elements of $\mathcal{I}_t(\boldsymbol{\phi}_0)$ will be:*

$$\begin{aligned}
\mathcal{I}_{\boldsymbol{\gamma\gamma}t}(\boldsymbol{\phi}) &= \mathrm{M}_{ll}(\boldsymbol{\eta})\begin{pmatrix}1 & r_{Mt}\\ r_{Mt} & r_{Mt}^2\end{pmatrix}\otimes\boldsymbol{\Omega}^{-1},\\
\mathcal{I}_{\boldsymbol{\omega\omega}t}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{ss}(\boldsymbol{\eta})}{2}\mathbf{D}_N'\left[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right]\mathbf{D}_N+\frac{\mathrm{M}_{ss}(\boldsymbol{\eta})-1}{4}\mathbf{D}_N'\left[vec(\boldsymbol{\Omega}^{-1})vec'(\boldsymbol{\Omega}^{-1})\right]\mathbf{D}_N,\\
\mathcal{I}_{\boldsymbol{\omega\eta}t}(\boldsymbol{\phi}) &= \frac{1}{2}\mathrm{M}_{sr}(\boldsymbol{\eta})\mathbf{D}_N'vec(\boldsymbol{\Omega}^{-1}),\\
\mathcal{I}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= V[\,\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\,\boldsymbol{\phi}]=-E[\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi})|\,\boldsymbol{\phi}],
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{M}_{ll}(\boldsymbol{\eta}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}\right|\boldsymbol{\phi}\Bigg\}=E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t(\boldsymbol{\theta})}{N}+\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\right|\boldsymbol{\phi}\Bigg\},\\
\mathrm{M}_{ss}(\boldsymbol{\eta}) &= \frac{N}{N+2}\left[1+V\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t}{N}\right|\boldsymbol{\phi}\Bigg\}\right]=E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)}\right|\boldsymbol{\phi}\Bigg\}+1,\\
\mathrm{M}_{sr}(\boldsymbol{\eta}) &= E\left[\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}-1\right\}\mathbf{e}_{rt}'(\boldsymbol{\phi})\right|\boldsymbol{\phi}\Bigg]=-E\left\{\frac{\varsigma_t(\boldsymbol{\theta})}{N}\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'}\right|\boldsymbol{\phi}\Bigg\}.
\end{aligned}$$

[5] See Fiorentini, Sentana and Calzolari (2003) for numerically reliable expressions for $s_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\eta t}(\boldsymbol{\phi})$ in the multivariate t case.

In the multivariate standardised student t case, in particular:

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

which under normality reduce to 1, 1 and 0, respectively (see Fiorentini, Sentana and Calzolari (2003)).

As for the Kotz distribution, we can combine the moments of the gamma and reciprocal gamma random variables to show that

$$\mathrm{M}_{ll}(\kappa) = \frac{1}{[(N+2)\kappa+2]^2}\left\{\frac{N(N+2)^2\kappa^2}{N-[(N+2)\kappa+2]} + 4[(N+2)\kappa+1]\right\}, \tag{6}$$

as long as $\kappa < (N-2)/(N+2)$ when $\kappa \neq 0$,

$$\mathrm{M}_{ss}(\kappa) = \frac{1}{[(N+2)\kappa+2]^2}\left\{(N+2)^2\kappa^2 + \frac{4}{N}[N+(N+2)\kappa+2] + 4(N+2)\kappa\right\},$$

and $\mathrm{M}_{sr}(\kappa) = 0\ \forall\kappa$, as in the Gaussian case.

Finally, we provide the relevant expressions for the case of the two-component scale mixture of normals in Appendix D.

The next result follows directly from Proposition 1

Corollary 1 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1};\boldsymbol{\phi}_0$ in (1) is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with density $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$ such that $\mathrm{M}_{ll}(\boldsymbol{\eta}_0)<\infty$, and both r_{Mt} and r_{Mt}^2 are strictly stationary processes with absolutely summable autocovariances, then*

$$\sqrt{T}(\hat{\mathbf{a}}_{ML}-\mathbf{a}_0) \to N[\mathbf{0},\mathcal{I}^{\mathbf{aa}}(\boldsymbol{\phi}_0)], \tag{7}$$

where

$$\mathcal{I}^{\mathbf{aa}}(\boldsymbol{\phi}) = [\mathcal{I}_{\mathbf{aa}}(\boldsymbol{\phi}) - \mathcal{I}_{\mathbf{ab}}(\boldsymbol{\phi})\mathcal{I}_{\mathbf{bb}}^{-1}(\boldsymbol{\phi})\mathcal{I}'_{\mathbf{ab}}(\boldsymbol{\phi})]^{-1} = \frac{1}{\mathrm{M}_{ll}(\boldsymbol{\eta})}\left(1+\frac{\mu_M^2}{\sigma_M^2}\right)\boldsymbol{\Omega},$$

$\mu_M = E(r_{Mt}|\boldsymbol{\phi})$ and $\sigma_M^2 = V(r_{Mt}|\boldsymbol{\phi})$, so that μ_M^2/σ_M^2 is the square Sharpe ratio of the reference portfolio.

Importantly, expression (7) is valid regardless of whether or not the shape parameters $\boldsymbol{\eta}$ are fixed to their true values $\boldsymbol{\eta}_0$, as in the infeasible ML estimator, $\hat{\mathbf{a}}_{IML}$ say, or jointly estimated with $\boldsymbol{\theta}$, as in the feasible one, $\hat{\mathbf{a}}_{FML}$ say. The reason is that the scores corresponding to the mean parameters, $\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi}_0)$, and the scores corresponding to variance and shape parameters, $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\phi}_0)$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, respectively, are asymptotically uncorrelated under our sphericity assumption in view of Proposition 1.

2.2.2 GMM estimators

MacKinlay and Richardson (1991) developed a robust test of mean-variance efficiency by using Hansen's (1982) GMM methodology. If we call $\mathbf{R}_t' \equiv (r_{Mt}, \mathbf{r}_t')$, the orthogonality conditions that they considered are

$$
\begin{gathered}
E\left[\mathbf{m}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}\right)\right]=\mathbf{0}, \\
\mathbf{m}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}\right)=\begin{pmatrix} 1 \\ r_{Mt} \end{pmatrix}\otimes\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}).
\end{gathered} \tag{8}
$$

The advantage of working within a GMM framework is that under fairly weak regularity conditions inference can be made robust to departures from the assumption of normality, conditional homoskedasticity, serial independence or identity of distribution. But since the above moment conditions exactly identify $\boldsymbol{\gamma}$, the unrestricted GMM estimators coincide with the Gaussian pseudo ML estimators, which in turn coincide with the equation by equation OLS estimators in the regression of each element of $\mathbf{r}_t$ on a constant and r_{Mt}. An alternative way of reaching the same conclusion is by noticing that the influence function $\mathbf{m}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}\right)$ is a full-rank linear transformation with time-invariant weights of the Gaussian pseudo-score $\mathbf{s}_{\gamma t}(\boldsymbol{\theta},\boldsymbol{\eta}=\mathbf{0})$.[6]

It is convenient to derive an expression for the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}_{GMM}$ under *i.i.d.* innovations:

Proposition 2 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1};\boldsymbol{\phi}$ in (1) is i.i.d. $(\mathbf{0},\mathbf{I}_N)$ with density function $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})$, where $\boldsymbol{\varrho}$ are some shape parameters, and both r_{Mt} and r_{Mt}^2 are strictly stationary processes with absolutely summable autocovariances, then*

$$
\sqrt{T}(\hat{\boldsymbol{\gamma}}_{GMM}-\boldsymbol{\gamma}_0)\to N\left[\mathbf{0},\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}_0)\right], \tag{9}
$$

where

$$
\begin{gathered}
\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi})=\mathcal{A}_{\gamma\gamma}^{-1}(\boldsymbol{\phi})\mathcal{B}_{\gamma\gamma}(\boldsymbol{\phi})\mathcal{A}_{\gamma\gamma}^{-1}(\boldsymbol{\phi}), \\
\mathcal{A}_{\gamma\gamma}(\boldsymbol{\phi})=-E\left[\mathbf{h}_{\gamma\gamma t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=E\left[\mathcal{A}_{\gamma\gamma t}(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{A}_{\gamma\gamma t}(\boldsymbol{\phi})=-E[\mathbf{h}_{\gamma\gamma t}(\boldsymbol{\theta};\mathbf{0})|\,r_{Mt},I_{t-1};\boldsymbol{\phi}]=\begin{pmatrix} 1 & r_{Mt} \\ r_{Mt} & r_{Mt}^2 \end{pmatrix}\otimes\boldsymbol{\Omega}^{-1}, \\
\mathcal{B}_{\gamma\gamma}(\boldsymbol{\phi})=V\left[\mathbf{s}_{\gamma t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=E\left[\mathcal{B}_{\gamma\gamma t}(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{B}_{\gamma\gamma t}(\boldsymbol{\phi})=V[\mathbf{s}_{\gamma t}(\boldsymbol{\theta};\mathbf{0})|\,r_{Mt},I_{t-1};\boldsymbol{\phi}]=\mathcal{A}_{\gamma\gamma t}(\boldsymbol{\phi}),
\end{gathered}
$$

[6] The obvious GMM estimator of $\boldsymbol{\omega}$ is given by $\hat{\boldsymbol{\Omega}}_{GMM}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\gamma}}_{GMM})\boldsymbol{\varepsilon}_t'(\hat{\boldsymbol{\gamma}}_{GMM})$, which is the sample analogue to the residual covariance matrix.

so that

$$\mathcal{C}_{\gamma\gamma}(\phi_0) = \begin{pmatrix} (1+\mu_{M0}^2/\sigma_{M0}^2) & -\mu_{M0}/\sigma_{M0}^2 \\ -\mu_{M0}/\sigma_{M0}^2 & 1/\sigma_{M0}^2 \end{pmatrix} \otimes \mathbf{\Omega}_0.$$

Importantly, note that $\mathcal{C}_{\gamma\gamma}(\phi_0)$ does not depend on the specific distribution for the innovations that we are considering, regardless of whether or not the conditional distribution of ε_t^* is spherical, as long as it is *i.i.d.*[7]

2.2.3 HLV elliptically symmetric semiparametric estimators

HLV proposed a semiparametric estimator of multivariate linear regression models that updates $\hat{\boldsymbol{\theta}}_{GMM}$ (or any other root-T consistent estimator) by means of a single scoring iteration without line searches. The crucial ingredient of their method is the so-called elliptically symmetric semiparametric efficient score (see Proposition 7 in Fiorentini and Sentana (2007)):

$$\mathring{\mathbf{s}}_{\theta t}(\phi_0) = \mathbf{s}_{\theta t}(\phi_0) - \mathbf{W}_s(\phi_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2}{(N+2)\kappa_0+2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\right\},$$

where

$$\mathbf{W}_s'(\phi) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \frac{1}{2}vec'(\mathbf{\Omega}^{-1})\mathbf{D}_N \end{bmatrix}$$

in the case of model (1). In fact, the special structure of $\mathbf{W}_s(\phi)$ implies that we can update the GMM estimator of $\boldsymbol{\gamma}$ by means of the following simple BHHH correction:

$$\left[\sum_{t=1}^{T} \mathbf{s}_{\gamma t}(\phi_0)\mathbf{s}_{\gamma t}'(\phi_0)\right]^{-1} \sum_{t=1}^{T} \mathbf{s}_{\gamma t}(\phi_0), \tag{10}$$

which does not require the computation of $\mathring{\mathbf{s}}_{\omega t}(\phi_0)$. In practice, of course, $\mathbf{s}_{\gamma t}(\phi_0)$ has to be replaced by a semiparametric estimate obtained from the joint density of ε_t^*. However, the elliptical symmetry assumption allows one to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h(\varsigma_t; \boldsymbol{\eta})$, avoiding in this way the curse of dimensionality (see appendix B1 in Fiorentini and Sentana (2007) for details).

Proposition 7 in Fiorentini and Sentana (2007) shows that the elliptically symmetric semiparametric efficiency bound will be given by:

$$\mathring{\mathcal{S}}(\phi_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\phi_0) - \mathbf{W}_s(\phi_0)\mathbf{W}_s'(\phi_0) \cdot \left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]}\right\},$$

which implies that $\mathring{\mathcal{S}}_{\gamma\gamma}(\phi_0) = \mathcal{I}_{\gamma\gamma}(\phi_0)$ in our case in view of the structure of $\mathbf{W}_s(\phi_0)$. This result confirms that the HLV estimator of $\boldsymbol{\gamma}$ is adaptive.[8]

[7] The asumption of constant conditional third and fourth moments implicit in the assumption of *i.i.d.* innovations also implies that the optimal GMM estimators of Meddahi and Renault (1998) do not offer any asymptotic efficiency gains over $\hat{\mathbf{a}}_{GMM}$.

[8] HLV also consider alternative estimators that iterate the semiparametric adjustment (10) until it becomes negligible. However, since they have the same asymptotic distribution, we shall not discuss them separately.

3 Asymptotic comparison of test procedures

3.1 Correctly specified innovation distributions

Let $\hat{\mathbf{a}}$ denote any of the asymptotically normal, root-T estimators of $\mathbf{a}$ analysed in the previous section, and denote its asymptotic covariance matrix by $V(\hat{\mathbf{a}})$. To test $H_0 : \mathbf{a} = \mathbf{0}$, we can in principle use any of the trinity of classical hypothesis tests, namely, Wald (W_T), Lagrange Multiplier (LM_T) and Likelihood Ratio/Distance Metric test (LR_T). For the sake of concreteness, though, we shall centre our discussion around the Wald test, which examines whether the homogeneity constraints imposed by H_0 are approximately satisfied by $\hat{\mathbf{a}}$.[9] More formally,

$$W_T = T \cdot \hat{\mathbf{a}}' V^{-1}(\hat{\mathbf{a}})\hat{\mathbf{a}}.$$

As is well known, W_T will be asymptotically distributed as a χ^2 with N degrees of freedom under the null, and as a non-central χ^2 with the same degrees of freedom and non-centrality parameter $\boldsymbol{\delta}' V^{-1}(\hat{\mathbf{a}})\boldsymbol{\delta}$ under the Pitman sequence of local alternatives $H_l : \mathbf{a} = \boldsymbol{\delta}/\sqrt{T}$ (see Newey and MacFadden (1994)). In contrast, W_T will diverge to infinity for fixed alternatives of the form $H_f : \mathbf{a} = \boldsymbol{\delta}$, which makes it a consistent test. In that case, we can use Theorem 1 in Geweke (1981) to show that

$$p\lim \frac{1}{T} W_T = \boldsymbol{\delta}' V^{-1}(\hat{\mathbf{a}})\boldsymbol{\delta}$$

coincides with Bahadur's (1960) definition of the approximate slope of the Wald test. This expression differs from the non-centrality parameter in that the covariance matrix is no longer evaluated under the null. However, since $V(\hat{\mathbf{a}})$ does not depend on $\mathbf{a}$ when the true distribution is elliptical for any of the estimators considered in the previous section, both comparison criteria coincide.

In addition, since $V(\hat{\mathbf{a}}_{GMM}) = \mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}_0)$ in view of (9), while $V(\hat{\mathbf{a}}_{IML}) = V(\hat{\mathbf{a}}_{FML}) = V(\hat{\mathbf{a}}_{HLV}) = \mathrm{M}_{ll}^{-1}(\boldsymbol{\eta}_0)\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}_0)$ in view of (7), we can use $\mathrm{M}_{ll}(\boldsymbol{\eta}_0)$ to measure the relative efficiency of the GMM-based test procedure regardless of the value of $\boldsymbol{\delta}$.

In this sense, Proposition 9 in Fiorentini and Sentana (2007) implies that $\mathrm{M}_{ll}(\boldsymbol{\eta}_0) = 1$ if and only if the true conditional distribution is indeed normal. Otherwise, $0 \leq \mathrm{M}_{ll}^{-1}(\boldsymbol{\eta}_0) < 1$. This means that while there is no asymptotic efficiency loss in estimating $\boldsymbol{\eta}$ when the true conditional distribution is Gaussian, the efficiency gains could be potentially very large for

[9] Another advantage of the Wald test, shared with the LM test, is that it is easy to robustify with respect to misspecification, unlike the LR test.

other elliptical distributions. In the multivariate student t case with $\nu_0 > 2$, in particular, the relative efficiency ratio becomes $(\nu_0 - 2)(\nu_0 + N + 2)/[\nu_0(\nu_0 + N)]$. For any given N, this ratio is monotonically increasing in ν_0, and approaches 1 from below as $\nu_0 \to \infty$, and 0 from above as $\nu_0 \to 2^+$. At the same time, this ratio is decreasing in N for a given ν_0, which reflects the fact that the student t information matrix is "increasing" in N. Figure 2a presents a plot of this efficiency ratio as a function of η for several values of N. Similarly, Figure 2b presents the efficiency ratio as a function of κ for different values of N in the case of the Kotz distribution, where we have obtained $\mathrm{M}_{ll}^{-1}(\kappa)$ from (6). In this sense, it is worth mentioning that the excess kurtosis coefficient of any elliptical distribution is bounded from below by $-2/(N+2)$, which is the excess kurtosis of a random vector that is uniformly distributed on the unit sphere. This explains why the lower limit of admissible values for κ gets closer and closer to 0 from below as N increases. Finally, Figure 2c contains the corresponding efficiency ratios for a two-component scale mixture of normals in which $\pi = \frac{1}{2}$ as a function of the relative variance parameter $\varkappa$. As expected, the GMM and ML/HLV estimators are equally efficient for $\varkappa = 1$, since in that case the mixture of normals is itself normal. Once again, though, the relative efficiency of the ML/HLV estimators increases as we move away from normality, the more so the bigger N is.

We can assess the power implications of such efficiency gains by computing the probability of rejecting the null hypothesis when it is false as a function of $\mathbf{a}$ under the assumption that the asymptotic non-central chi-square distributions of the Wald tests implied by (7) or (9) provide reliable rejection probabilities in finite samples. The results for $T = 500$ at the usual 5% level are plotted in Figure 3 under the fairly innocuous assumptions that $\mathbf{\Omega} = \mathbf{I}_N$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$ and $\mathbf{a} = a\ell_N$, with $\ell_N' = (1, \ldots, 1)'$ and $a \in [0, .2]$. We consider two examples of elliptical distributions whose $\mathrm{M}_{ll}(\boldsymbol{\eta})$ correspond to those of a student t with 8 and 20 degrees of freedom, respectively. Not surprisingly, the power of all tests increases as we depart from the null. Similarly, their power also increases with the number of series due to the lack of cross-sectional correlation of the regression residuals. More importantly, the power of the efficient tests is always larger than the power of the GMM tests, although the differences are unsurprisingly small when the true distribution is not too far away from the normal.

In empirical applications, it is customary to pay attention not only to the joint Wald test of $H_0 : \mathbf{a} = \mathbf{0}$, but also to individual tests of the form $H_0 : a_i = 0$ for some i between 1 and N. Given that the asymptotic power of such partial tests under either local or fixed alternatives will depend on the non-centrality parameter $a_i^2/V(\hat{a}_i)$, the discussion in the previous paragraphs applies directly to those individual Wald tests too (see Sentana (2008) for a discussion on the advantages and disadvantages of joint versus individual tests on these contexts).

3.2 Misspecified elliptical distributions for the innovations

In this section, we derive the asymptotic distribution of the infeasible and feasible ML estimators introduced in section 2.2.1 when the true conditional distribution of $\mathbf{r}_t$ given r_{Mt} and their past is *i.i.d.* elliptical, but does not coincide with the distribution assumed for estimation purposes. For the sake of concreteness, we assume in what follows that the feasible and infeasible (pseudo) ML estimators are based on the erroneous assumption that $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1};\boldsymbol{\theta},\eta \sim i.i.d.\ t(\mathbf{0}, \mathbf{I}_N, \nu_0)$. Nevertheless, our results can be trivially extended to any other spherically-based likelihood estimators, as the only advantage of the student t likelihood four our purposes is the fact that its limiting relationship to the Gaussian distribution can be made explicit. In this context, the "infeasible" t-based PML estimator should be understood as the one that fixes the parameter η to some arbitrary value $\bar{\eta}$ between 0 and $\frac{1}{2}$.

For simplicity, we shall also define the pseudo-true values of $\boldsymbol{\theta}$ and η as consistent roots of the expected t pseudo log-likelihood score, which under appropriate regularity conditions will maximise the expected value of the t pseudo log-likelihood function. Specifically, if we define the pseudo-true values of $\boldsymbol{\phi}$ as the values of $\mathbf{a}, \mathbf{b}, \boldsymbol{\Omega}$, and η that will set to zero the expected value of the score vector, $\mathbf{s}_t(\boldsymbol{\phi}_0)$, where the expected value is taken with respect to the true distribution of the data, then we can derive the following result, which particularises to our context Proposition 15 in Fiorentini and Sentana (2007):

Proposition 3 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1};\boldsymbol{\varphi}_0$ in (1) is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ but not t and $\kappa_0 \leq 0$, where $\boldsymbol{\varphi}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\omega}_0, \boldsymbol{\varrho}_0)$, then:*

1. *The pseudo-true value of feasible student t-based ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \eta)'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\gamma}_\infty$ and $\boldsymbol{\omega}_\infty$ are equal to their corresponding true values $\boldsymbol{\gamma}_0$ and $\boldsymbol{\omega}_0$, respectively, and $\eta_\infty = 0$.*

2. $\sqrt{T}(\hat{\boldsymbol{\gamma}}_{FML} - \hat{\boldsymbol{\gamma}}_{GMM}) = o_p(1)$.

Intuitively, the reason is that since η must be estimated subject to the non-negativity restriction $\eta \geq 0$, the most platykurtic student t distribution that one can obtain is the normal distribution, in which case the feasible student t-based PML estimator coincides with the GMM one.

The following result derives the asymptotic distribution of the feasible t-based PML estimator of $\boldsymbol{\theta}$ in the more realistic case of leptokurtic disturbances. To keep the algebra simple, we will reparametrise $\boldsymbol{\Omega}$ as $\tau\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})$, so that $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\upsilon}, \tau)$, where $\boldsymbol{\upsilon}$ are $N(N+1)/2-1$ parameters that ensure that $|\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})| = 1\ \forall\boldsymbol{\upsilon}$. In other words, our reparametrisation will be such that

$$\tau = |\boldsymbol{\Omega}|^{1/N} \tag{11}$$

and

$$\boldsymbol{\Upsilon}(\boldsymbol{\upsilon}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N}. \tag{12}$$

Nevertheless, the t-based ML estimator of $\boldsymbol{\gamma}$ will be unaffected by this change.

Proposition 4 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ but not t with $\kappa_0 > 0$, where $\boldsymbol{\varphi}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_0, \boldsymbol{\varrho}_0)$, then:*

1. *The pseudo-true value of feasible student-t based ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \boldsymbol{\upsilon}, \tau, \eta)'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\gamma}_\infty$ and $\boldsymbol{\upsilon}_\infty$ are equal to their corresponding true values $\boldsymbol{\gamma}_0$ and $\boldsymbol{\upsilon}_0$.*

2. *$\mathcal{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ and $\mathcal{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ will be block diagonal between $(\boldsymbol{\gamma}, \boldsymbol{\upsilon})$ and (τ, η), where both*

$$\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0]$$

and

$$\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0]$$

will share the structure of $\mathcal{I}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ in Proposition 1, with

$$\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = E\left\{\delta^2[\varsigma_t(\boldsymbol{\vartheta}), \eta] \cdot [\varsigma_t(\boldsymbol{\vartheta})/N] \big| \boldsymbol{\varphi}\right\}$$

$$\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = N(N+2)^{-1}\left[1 + V\left\{\delta[\varsigma_t(\boldsymbol{\gamma}, \boldsymbol{\upsilon}, \tau), \eta] \cdot [\varsigma_t(\boldsymbol{\vartheta})/N] | \boldsymbol{\varphi}\right\}\right],$$

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

$$\mathcal{O}_{\eta\eta}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = V[\, s_{\eta t}(\boldsymbol{\phi}) | \boldsymbol{\varphi}],$$

$$\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}), \eta]/\partial\varsigma \cdot [\varsigma_t(\boldsymbol{\vartheta})/N] + \delta[\varsigma_t(\boldsymbol{\theta}), \eta] | \boldsymbol{\varphi}\right\},$$

$$\mathrm{M}_{ss}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}), \eta]/\partial\varsigma \cdot \varsigma_t^2(\boldsymbol{\vartheta})/[N(N+2)] \big| \boldsymbol{\varphi}\right\} + 1,$$

$$\mathrm{M}_{sr}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E\left\{[\varsigma_t(\boldsymbol{\vartheta})/N] \cdot \partial\delta[\varsigma_t(\boldsymbol{\vartheta}), \eta]/\partial\eta | \boldsymbol{\varphi}\right\},$$

$$\mathcal{H}_{\eta\eta}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E[\, h_{\eta\eta t}(\boldsymbol{\phi}) | \boldsymbol{\varphi}].$$

Intuitively, what this proposition shows is that

$$
\begin{aligned}
E\left\{\mathbf{s}_{\boldsymbol{\gamma}t}[\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_\infty(\eta), \eta] \middle| \boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_0, \boldsymbol{\varrho}_0\right\} &= \mathbf{0}, \\
E\left\{\mathbf{s}_{\boldsymbol{\upsilon}t}[\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_\infty(\eta), \eta] \middle| \boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_0, \boldsymbol{\varrho}_0\right\} &= \mathbf{0},
\end{aligned}
$$

for any elliptical distribution for the innovations, which implies in particular that the t-based PML estimators of $\mathbf{a}$ and $\mathbf{b}$ will be consistent. In contrast, when $\kappa > 0$ we cannot find any distribution for $\boldsymbol{\varepsilon}_t^*$ other than the multivariate t for which

$$
\begin{aligned}
E\left[s_{\tau t}(\boldsymbol{\phi})|\boldsymbol{\varphi}_0\right] &= 0, \\
E\left[s_{\eta t}(\boldsymbol{\phi})|\boldsymbol{\varphi}_0\right] &= 0,
\end{aligned}
$$

which means that the overall scale parameter τ will be inconsistently estimated.

The asymptotic distribution of the feasible t-based PML estimator of $\boldsymbol{\gamma}$ follows immediately from Proposition 4:

Corollary 2 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ but not t with $\kappa_0 > 0$, where $\boldsymbol{\varphi}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \lambda_0, \boldsymbol{\varrho}_0)$, then:*

$$
\sqrt{T}\left(\hat{\boldsymbol{\gamma}}_{FML} - \boldsymbol{\gamma}_0\right) \to N\left[\mathbf{0}, \frac{\mathrm{M}_{ll}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)}{\lambda_\infty\left[\mathrm{M}_{ll}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\right]^2} \cdot \mathcal{C}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\varphi}_0)\right], \tag{13}
$$

where $\lambda_\infty = \tau_0/\tau_\infty$.

The analysis of the "infeasible" t-based PML estimator, which fixes η to some value $\bar{\eta}$, is entirely analogous, except for the fact that the pseudo-true value of τ becomes $\tau_\infty(\bar{\eta})$, as opposed to $\tau_\infty = \tau_\infty(\eta_\infty)$.

A natural question in this context is a comparison of the efficiency of the t-based pseudo ML estimator and the GMM estimator when the distribution is elliptical but not t. We answer this question by assuming that the conditional distribution is either normal, Kotz, or the two-component scale mixture of normals discussed in section 2.1. It turns out that in all three cases we can obtain analytical expressions for the inefficiency ratio $\mathrm{M}_{ll}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}_{ll}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)]^2\}$ (see Appendix B).

The top panels of Figure 4 present the relative efficiency of these two estimators of $\boldsymbol{\gamma}$ as a function of $\bar{\eta}$ for four cross-sectional dimensions, while the bottom panels contain the corresponding pseudo-true values of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$. In addition, the straight lines indicate the position of the pseudo-true values when we also estimate η. As expected, if the true conditional distribution is Gaussian (Figure 4a), then the "infeasible" ML estimator that makes the erroneous assumption that it is a student t with $\bar{\eta}^{-1}$ degrees of freedom is inefficient relative to the GMM estimator, the more so the larger the value of $\bar{\eta}$. Nevertheless, this inefficiency becomes smaller and less sensitive to $\bar{\eta}$ as the number of assets increases. But of course $\eta_\infty = 0$ in this case in view of Proposition 3, which suggests that estimating η is clearly beneficial under misspecification. In fact, the "infeasible" t-based PML estimator seems to be strictly more efficient than the GMM one at the pseudo-true value of η when the true conditional distribution is leptokurtic. This is indeed true for any value of $\bar{\eta}$ for a Kotz distribution with $\kappa_0 = 1/8$ (Figure 4c), which is equal to the excess kurtosis of a t with 20 degrees of freedom, as well as for a two-component mixture of normals with $\pi = 1/2$ and $\kappa_0 = 1/4$ (Figure 4e), which coincides with the excess kurtosis of the more empirically realistic t distribution with 12 degrees of freedom. It is notheworthy that as N increases the "infeasible" t-based PML estimator tends to achieve the full efficiency of the ML estimator for any $\bar{\eta} > 0$.[10] Whether such efficiency gains always accrue when η is estimated is left for future research.[11]

3.3 Asymmetric innovations

To focus our discussion, we assume in this section that $\boldsymbol{\varepsilon}_t^*$ is distributed as an *i.i.d.* multivariate asymmetric t. Following Mencía and Sentana (2008), if we choose

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\beta}\left[\xi_t^{-1} - c(\boldsymbol{\beta},\eta)\right] + \sqrt{\frac{\zeta_t}{\xi_t}}\boldsymbol{\Xi}^{1/2}\mathbf{u}_t \tag{14}$$

where $\mathbf{u}_t$ is uniformly distributed on the unit sphere in $\mathbb{R}^N$, ζ_t is a χ^2 random variable with N degrees of freedom, ξ_t is Gamma random variable with parameters $(2\eta)^{-1}$ and $\delta^2/2$ with $\delta = (1-2\eta)\eta^{-1}c(\boldsymbol{\beta},\eta)$, $\boldsymbol{\beta}$ is a $N \times 1$ parameter vector, and $\boldsymbol{\Xi}$ is a $N \times N$ positive definite matrix given by

$$\boldsymbol{\Xi} = \frac{1}{c(\boldsymbol{\beta},\eta)}\left[I_N + \frac{c(\boldsymbol{\beta},\eta) - \mathbf{1}}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right],$$

with

$$c(\boldsymbol{\beta},\eta) = \frac{-(1-4\eta) + \sqrt{(1-4\eta)^2 + 8\boldsymbol{\beta}'\boldsymbol{\beta}(1-4\eta)\eta}}{4\boldsymbol{\beta}'\boldsymbol{\beta}\eta},$$

[10] The values corresponding to $N = \infty$ in Figures 4 and 5 are intended to reflect the maximum efficiency gains that could be obtained by increasing the number of series; and hence, they are derived under sequential limits, i.e., T converges to infinity with a fixed N and then N converges to infinity. In this sense, $\lim_{N\to\infty}\mathrm{M}_{ll}(\boldsymbol{\eta}) = (1-\kappa)^{-1}$ in the case of Kotz innovations and discrete scale mixture of normals innovations.

[11] Another pending issue is whether η_∞ is always larger than $\max(0,\kappa_0)/[4\max(0,\kappa_0)+2]$, which is the value of η that matches the excess kurtosis of the t distribution with the excess kurtosis of the true distribution, as Figures 4b and 4c seem to suggest.

then $E[\boldsymbol{\varepsilon}_t^*] = \mathbf{0}$ and $V[\boldsymbol{\varepsilon}_t^*] = \mathbf{I}_N$. In this sense, note that $\lim_{\boldsymbol{\beta}'\boldsymbol{\beta}\longrightarrow 0} c(\boldsymbol{\beta},\eta) = 1$, so that the above distribution collapses to the usual multivariate symmetric t when $\boldsymbol{\beta} = \mathbf{0}$. Therefore, we allow for asymmetries by introducing the vector of parameters $\boldsymbol{\beta}$.

To study the consistency of the symmetric t-based PML estimator when the DGP is asymmetric, it is once again convenient to look at its score. Specifically, given the definition of (14), we can write

$$\mathbf{s}_{\mathbf{a}t}(\boldsymbol{\gamma}_0, \boldsymbol{\omega}_0, \eta) = \boldsymbol{\Omega}_0^{-1/2} \frac{N\eta + 1}{1 - 2\eta + \eta(\zeta_t/\xi_t)} \left\{ \boldsymbol{\beta}\left[\xi_t^{-1} - c(\boldsymbol{\beta},\eta)\right] + \sqrt{\frac{\zeta_t}{\xi_t}} \boldsymbol{\Xi}^{1/2}\mathbf{u}_t \right\} \tag{15}$$

The expected value of $\boldsymbol{\varepsilon}_t^*$ in (14) is clearly zero by construction. Similarly, the expected value of (15) is also zero when $\boldsymbol{\beta}_0 = \mathbf{0}$ since $\mathbf{u}_t$ and (ζ_t/ξ_t) are independent. But when $\boldsymbol{\beta}_0 \neq 0$, the expected value of (15) will be generally different from zero because ξ_t^{-1} appears both in the numerator and denominator. Consequently, the mean parameters $\mathbf{a}$ will be inconsistently estimated. In contrast, $\mathbf{b}$ will be consistently estimated precisely because the estimator of $\mathbf{a}$ will fully mop up the bias in the mean. More formally, re-write model (1) as

$$\mathbf{r}_t = \boldsymbol{\Omega}^{1/2}\boldsymbol{\delta} + \mathbf{b}r_{Mt} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\Omega}^{-1/2}\mathbf{a} = \boldsymbol{\delta}$. This homeomorphic reparametrisation satisfies the conditions of Proposition 17 in Fiorentini and Sentana (2007), which implies the consistency of $\mathbf{b}$. Unfortunately, mean-variance efficiency tests are based on $\mathbf{a}$, not $\mathbf{b}$.

For analogous reasons, the HLV estimator of $\mathbf{a}$ also becomes inconsistent under asymmetry. Intuitively, the problem is that it will not be true any more that the N-dimensional density of $\boldsymbol{\varepsilon}_t^*$ could be written as a function of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$ alone. Therefore, a semiparametric estimator of $\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi}_0)$ that combines the elliptical symmetry assumption with a non-parametric specification for $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ will be contaminated by the skewness of the data.

In contrast, the GMM estimator always yields a consistent estimator of $\mathbf{a}$, on the basis of which we can develop a GMM-based Wald test with the correct asymptotic size since (9) remains valid under asymmetry.

3.4 Elliptical distributions for returns

In this section we explicitly study the framework analysed by MacKinlay and Richardson (1991) and Kan and Zhou (2006), who considered a *joint* distribution of excess returns for the N assets $\mathbf{r}_t$ and the reference portfolio, r_{Mt}. When the joint distribution of $\mathbf{R}_t$ is *i.i.d.*

Gaussian, the distribution of $\mathbf{r}_t$ conditional on r_{Mt} must also be normal, with a mean $\mathbf{a}+\mathbf{b}r_{Mt}$ that is a linear function of r_{Mt}, and a covariance matrix $\mathbf{\Omega}$ that does not depend on r_{Mt}. However, while the linearity of the conditional mean will be preserved when $\mathbf{R}_t$ is elliptically distributed but non-Gaussian, the conditional covariance matrix will no longer be independent of r_{Mt}. For instance, if we assume that $\mathbf{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})] \sim i.i.d.\ t(\mathbf{0}, \mathbf{I}_{N+1}, \eta)$, where

$$\boldsymbol{\mu}(\boldsymbol{\rho}) = \begin{pmatrix} \mu_M \\ \mathbf{a}+\mathbf{b}\mu_M \end{pmatrix}, \tag{16}$$

$$\mathbf{\Sigma}(\boldsymbol{\rho}) = \begin{pmatrix} \sigma_M^2 & \sigma_M^2\mathbf{b}' \\ \sigma_M^2\mathbf{b} & \sigma_M^2\mathbf{b}\mathbf{b}'+\mathbf{\Omega} \end{pmatrix}, \tag{17}$$

and $\boldsymbol{\rho}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}', \mu_M, \sigma_M^2)$, then

$$\begin{aligned} E\left[\mathbf{r}_t|r_{Mt};\boldsymbol{\rho},\eta\right] &= \mathbf{a}+\mathbf{b}r_{Mt}, \\ V\left[\mathbf{r}_t|r_{Mt};\boldsymbol{\rho},\eta\right] &= \left(\frac{\nu-2}{\nu-1}\right)\left[1+\frac{(r_{Mt}-\mu_M)^2}{(\nu-2)\,\sigma_M^2}\right]\mathbf{\Omega} \equiv \mathbf{\Psi}_t(\boldsymbol{\rho},\eta), \end{aligned}$$

which means that model (1) will be misspecified due to contemporaneous, conditionally heteroskedastic innovations. In other words, the variances and covariances of the regression residuals will be a function of the regressor.

As MacKinlay and Richardson (1991) pointed out, the GMM estimator of $\boldsymbol{\gamma}$ remains consistent in this case. In addition, we know from Lemma D3 in Peñaranda and Sentana (2004) that if $\mathbf{R}_t$ is independently and identically distributed as an elliptical random vector with mean $\boldsymbol{\mu}(\boldsymbol{\rho})$, covariance matrix $\mathbf{\Sigma}(\boldsymbol{\rho})$, and bounded fourth moments, then the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{m}}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}_0\right)$ will be given by

$$\mathbf{S}_U(\boldsymbol{\gamma}_0) = \begin{pmatrix} 1 & \mu_{M0} \\ \mu_{M0} & (\kappa_0+1)\,\sigma_{M0}^2+\mu_{M0}^2 \end{pmatrix} \otimes \mathbf{\Omega}_0,$$

where $\bar{\mathbf{m}}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}_0\right)$ is the sample mean of $\mathbf{m}_U\left(\mathbf{R}_t;\boldsymbol{\gamma}_0\right)$ in (8). Hence,

$$V(\hat{\mathbf{a}}_{GMM}) = \left[1+\frac{\mu_{M0}^2}{\sigma_{M0}^2}\,(1+\kappa_0)\right]\mathbf{\Omega}_0. \tag{18}$$

In this sense, note that the only difference with respect to (9) is that the square Sharpe ratio of the reference portfolio μ_{M0}^2/σ_{M0}^2 is multiplied by the factor $(1+\kappa_0)$. In practice, we will estimate $V(\hat{\mathbf{a}}_{GMM})$ by using heteroskedastic robust standard errors a la White (1980). Specifically, we should use the sandwich expression $\mathcal{C}_{\gamma\gamma}(\boldsymbol{\phi}) = \mathcal{A}_{\gamma\gamma}^{-1}(\boldsymbol{\phi})\mathcal{B}_{\gamma\gamma}(\boldsymbol{\phi})\mathcal{A}_{\gamma\gamma}^{-1}(\boldsymbol{\phi})$, but this time with

$$\hat{\mathcal{B}}_{\gamma\gamma}(\boldsymbol{\phi}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\gamma t}(\boldsymbol{\theta};\mathbf{0})\mathbf{s}'_{\gamma t}(\boldsymbol{\theta};\mathbf{0}) \tag{19}$$

while we will continue to use

$$\hat{\mathcal{A}}_{\gamma\gamma}(\boldsymbol{\phi}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} 1 & r_{Mt} \\ r_{Mt} & r_{Mt}^2 \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}. \tag{20}$$

At the other extreme of the efficiency range, we can consider the joint ML estimator that makes the correct assumption that $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})] \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_{N+1}, \boldsymbol{\eta})$, whose asymptotic distribution can be obtained from the following result:

Proposition 5 *Let $\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t(\boldsymbol{\rho})$, where $\boldsymbol{\epsilon}_t(\boldsymbol{\rho}) = \mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})$, $\boldsymbol{\mu}(\boldsymbol{\rho})$ and $\boldsymbol{\Sigma}(\boldsymbol{\rho})$ are defined in (16) and (17), respectively, and $\boldsymbol{\rho}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}', \mu_M, \sigma_M^2)$. If $\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}_0)|I_{t-1}; \boldsymbol{\rho}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_{N+1}, \boldsymbol{\eta}_0)$ with density $\exp[c_{N+1}(\boldsymbol{\eta}) + g_{N+1}(\varsigma_t, \boldsymbol{\eta})]$, then the only non-zero elements of the information matrix will be:*

$$\begin{aligned}
\mathcal{I}_{\gamma\gamma}(\boldsymbol{\phi}) &= \left[\mathrm{M}_{ll}(\boldsymbol{\eta})\begin{pmatrix} 1 & \mu_M \\ \mu_M & \mu_M^2 \end{pmatrix} + \mathrm{M}_{ss}(\boldsymbol{\eta}_0)\begin{pmatrix} 0 & 0 \\ 0 & \sigma_M^2 \end{pmatrix}\right] \otimes \boldsymbol{\Omega}^{-1}, \\
\mathcal{I}_{\boldsymbol{\omega\omega}}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{ss}(\boldsymbol{\eta})}{2}\mathbf{D}_N'\left[\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right]\mathbf{D}_N + \frac{\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1}{4}\mathbf{D}_N'\left[vec(\boldsymbol{\Omega}^{-1})vec'(\boldsymbol{\Omega}^{-1})\right]\mathbf{D}_N', \\
\mathcal{I}_{\mu_M\mu_M}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{ll}(\boldsymbol{\eta})}{\sigma_M^2}, \\
\mathcal{I}_{\sigma_M^2\sigma_M^2}(\boldsymbol{\phi}) &= \frac{3\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1}{4\sigma_M^4}, \\
\mathcal{I}_{\boldsymbol{\omega}\sigma_M^2}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1}{4\sigma_M^2}\mathbf{D}_N' vec(\boldsymbol{\Omega}^{-1}), \\
\mathcal{I}_{\boldsymbol{\omega\eta}}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{sr}(\boldsymbol{\eta})}{2}\mathbf{D}_N' vec(\boldsymbol{\Omega}^{-1}), \\
\mathcal{I}_{\sigma_M^2\boldsymbol{\eta}}(\boldsymbol{\phi}) &= \frac{\mathrm{M}_{sr}(\boldsymbol{\eta})}{2\sigma_M^2}, \\
\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}) &= V[\,\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = -E[\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi})|\,\boldsymbol{\phi}],
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{M}_{ll}(\boldsymbol{\eta}) &= E\left\{\delta_{N+1}^2[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\frac{\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho})}{N+1}\right|\boldsymbol{\phi}\Bigg\} \\
&= E\left\{\frac{2\partial\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]}{\partial\varsigma}\frac{\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho})}{N+1} + \delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\right|\boldsymbol{\phi}\Bigg\}, \\
\mathrm{M}_{ss}(\boldsymbol{\eta}) &= \frac{N+1}{N+3}\left[1 + V\left\{\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\frac{\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho})}{N+1}\right|\boldsymbol{\phi}\Bigg\}\right] \\
&= E\left\{\frac{2\partial\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]}{\partial\varsigma}\frac{[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho})]^2}{(N+1)(N+3)}\right|\boldsymbol{\phi}\Bigg\} + 1, \\
\mathrm{M}_{sr}(\boldsymbol{\eta}) &= E\left[\left\{\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\frac{\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho})}{N+1} - 1\right\}\mathbf{e}_{rt}'(\boldsymbol{\phi})\right|\boldsymbol{\phi}\Bigg] \\
&= -E\left\{\frac{\varsigma_t(\boldsymbol{\theta})}{N+1}\frac{\partial\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'}\right|\boldsymbol{\phi}\Bigg\},
\end{aligned}$$

and the subscript $N+1$ in δ emphasises the cross-sectional dimension.

Specifically, we can use this Proposition to extend the result in equation (31) in Kan and Zhou (2006) and show that

$$V(\hat{\mathbf{a}}_{JML}) = \frac{1}{\mathrm{M}_{ll}(\boldsymbol{\eta}_0)}\left[1 + \frac{\mathrm{M}_{ll}(\boldsymbol{\eta}_0)}{\mathrm{M}_{ss}(\boldsymbol{\eta}_0)}\frac{\mu_{M0}^2}{\sigma_{M0}^2}\right]\boldsymbol{\Omega}_0, \tag{21}$$

where $\hat{\boldsymbol{\theta}}_{JML}$ denotes the joint ML estimator that makes the correct assumption that $\boldsymbol{\Sigma}^{-1/2}(\rho)[\mathbf{R}_t - \boldsymbol{\mu}(\rho)] \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_{N+1}, \boldsymbol{\eta})$, and both $\mathrm{M}_{ll}(\boldsymbol{\eta}_0)$ and $\mathrm{M}_{ss}(\boldsymbol{\eta}_0)$ correspond to this $(N+1)$-dimensional distribution. However, $\hat{\mathbf{a}}_{JML}$ assumes omniscience on the part of the researcher, which is unrealistic.

The following proposition shows the consistency of the t-based estimators which make the erroneous assumption that $V[\mathbf{r}_t|r_{Mt}] = \tau\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})$, where τ and $\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})$ are defined in (11) and (12), and provides expressions for the conditional variance of the score and expected Hessian matrix under such misspecification:

Proposition 6 *If $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})]|I_{t-1}; \boldsymbol{\varphi}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_{N+1}, \boldsymbol{\varrho}_0)$ with $\kappa_0 > 0$, where $\boldsymbol{\mu}(\boldsymbol{\rho})$ and $\boldsymbol{\Sigma}(\boldsymbol{\rho})$ are defined in (16) and (17) respectively, $\boldsymbol{\rho}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}', \mu_M, \sigma_M^2)$ and $\boldsymbol{\varphi} = (\boldsymbol{\rho}', \boldsymbol{\varrho}')'$, then:*

1. *The pseudo-true value of feasible student-t based PML estimator of $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \boldsymbol{\upsilon}, \tau, \eta)'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\gamma}_\infty$ and $\boldsymbol{\upsilon}_\infty$ are equal to their corresponding true values $\boldsymbol{\gamma}_0$ and $\boldsymbol{\upsilon}_0$.*

2. *$\mathcal{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ and $\mathcal{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)|\boldsymbol{\varphi}_0]$ will be block diagonal between $(\boldsymbol{\gamma}, \boldsymbol{\upsilon})$ and (τ, η), where both*

$$\mathcal{O}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0]$$

and

$$\mathcal{H}_t(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0]$$

will share the structure of $\mathcal{I}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ in Proposition 1, with

$$
\begin{aligned}
\mathrm{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\rho}),\eta]\cdot[\varsigma_t(\boldsymbol{\rho})/N]\right|\boldsymbol{\varphi}\Big\} \\
\mathrm{M}^O_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= N(N+2)^{-1}\left[1+V\left\{\delta[\varsigma_t(\boldsymbol{\rho}),\eta]\cdot[\varsigma_t(\boldsymbol{\rho})/N]\right|\boldsymbol{\varphi}\right\}\right], \\
\mathrm{M}^O_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= E\left[\left\{\delta[\varsigma_t(\boldsymbol{\rho}),\eta]\cdot[\varsigma_t(\boldsymbol{\rho})/N]-1\right\}s_{\eta t}(\boldsymbol{\phi})\right|\boldsymbol{\varphi}\right], \\
\mathcal{O}_{\eta\eta}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= V[\,s_{\eta t}(\boldsymbol{\phi})|\,\boldsymbol{\varphi}], \\
\mathrm{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\rho}),\eta]/\partial\varsigma\cdot[\varsigma_t(\boldsymbol{\rho})/N]+\delta[\varsigma_t(\boldsymbol{\theta}),\eta]\right|\boldsymbol{\varphi}\right\}, \\
\mathrm{M}^H_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\rho}),\eta]/\partial\varsigma\cdot\varsigma_t^2(\boldsymbol{\rho})/[N(N+2)]\right|\boldsymbol{\varphi}\Big\}+1, \\
\mathrm{M}^H_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= -E\left\{[\varsigma_t(\boldsymbol{\rho})/N]\cdot\partial\delta[\varsigma_t(\boldsymbol{\rho}),\eta]/\partial\eta|\boldsymbol{\varphi}\right\}, \\
\mathcal{H}_{\eta\eta}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= -E[\,h_{\eta\eta t}(\boldsymbol{\phi})|\,\boldsymbol{\varphi}].
\end{aligned}
$$

The top panel of Figure 5 presents the efficiency of the t-based PML estimators of $\boldsymbol{\gamma}$ in relation to the corresponding GMM estimator as a function of $\bar{\eta}$ when $\mathbf{R}_t$ is distributed as a multivariate t with 8 degrees of freedom ($\eta_0 = .125$) for three cross-sectional dimensions, while the bottom panel contains the corresponding pseudo-true values of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$. In addition, the vertical straight lines in the top panel indicate the position of the pseudo-true values η_∞ when we also estimate this parameter, while the horizontal ones describe the efficiency of the GMM estimator of γ in (18) relative to that of the joint ML estimator in (21).[12] As in Figures 4b and 4c, the "infeasible" t-based PML estimator of $\boldsymbol{\gamma}$ is more efficient than the GMM estimator for all values of $\bar{\eta}$, the more so the larger N is. Furthermore, the feasible t-based PML estimator that also estimates η gets close to achieving the full efficiency of the joint ML estimator, especially for large N. Finally, another noteworthy fact is the very small asymptotic bias of the t-based PML estimator of η.

In principle, Proposition 6, and in particular the block diagonal structure of $\mathcal{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will continue to hold if we replace the t-based ML estimator by any other estimator based on a specific *i.i.d.* elliptical distribution for $\mathbf{r}_t|r_{Mt}$. But since the HLV estimator is asymptotically equivalent to a parametric estimator that uses a flexible elliptical distribution as we increase the number of parameters, Proposition 6 suggests that the HLV estimator of $\boldsymbol{\gamma}$ will continue to be consistent. In fact, an argument analogous to the one made by Hodgson (2000) in a closely related univariate context would imply that the HLV estimator is as efficient as the parametric estimator that used the true *unconditional* distribution of the innovations

[12] These graphs are based on the expressions in Proposition 6, with the relevant expectations computed by Monte Carlo integration with 10^6 drawings.

$\boldsymbol{\varepsilon}_t = \mathbf{r}_t - \mathbf{a}_0 - \mathbf{b}_0 r_{Mt}$. Nevertheless, inferences about $\mathbf{a}$ and $\mathbf{b}$ would have to be adjusted to reflect the contemporaneous conditional heteroskedasticity of $\boldsymbol{\varepsilon}_t$.

4 Monte Carlo analysis

In this section we assess the finite sample size and power properties of the GMM, HLV and feasible t-based ML test statistics of the null hypothesis $H_0 : \mathbf{a} = \mathbf{0}$ for six different distributional assumptions for the innovations, namely Gaussian, student-t with 8 and 4 degrees of freedom, Kotz with $\kappa = 1/8$, two-component scale mixture of normals with the same kurtosis, and asymmetric-t innovations.[13] We also consider a t_8 and t_4 distributional assumptions for the returns, $\mathbf{R}_t$. In all cases, we carry out 10,000 replications with $T = 500$, $N = 5$, $\boldsymbol{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$ and $\mathbf{b} = \mathbf{0}$ both under the null hypothesis, and under the alternative that $\mathbf{a} = 4\mu_M \times \boldsymbol{\ell}_5$.[14]

We sample Gaussian and Student t random numbers using standard MATLAB routines. To sample the Kotz innovations, we exploit the fact that $\boldsymbol{\varepsilon}_t^* = \sqrt{\xi_t}\mathbf{u}_t$, where ξ_t is a univariate Gamma with mean N and variance $N[(N+2)\kappa+2]$. Similarly, we use (2) to sample the discrete mixture of normals. Finally, to draw asymmetric t innovations we first generate a univariate Gamma and N independent standard Gaussian variates, and then use the decomposition presented in (14).

As mentioned in section 2.2.2, the GMM estimators of $\boldsymbol{\gamma}$ coincide with the equation by equation coefficient estimates in the OLS regression of r_{it} on a constant and r_{Mt}. Similarly, a GMM estimator of $\boldsymbol{\Omega}$ can be easily obtained from the covariance matrix of the OLS regression residuals, as explained in footnote 5. We use the expressions in Proposition 2 to compute its covariance matrix under the maintained assumption of *i.i.d.* innovations. In contrast, we combine (19) with (20) to obtain heterokedasticity robust standard errors.

Following Fiorentini, Sentana and Calzolari (2003), we obtain a consistent estimator of the reciprocal degrees of freedom parameter η on the basis of the GMM estimators as

$$\hat{\eta}_{SMM} = \frac{\max[0, \bar{\kappa}_T(\hat{\boldsymbol{\theta}}_{GMM})]}{4\max[0, \bar{\kappa}_T(\hat{\boldsymbol{\theta}}_{GMM})] + 2}, \tag{22}$$

[13] In these cases, a sample of r_{Mt} is drawn from a Gaussian distribution for each replication.

[14] The value of $\mathbf{b}$ does not affect the asymptotic distribution of the different estimators of $\mathbf{a}$ and the corresponding test procedures, while the value of σ_M simply scales up or down all the return series, and consequently $\boldsymbol{\Omega}$, μ_M and $\mathbf{a}$.

where

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

is Mardia's (1970) sample coefficient of multivariate excess kurtosis of the estimated standardised residuals. Then, we use $\hat{\eta}_{SMM}$ as initial value to obtain the sequential ML estimator of η proposed by Fiorentini and Sentana (2007), $\hat{\eta}_{SML}$ say, which maximises the t-based log-likelihood function with respect to η keeping $\boldsymbol{\theta}$ fixed at $\hat{\boldsymbol{\theta}}_{GMM}$.

Having obtained $\hat{\boldsymbol{\theta}}_{GMM}$ and $\hat{\eta}_{SML}$, we compute a one-step ML estimator of $\boldsymbol{\theta}$ by means of the BHHH correction

$$\left[\sum_{t=1}^{T}\mathbf{s}_{\gamma t}(\boldsymbol{\theta})\mathbf{s}'_{\gamma t}(\boldsymbol{\theta})\right]^{-1}\sum_{t=1}^{T}\mathbf{s}_{\gamma t}(\boldsymbol{\theta}), \tag{23}$$

with the analytical expressions for the t-score derived in section 2.2.1.[15] Next, we carry out a few EM iterations over $\boldsymbol{\theta}$ using this one-step ML estimator as initial value (see Appendix C), and finally switch to a quasi-Newton procedure until convergence. The (non-robust) asymptotic covariance matrix is computed using the expressions in Proposition 1, while for the robust standard errors we use the expressions in Proposition 4.

As for the HLV estimator and its asymptotic covariance matrix, we follow the computational approach described in Appendix B1 of Fiorentini and Sentana (2007).

4.1 Sampling distribution of the different estimators

Although we are mostly interested in the test statistics, it is convenient to study first the finite sample distributions of the estimators of $\mathbf{a}$, which are not affected by the estimation of their asymptotic covariances matrices.

In this sense, Figure 6 presents box-plots of the feasible t-based PML, HLV and GMM estimators for the eight different DGP's that we have considered. As usual, the central boxes describe the first and third quartiles of the sampling distributions, as well as their median. The maximum length of the whiskers is one interquartile range.

By and large, the behaviour of the different estimators is in accordance to what the asymptotic results would suggest. The only "surprises" are the fact that the dispersion of the distribution of the HLV estimator is systematically larger than the distribution of the ML estimator under correct specification of the latter, and that this result continues to hold even when the innovations follow a discrete mixture of normals. The other interesting results occur

[15] This one-step ML estimator is asymptotically equivalent to $\hat{\gamma}_{ML}$. An alternative asymptotically equivalent estimator of $\hat{\gamma}_{ML}$ will update the whole of $\hat{\boldsymbol{\theta}}_{GMM}$ by means of a simple BHHH correction based on $\mathbf{s}_{\boldsymbol{\theta}t}$.

when the joint distribution of $\mathbf{r}_t$ and r_{Mt} is elliptical, so that the conditional mean of $\mathbf{r}_t$ given r_{Mt} continues to be linear in r_{Mt} but the conditional variance is no longer constant. In this case not only the HLV and ML estimators of $\mathbf{a}$ remain consistent despite this misspecification, as we discussed in section 3.4, but they are also more efficient than the GMM estimator.

4.2 Sampling distribution of the associated test statistics

The first question that we need to address is whether the asymptotic distribution under the null attributed to the joint and individual Wald test statistics introduced in section 3.1 is reliable in finite samples. To do so, we employ the p-value discrepancy plots proposed by Davidson and MacKinnon (1998). Let w_j denote the simulated value of a given test statistic, and let p_j be the asymptotic p-value of w_j, that is the probability of observing a value of the test statistic at least as large as w_j according to its asymptotic distribution under the null. Let also $\hat{F}(x)$ for $x \in (0,1)$ be the empirical distribution function of p_j i.e. the sample proportion of $p_j's$ which are not greater than x. A p-value discrepancy plot is a plot of $[\hat{F}(x) - x]$ against x, i.e. a plot of the difference between actual and nominal size for a range of nominal sizes. If the candidate distribution for w_j is correct, then the p-value discrepancy should be close to zero.

The top left panels of Figures 7a-7h show p-value discrepancy plots of the joint tests of $H_0 : \mathbf{a} = \mathbf{0}$ for the eight DGPs that we have considered ("Wald statistics"), while the bottom left panels show the corresponding plots for the individual tests of $H_0 : a_i = 0$ ("t statistics"). The most striking fact that we find is that the HLV-based joint and individual tests have systematically the largest size distortions irrespective of whether the assumptions that justify them are correct. In contrast, the GMM tests that use expression (9) to compute the asymptotic weighting matrix have finite sample sizes that are close to their asymptotically equivalent in all cases, including when the correct expression should be (18). As for the tests that use the t-based PML estimator, there is also little to choose between the robust and non-robust versions, which are both well behaved even when the conditional distribution is heteroskedastic. The only exception seems to be the discrete mixture of normals example (Figure 7e), in which case the non-robust test is surprisingly better behaved than the robust one. As expected, though, when the distribution of the innovations is asymmetric (Figure 7f), the HLV and ML tests present considerable size distortions.

We can complement our finite sample analysis with *size-power curves*, which is another graphical method proposed by Davidson and MacKinnon (1998) to display the simulation evidence on the power of the different tests. We can define $\hat{F}^{*}(x)$ for $x \in (0,1)$ as the empirical distribution function of the asymptotic p-values under the null when the data are generated under the alternative. A size-power curve is a plot of test power versus actual test size for a range of test sizes.

The right panels of Figures 7a-7h show size-power curves for the same eight DGPs. Not surprisingly, the size-adjusted powers of the robust tests are very close to the corresponding non-robust tests in all cases. Contrary to the asymptotic results, though, GMM tests seem to have more power than the others under Gaussian innovations. In all other cases, in contrast, the HLV-based tests are more powerful than the GMM ones, but less so that the ones that use the feasible t-based PML estimator. In addition, the differences in power between HLV and t-based PML tests are very small in the case of Kotz and discrete mixture of normals innovations, despite the fact that the t-based estimator is suboptimal.

5 Empirical application

In this section we use the alternative estimators previously discussed to test the mean-variance efficiency of the US aggregate stock market portfolio using monthly data over the period July 1962 to June 2007 (540 observations). As for $\mathbf{r}_t$, we consider two different sets of $N = 5$ portfolios from Ken French's Data Library: one grouped by industry, and another one sorted by their book-to-market ratio. Specifically, each NYSE, AMEX, and NASDAQ stock is assigned to an industry portfolio at the end of June of year t based on its four-digit SIC code at the time.[16] Similarly, quintile portfolios are formed on BE/ME at the end of each June using NYSE breakpoints. The BE used in June of year t is the book equity for the last fiscal year end in $t-1$, while ME is price times shares outstanding at the end of December of $t-1$. The excess return on the market portfolio corresponds to the value weighted return measure on all NYSE, AMEX and NASDAQ stocks in CRSP, while the safe asset is the 1-month TBill return from Ibbotson and Associates (see <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html> for further details).

[16] Industry definitions: Cnsmr: Consumer Durables, NonDurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops). Manuf: Manufacturing, Energy, and Utilities. HiTec: Business Equipment, Telephone and Television Transmission. Hlth: Healthcare, Medical Equipment, and Drugs. Other: Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance.

The most obvious characteristic of these portfolios for our purposes is their leptokurtosis. The LM test of normality against the alternative of multivariate student t proposed by Fiorentini, Sentana and Calzolari (2003) yields a value of 3173.71 for the industry portfolios residuals from (1), and 1997.83 for the book to market ones. This confirms our empirical motivation for estimation and testing procedures that exploit such a prevalent feature of the data.

Table 1a presents the parameter estimates and (asymptotic) robust standard errors for the GMM, HLV and t-based ML estimators of model (1), while Table 1b reports the corresponding joint tests of $H_0 : \mathbf{a} = \mathbf{0}$. The results for industry portfolios indicate that the Student t-based test clearly rejects the efficiency of the market portfolio. They also show that the GMM test is borderline, while the HLV-based test fails to reject, which is in line with the results reported by Vorkink (2003).

Given the expressions for the test statistics in sections 2 and 3, the contradicting conclusions obtained with the ML and HLV tests must be due to three causes. First, the point estimates of $\mathbf{a}$ are somewhat different. Second, the point estimates of the idiosyncratic covariance matrix $\boldsymbol{\Omega}$ also differ, although even less so. More importantly, the scalar factors that multiply $\boldsymbol{\Omega}^{-1}$ are noticeably different too. In particular, they are 1.87 and 1.98 for the robust and non-robust versions of the ML tests, but only 1.43 for the HLV test. Both our Monte Carlo results and the results reported in Fiorentini and Sentana (2007) indicate the unreliable nature of the non-parametric estimates of M_{ll} in finite samples.

In contrast, all three tests reject the mean-variance efficiency of the market portfolio relative to the book-to-market sorted portfolios of Fama and French (1993). Still, we also find important differences in the estimates of the scalar factors mentioned in the previous paragraph.

As we saw in section 3.3, though, both parametric and semiparametric elliptically-based procedures are sensitive to the assumption of elliptical symmetry. For that reason, we follow Mencía and Sentana (2008), and test the null hypothesis of multivariate student t innovations against Generalised Hyperbolic (GH) alternatives, which include the multivariate asymmetric t distribution in (14) as a special case. In fact, the only difference between a GH distribution and an asymmetric t distribution is that in the second case the scalar mixing variable ξ_t can be any Generalised Inverse Gaussian (GIG) with parameters $\nu/2, \gamma$ and 1, or $\xi_t \sim GIG(\nu/2, \gamma, 1)$ for short (see Jørgensen (1982)). In this sense, a multivariate t distribution is obtained precisely when ξ_t is a Gamma random variable with parameters $(2\eta)^{-1}$ and $\delta^2/2$.

Mencía and Sentana (2008) proposed joint LM tests, as well as tests for asymmetry and kurtosis separately. Their kurtosis statistic tests that $(1+\gamma)^{-1} \equiv \psi = 1$ under the maintained hypothesis of $\boldsymbol{\beta} = \mathbf{0}$, where γ is the second tail shape parameter of the GIG distribution and $\boldsymbol{\beta}$ is the $N \times 1$ vector of coefficients that appears in (14). In effect, this amounts to testing that the tail behavior of the multivariate t distribution adequately reflects the kurtosis in the data. In turn, the asymmetry statistic tests that $\boldsymbol{\beta} = \mathbf{0}$ under the maintained assumption that $\psi = 1$.

However, since the ML estimates of η, which is the reciprocal of the degrees of freedom of the multivariate t distribution, in the two data sets that we consider are above .25, we must use the modified expressions that Mencia and Sentana (2008) suggest for this case.

Table 2 reports the values of the tests statistics for the industry and book-to-market sorted portfolios, together with their p-values. As can be seen, we cannot reject the null hypothesis that the distribution of $\mathbf{r}_t$ conditional on r_{Mt} is multivariate student t at conventional levels.

Finally, we perform a simple conditional homoskedasticity test by regressing the squared OLS residuals from the regression of r_{it} on a constant and r_{Mt}, $\hat{\varepsilon}_{it}^2$ say, on a constant, the market excess return r_{Mt} and its squared r_{Mt}^2 for $i = 1, \ldots, N$ (see White (1980)). The results in Table 3 suggest that the distribution of the innovations conditional on r_{Mt} is rather heteroskedastic, as we reject the null hypothesis at 5% significance level in almost all cases. This result confirms the need to use the robust estimates of the asymptotic covariance matrix of the t-based ML procedures in Proposition 6, as well as the problems that the HLV standard errors face, since they are based on Proposition 1 instead.

6 Conclusions

In this paper we study the efficiency-consistency trade-offs of three approaches to test the mean-variance efficiency of a candidate portfolio with returns r_{Mt} in excess of the riskless asset with respect to a set of N assets with excess returns $\mathbf{r}_t$. In particular, we consider tests based on the GMM approach advocated by MacKinlay and Richardson (1991), the elliptically symmetric semiparametric methods proposed by HLV, and a feasible parametric procedure that makes the assumption that, conditional on the reference portfolio, the excess returns of the original assets are independent and identically distributed as a multivariate t.

We would like to emphasise, though, that most of our results apply not only to the multivariate t, but also to any other elliptically-based likelihood estimator. The main advantage of the student t for our purposes is that we can make explicit its limiting relationship to the Gaussian distribution.

Our main asymptotic results are:

1. Under correct specification, the feasible parametric and HLV procedures are adaptive, in the sense that they are as efficient as if one had full knowledge of the true conditional distribution, including its shape parameters.

2. The t-based PML estimator provides asymptotically valid mean-variance efficiency tests when the conditional distribution is *i.i.d.* elliptical but not t, or when it is elliptical but heteroskedastic, although in both cases the asymptotic covariance matrices have to be adjusted appropriately. The same seems to be true of the HLV procedure, which confirms related results by Hodgson (2000) in a univariate context.

3. Test procedures that use the t-based PML estimators seem to be systematically more efficient from an asymptotic point of view than those based on the GMM estimators when the conditional distribution of $\mathbf{r}_t$ given r_{Mt} is elliptical, irrespective of whether or not it is t or conditionally homoskedastic.

4. Only the GMM estimator provides reliable inferences in the presence of asymmetries.

Although our Monte Carlo results are broadly in line with these theoretical conclusions, they also point out two interesting facts. First, we find that the HLV tests typically have much larger size distortions in finite samples than the other tests. Secondly, they have smaller size-adjusted power than the t-based PML tests, although the differences are very small when the latter are asymptotically suboptimal.

Finally, we apply these different procedures to test the mean-variance efficiency of the US aggregate stock market portfolio using monthly data over the period July 1962 to June 2007. The results that we obtain for industry portfolios indicate that the student t-based test clearly rejects the efficiency of the market portfolio, while the GMM test is borderline, and the HLV based test fails to reject. Given our Monte Carlo results, this contradicting behaviour is probably due to the lack of reliability of the nonparametric estimates of the asymptotic covariance matrix implicit in the HLV procedure. In contrast, all three tests reject the mean-variance efficiency of the market portfolio relative to the book-to-market sorted portfolios of Fama and French (1993). Importantly, we also find that while the assumption of Gaussianity is overwhelmingly rejected in both data sets, the evidence against a multivariate t distribution for the innovations is weak.

The fact that the number of assets that we consider in our Monte Carlo experiments and in our empirical application is fairly small probably means that they are unaffected by the criticism raised by Gibbons, Ross and Shanken (1989) in relation to the sensitivity of the asymptotic (in T) distribution of mean-variance efficiency tests to the cross-sectional dimension N. However, situations in which N/T cannot be regarded as negligible would require different asymptotic approximations to the one used in this paper.

A tedious, but rather trivial extension of our results would be to consider a situation in which we want to test the mean-variance efficiency of several reference portfolios simultaneously. As Gibbons, Ross and Shanken (1989) show, in a classical Gaussian log-likelihood context, such a test would simply assess the significance of the intercept in the regression of $\mathbf{r}_t$ on the vector of reference portfolios $\mathbf{r}_{Mt}$.

Similarly, we could also allow both $\mathbf{a}$ and $\mathbf{b}$ to linearly depend on a vector of predictor variables known at time $t-1$, $\mathbf{x}_{t-1}$ say, and in this way test for conditional mean variance efficiency, as discussed in Beaulieu, Dufour and Khalaf (2007) and others.

Hodgson, Linton and Vorkink (2006) consider the application of the elliptically symmetric semiparametric inference method developed in HLV to testing whether forward exchange rates provide unbiased forecasts of future changes in spot exchange rates. In the case of a single currency and contract period, the unbiased hypothesis implies that the slope is 1 and the intercept is 0 in the regression of future exchange rate movements on a constant and the current forward premium. It would be interesting to extend our analysis to cover that situation as well.

A closely related application would be spanning tests (see De Roon and Nijman (2001) for a recent survey), in which the null hypothesis also involves restrictions on both intercepts and slopes of a multivariate regression model (see Peñaranda and Sentana (2004) for a comparison of alternative GMM procedures).

We could increase the efficiency of the GMM estimator of $\mathbf{a}$ discussed in section 2.2.2 and the power of the associated test procedures by including additional moment restrictions that exploit the elliptical distribution of the innovations. For instance, we could follow Renault and Sentana (2003), and consider moment conditions of the form:

$$E\left\{\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}) \otimes vech[\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma})\boldsymbol{\varepsilon}_t'(\boldsymbol{\gamma})]\right\} = \mathbf{0}. \tag{24}$$

GMM estimators that combine (8) with this moment condition will typically have a lower asymptotic variance than $\hat{\mathbf{a}}_{GMM}$. In fact, we could regard the HLV estimator as a GMM estimator that optimally exploits the ellipticity of $\boldsymbol{\varepsilon}_t^*$, which means that in principle such augmented GMM procedures could achieve the elliptically symmetric semiparametric efficiency bound $\mathcal{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}_0)$. Like the HLV estimator, though, such GMM estimators will also become inconsistent if (24) does not hold, but their main advantage is that GMM integrates estimation and testing.

To test the validity of the specific distributional assumption for $\boldsymbol{\varepsilon}_t^*$ made for the purposes of obtaining $\hat{\mathbf{a}}_{ML}$ in our empirical application, we have used the LM specification tests of Mencia and Sentana (2008), who use the generalised hyperbolic family as the nesting distribution. And although there are many other tests of ellipticity in the statistical literature (see e.g. Beran (1979)), for the purposes of testing mean-variance efficiency we could also use the Hausman specification tests proposed by Fiorentini and Sentana (2007), which compare the consistent but inefficient estimator $\hat{\mathbf{a}}_{GMM}$ with the efficient but potentially inconsistent estimators $\hat{\mathbf{a}}_{HLV}$ and $\hat{\mathbf{a}}_{ML}$. An alternative procedure would be a moment test that checks whether the information matrix equality for M_{ll} implicit in Proposition 1 holds, as suggested by Fiorentini and Sentana (2007).

All these issues constitute interesting avenues for further research.

Appendix

A Proofs

Proposition 1:

The result follows directly from Proposition 1 in Fiorentini and Sentana (2007) by using the fact that in the case of model (1)

$$\mathbf{Z}'_{lt}(\boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1/2}\frac{\partial(\mathbf{a}+\mathbf{b}r_{Mt})}{\partial\boldsymbol{\theta}'} = \boldsymbol{\Omega}^{-1/2}\left[\begin{array}{cc}(1, r_{Mt})\otimes\mathbf{I}_N & \mathbf{0}\end{array}\right] \tag{A1}$$

and

$$\mathbf{Z}'_{st}(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\frac{\partial vec(\boldsymbol{\Omega})}{\partial\boldsymbol{\theta}'} = \frac{1}{2}(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\left(\begin{array}{cc}\mathbf{0} & \mathbf{D}_N\end{array}\right). \tag{A2}$$

Corollary 1

The asymptotic normality of the ML estimator of $\mathbf{a}$ follows from standard arguments by combining a central limit theorem for the score with a uniform law of large numbers for the Hessian matrix under the explicit assumptions that $\boldsymbol{\varepsilon}_t^*$ is *i.i.d.* and both r_{Mt} and r_{Mt}^2 are strictly stationary process with absolutely summable autocovariances. The expression for the asymptotic covariance matrix is a direct product of the partitioned inverse formula.

Proposition 2

The expressions for the matrices $\mathcal{A}_{\gamma\gamma t}(\boldsymbol{\phi})$, $\mathcal{B}_{\gamma\gamma t}(\boldsymbol{\phi})$ and $\mathcal{C}_{\gamma\gamma t}(\boldsymbol{\phi})$ follow directly from replacing (A1) and (A2) in Proposition 2 in Fiorentini and Sentana (2007). The asymptotic normality of the GMM estimator of $\boldsymbol{\gamma}$ can be obtained using the arguments in the proof of Corollary 1.

Proposition 3

The first part of the Proposition follows directly from the first part of Proposition 15 in Fiorentini and Sentana (2007). The second part of the distribution also follows directly from the second and third parts of the same proposition because mesokurtic elliptical distributions satisfy their condition (39), as Fiorentini and Sentana (2007) explain in their proof.

Proposition 4

The first part of the Proposition follows directly from the first part of Proposition 16 in Fiorentini and Sentana (2007). Specifically, let us initially keep η fixed to some positive value.

Since $\boldsymbol{\varepsilon}_t$ is elliptical, it can be written as $\boldsymbol{\varepsilon}_t^* = \sqrt{\varsigma_t}\mathbf{u}_t$ where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$ and ς_t is a non-negative random variable independent of $\mathbf{u}_t$. Since

$$\varsigma_t(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau) = \frac{1}{\tau}\boldsymbol{\varepsilon}_t'(\boldsymbol{\gamma}_0)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\upsilon}_0)\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}_0) = \frac{\tau_0}{\tau}\varsigma_t$$

where $\varsigma_t = \varsigma_t(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_0)$, we can write the blocks of the score corresponding to $\boldsymbol{\gamma}$, $\boldsymbol{\upsilon}$ and τ as

$$\mathbf{s}_{\gamma t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau, \eta) = \begin{pmatrix} 1 \\ r_{Mt} \end{pmatrix} \otimes \frac{1}{\sqrt{\tau}}\boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{\upsilon}_0)\delta\left[(\tau_0/\tau)\varsigma_t, \eta\right]\sqrt{(\tau_0/\tau)}\sqrt{\varsigma_t}\mathbf{u}_t \tag{A3}$$

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\upsilon} t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau, \eta) &= \frac{1}{2}\frac{\partial vec'[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon}_0)]}{\partial \boldsymbol{\upsilon}}\left[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon}_0)^{-1/2} \otimes \boldsymbol{\Upsilon}(\boldsymbol{\upsilon}_0)^{-1/2}\right] \\ &\quad \times vec\left\{\delta\left[(\tau_0/\tau)\varsigma_t, \eta\right]\frac{\tau_0}{\tau}\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N\right\}\end{aligned} \tag{A4}$$

and

$$s_{\tau t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau, \eta) = \frac{1}{2\tau}vec'(\mathbf{I}_N)vec\left\{\delta\left[(\tau_0/\tau)\varsigma_t, \eta\right]\frac{\tau_0}{\tau}\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N\right\}. \tag{A5}$$

Then, it follows that $E\left[\mathbf{s}_{\gamma t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau, \eta)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0\right] = \mathbf{0}$ regardless of τ and η because of the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t) = \mathbf{0}$.

If we define $\tau_\infty(\eta)$ as the value that solves the implicit equation

$$E\left[\frac{N\eta + 1}{1 - 2\eta + \eta(\tau_0/\tau)\varsigma_t}\frac{\tau_0}{\tau}\frac{\varsigma_t}{N} - 1\middle|\boldsymbol{\varphi}_0\right] = 0 \tag{A6}$$

then it is straightforward to show that

$$\begin{aligned} E\left[\mathbf{s}_{\boldsymbol{\upsilon} t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_\infty(\eta), \eta)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0\right] &= \mathbf{0} \\ E\left[s_{\tau t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_\infty(\eta), \eta)|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0\right] &= 0. \end{aligned}$$

by using the fact that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$.

If we choose η_∞ as the solution to the implicit equation

$$E\left[s_{\eta t}(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_\infty(\eta), \eta)|\boldsymbol{\varphi}_0\right] = 0, \tag{A7}$$

then it is clear that $\boldsymbol{\upsilon}_0, \tau_\infty(\eta_\infty)$ and η_∞ will be the pseudo true values of the parameters.

To obtain the variance of the t-score and the expected value of the t-hessian under misspecification it is convenient to rewrite the score as

$$\mathbf{s}_{\boldsymbol{\vartheta} t}(\boldsymbol{\vartheta}, \eta) = \begin{bmatrix} \mathbf{Z}_{\gamma t}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\upsilon} t}(\boldsymbol{\vartheta}) \\ \mathbf{0} & \mathbf{Z}_{\tau t}(\boldsymbol{\vartheta}) \end{bmatrix} \times \left[\mathbf{e}_{lt}(\boldsymbol{\vartheta}, \eta), \mathbf{e}_{st}(\boldsymbol{\vartheta}, \eta)\right]$$

where

$$\begin{aligned}
\mathbf{Z}_{\gamma t}(\boldsymbol{\vartheta}) &= \begin{pmatrix} 1 \\ r_{Mt} \end{pmatrix} \otimes \frac{1}{\sqrt{\tau}} \boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{\upsilon}) \\
\mathbf{Z}_{\upsilon t}(\boldsymbol{\vartheta}) &= \frac{1}{2} \frac{\partial vec'[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})]}{\partial \boldsymbol{\upsilon}} \left[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})^{-1/2} \otimes \boldsymbol{\Upsilon}(\boldsymbol{\upsilon})^{-1/2} \right] \\
\mathbf{Z}_{\tau t}(\boldsymbol{\vartheta}) &= \frac{1}{2} \frac{1}{\tau} vec'(\mathbf{I}_N)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{e}_{lt}(\boldsymbol{\vartheta}, \eta) &= \delta \left[\varsigma_t(\boldsymbol{\vartheta}), \eta \right] \sqrt{\varsigma_t(\boldsymbol{\vartheta})} \mathbf{u}_t \\
\mathbf{e}_{st}(\boldsymbol{\vartheta}, \eta) &= vec \left\{ \delta \left[\varsigma_t(\boldsymbol{\vartheta}), \eta \right] \varsigma_t(\boldsymbol{\vartheta}) \mathbf{u}_t \mathbf{u}_t' - \mathbf{I}_N \right\}.
\end{aligned}$$

Then, we can follow exactly the same steps as in the proof of Proposition 1 in Fiorentini and Sentana (2007) by exploiting that (A6) and (A7) hold at the pseudo-true parameter values $\boldsymbol{\phi}_\infty$.

Finally, tedious algebraic manipulations show that $\mathcal{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ and $\mathcal{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ will be block diagonal between $(\boldsymbol{\gamma}, \boldsymbol{\upsilon})$ and (τ, η) if $E[\partial d_t(\boldsymbol{\vartheta})/\partial \boldsymbol{\upsilon} | \boldsymbol{\varphi}_0] = \mathbf{0}$. But this trivially holds in our parametrization because $|\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})| = 1$ for all $\boldsymbol{\upsilon}$.

Proposition 5

If we use the subscript J to denote the joint log-likelihood function of $\mathbf{R}_t$, expression (2) in Fiorentini and Sentana (2007) implies that

$$\begin{aligned}
\mathbf{s}_{J\boldsymbol{\rho}t}(\boldsymbol{\rho}, \boldsymbol{\eta}) = & \frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\rho}) \delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho}) \boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}] \cdot \boldsymbol{\epsilon}_t(\boldsymbol{\rho}) \\
& + \frac{1}{2} \frac{\partial vec' \left[\boldsymbol{\Sigma}_t(\boldsymbol{\rho}) \right]}{\partial \boldsymbol{\rho}} [\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\rho}) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\rho})] \\
& \times vec \left\{ \delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho}) \boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}] \cdot \boldsymbol{\epsilon}_t(\boldsymbol{\rho}) \boldsymbol{\epsilon}_t'(\boldsymbol{\rho}) - \boldsymbol{\Sigma}_t(\boldsymbol{\rho}) \right\}.
\end{aligned}$$

In our case,

$$\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} = \frac{\partial}{\partial \boldsymbol{\rho}'} \begin{pmatrix} \mu_M \\ \mathbf{a} + \mathbf{b}\mu_M \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{0}' & \mathbf{0}' & 1 & 0 \\ \mathbf{I}_N & \mu_M \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

As for

$$\frac{\partial vec \left[\boldsymbol{\Sigma}_t(\boldsymbol{\rho}) \right]}{\partial \boldsymbol{\rho}'} = \frac{\partial vec}{\partial \boldsymbol{\rho}'} \begin{pmatrix} \sigma_M^2 & \sigma_M^2 \mathbf{b}' \\ \sigma_M^2 \mathbf{b} & \sigma_M^2 \mathbf{b}\mathbf{b}' + \boldsymbol{\Omega} \end{pmatrix},$$

it is more convenient to obtain its elements by blocks, so that

$$\frac{\partial}{\partial \boldsymbol{\rho}'} \begin{pmatrix} \sigma_M^2 \\ \sigma_M^2 \mathbf{b} \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{0}' & \mathbf{0}' & 0 & 1 \\ \mathbf{0} & \sigma_M^2 \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{b} \end{pmatrix}$$

and

$$\frac{\partial vec(\sigma_M^2 \mathbf{b}\mathbf{b}' + \boldsymbol{\Omega})}{\partial \boldsymbol{\rho}'} = \left[\begin{array}{ccccc} \mathbf{0} & (\mathbf{I}_{N^2} + \mathbf{K}_{NN})(\sigma_M^2 \mathbf{b} \otimes \mathbf{I}_N) & \mathbf{D}_N & \mathbf{0} & (\mathbf{b} \otimes \mathbf{b}) \end{array} \right],$$

and then re-arrange them appropriately.

It is also easy to see that

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho}) = \begin{pmatrix} \sigma_M^{-2} + \mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b} & -\mathbf{b}'\boldsymbol{\Omega}^{-1} \\ -\boldsymbol{\Omega}^{-1}\mathbf{b} & \boldsymbol{\Omega}^{-1} \end{pmatrix},$$

by exploiting the Cholesky decomposition of $\boldsymbol{\Sigma}(\boldsymbol{\rho})$ in (A8).

We can also tediously prove that

$$\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho})\right] \frac{\partial vec(\Sigma(\boldsymbol{\rho}))}{\partial \sigma_M^2} = \left[\begin{array}{c} 1/\sigma_M^4 \\ \mathbf{0} \end{array} \right],$$

and

$$\frac{\partial vec'[\boldsymbol{\Sigma}(\boldsymbol{\rho})]}{\partial \mathbf{b}} \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho})\right] = \left(\begin{array}{cc} -2\boldsymbol{\Omega}^{-1}\mathbf{b} & \boldsymbol{\Omega}^{-1} \otimes \mathbf{e}_{1,N+1} \end{array} \right)$$

where $\mathbf{e}_{1,N+1}$ is a vector whose first element is one and has zeros in its remaining N positions.

On this basis, we can write

$$\begin{aligned} \mathbf{s}_{J\gamma t}(\boldsymbol{\rho}, \boldsymbol{\eta}) &= \left[\begin{pmatrix} 1 \\ \mu_M \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1} \right] \delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \eta] \left[\begin{array}{cc} -\mathbf{b} & \mathbf{I}_N \end{array} \right] \boldsymbol{\epsilon}_t(\boldsymbol{\rho}) \\ &+ \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ -2\boldsymbol{\Omega}^{-1}\mathbf{b} & \boldsymbol{\Omega}^{-1} \otimes \mathbf{e}_{1,N+1} \end{pmatrix} vec\left\{\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \eta]\boldsymbol{\epsilon}_t(\boldsymbol{\rho})\boldsymbol{\epsilon}_t'(\boldsymbol{\rho}) - \boldsymbol{\Sigma}(\boldsymbol{\rho})\right\}. \end{aligned}$$

and

$$\mathbf{s}_{J\boldsymbol{\omega}t}(\boldsymbol{\rho}, \boldsymbol{\eta}) = \frac{1}{2}\mathbf{D}_N' \left[\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right] vec\left\{\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\boldsymbol{\epsilon}_{\mathbf{r}t}(\boldsymbol{\rho})\boldsymbol{\epsilon}_{\mathbf{r}t}'(\boldsymbol{\rho}) - \boldsymbol{\Omega}\right\},$$

where $\boldsymbol{\epsilon}_{\mathbf{r}t}(\rho) = \mathbf{r}_t - \mathbf{a} - \mathbf{b}\mu_M$.

In addition,

$$s_{J\mu_M}(\boldsymbol{\rho}, \boldsymbol{\eta}) = \frac{1}{2\sigma_M^2}\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\epsilon_t(\mu_M),$$

and

$$s_{J\sigma_M^2}(\boldsymbol{\rho}, \boldsymbol{\eta}) = \frac{1}{2\sigma_M^4}\left\{\delta_{N+1}[\boldsymbol{\epsilon}_t^{*\prime}(\boldsymbol{\rho})\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}), \boldsymbol{\eta}]\epsilon_t^2(\mu_M) - \sigma_M^2\right\},$$

where $\epsilon_{Mt}(\mu_M) = r_{Mt} - \mu_M$.

Finally, the result follows tediously from Proposition 1 in Fiorentini and Sentana (2007) if we exploit the fact that

$$\frac{\partial vec'[\boldsymbol{\Sigma}(\boldsymbol{\rho})]}{\partial \mathbf{b}} vec[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho})] = \mathbf{0}$$

and

$$\frac{\partial vec'[\boldsymbol{\Sigma}(\boldsymbol{\rho})]}{\partial \sigma_M^2} vec[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho})] = \frac{1}{\sigma_M^2}.$$

Interestingly, note that under Gaussianity $\mathcal{I}_{\boldsymbol{\omega}\sigma_M^2}(\boldsymbol{\phi}) = \mathbf{0}$, which confirms that the estimators of the parameter of the marginal model for r_{Mt} and the conditional model for $\mathbf{r}_t$ will be independent.

Proposition 6

Since $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})]|I_{t-1}; \boldsymbol{\varphi}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_{N+1}, \boldsymbol{\varrho}_0)$, we can write

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho})] = e_t \begin{pmatrix} u_{0t} \\ \sqrt{1-u_{0t}^2}\tilde{\mathbf{u}}_t \end{pmatrix}$$

where e_t is a positive random variable such that $E(e_t^2) = N+1$, u_{0t}^2 is a beta random variable with parameters $(1/2, N/2)$ and $\tilde{\mathbf{u}}_t$ is an independent uniform on the unit sphere surface in $\mathbb{R}^N$.

Given that the Cholesky decomposition of $\boldsymbol{\Sigma}(\boldsymbol{\rho})$ can be written as

$$\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\rho}) = \begin{pmatrix} \sigma_M & \mathbf{0} \\ \mathbf{b}\sigma_M & \boldsymbol{\Omega}^{1/2} \end{pmatrix} \tag{A8}$$

with $\boldsymbol{\Omega}^{1/2}$ denoting the Cholesky decomposition of $\boldsymbol{\Omega}$, we can write

$$\mathbf{R}_t - \boldsymbol{\mu}(\boldsymbol{\rho}) = \begin{pmatrix} \sigma_{M0} e_t u_{0t} \\ \mathbf{b}_0 \sigma_{M0} e_t u_{0t} + \boldsymbol{\Omega}_0^{1/2} e_t \sqrt{1-u_{0t}^2}\tilde{\mathbf{u}}_t \end{pmatrix},$$

where u_{0t} is a random variable on $(-1, 1)$ with density $(1-u_{0t}^2)^{N/2-1}/B(1/2, N/2)$. This follows from the symmetry of u_{0t} and the fact that the density of $|u_{0t}|$ is $2(1-u_{0t}^2)^{N/2-1}/B(1/2, N/2)$ because the density of u_{0t}^2 is $(u_{0t}^2)^{-1/2}(1-u_{0t}^2)^{N/2-1}/B(1/2, N/2)$. As a result, $\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}_0; \mathbf{R}_t) = \mathbf{r}_t - \mathbf{a}_0 - \mathbf{b}_0 r_{Mt} = \boldsymbol{\Omega}_0^{1/2} e_t \sqrt{1-u_{0t}^2}\tilde{\mathbf{u}}_t$ and

$$\boldsymbol{\varepsilon}_t'(\boldsymbol{\gamma}_0; \mathbf{R}_t)\boldsymbol{\Omega}_0^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}_0; \mathbf{R}_t) = e_t^2(1-u_{0t}^2)$$

because $\tilde{\mathbf{u}}_t'\tilde{\mathbf{u}}_t = 1$.

Let's now consider the following misspecified model

$$\boldsymbol{\Omega}^{-1/2}(\mathbf{r}_t - \mathbf{a} - \mathbf{b} r_{Mt})|r_{Mt}, \boldsymbol{\phi} \sim i.i.d. t(\mathbf{0}, \mathbf{I}_N, \eta)$$

and assume $\varsigma_t(\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau) = \boldsymbol{\varepsilon}_t'(\boldsymbol{\gamma}_0)\tau^{-1}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\upsilon}_0)\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}_0) = (\tau_0/\tau)e_t^2(1-u_{0t}^2)$. Hence, the blocks of the score corresponding to $\boldsymbol{\gamma}$, $\boldsymbol{\upsilon}$ and τ are given by (A3), (A4) and (A5) with $e_t^2(1-u_{0t}^2)$ replacing ς_t. Then, the first part of this proposition can be obtained using the arguments in the proof of the first part of Proposition 4.

The proof of the second part is analogous to the proof of the second part of Proposition 4. Note, in particular, that having contemporaneous, conditionally heteroskedastic innovations is innocuous to obtain the relevant expressions since all the scalar terms $\mathrm{M}_i^j(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{ f_i^j[\varsigma_t(\boldsymbol{\rho})] \middle| \boldsymbol{\varphi} \right\}$ appearing in $\mathcal{O}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{H}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ satisfy

$$E\left\{ f_i^j[\varsigma_t(\boldsymbol{\rho})] \middle| \boldsymbol{\varphi}, r_{Mt} \right\} = E\left\{ f_i^j[\varsigma_t(\boldsymbol{\rho})] \middle| \boldsymbol{\varphi} \right\}.$$

Finally, our parametrization implies that $\mathcal{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will be block diagonal between $(\boldsymbol{\gamma},\boldsymbol{\upsilon})$ and (τ,η), as in Proposition 4.

References

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

Bahadur, R. (1960). "Stochastic comparison of tests", *The Annals of Mathematical Statistics*, 31, 276-295.

Beaulieu, M.C., J.M. Dufour, and L. Khalaf (2007): Testing mean-variance efficiency in CAPM with possibly non-gaussian errors: an exact simulation-based approach", *Journal of Business and Economic Statistics*, forthcoming.

Beran, R. (1979). "Testing for ellipsoidal symmetry of a multivariate density", *Annals of Statistics*, 7, 150-162.

Berk, J. (1997). "Necessary conditions for the CAPM", *Journal of Economic Theory*, 73, 245-257.

Black, F., M.C. Jensen, and M. Scholes (1972). "The capital asset pricing model: some empirical tests", in Michael C. Jensen (ed.), *Studies in the Theory of Capital Markets*, Praeger Publishers.

Campbell, J., W. Lo, and A.C. MacKinlay (1997). *The econometrics of financial markets,* Princeton: Princeton University Press.

Chamberlain, G. (1983). "A characterization of the distributions that imply mean-variance utility functions", *Journal of Economic Theory*, 29, 185-201.

Cochrane, J. (2001). *Asset pricing*, Princeton: Princeton University Press.

Cressie, N.A.C., A.S. Davis, J.L. Folks and G.E. Policello (1981). "The moment generating function and negative integer moments", *American Statistician* 35, 148-150.

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

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

De Roon, F.A. and T.E. Nijman (2001). "Testing for mean-variance spanning: a survey", *Journal of Empirical Finance*, 8-2, 111-156.

Erdélyi, A. (1981). *Higher transcendental functions, vol. 1*, Robert E. Krieger Publishing Company, Melbourne, FL.

Errunza, V., K. Hogan and M. Hung (1999). "Can the gains from international diversification be achieved without trading abroad", *Journal of Finance*, 54, 2075-2107.

Fama, E.F. and French, K.R. (1993): "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics* 33, 3-56.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). "Maximum likelihood estimation and inference on multivariate conditionally heteroscedastic dynamic regression models with student t innovations", *Journal of Business and Economic Statistics*, 24, 532-546.

Fiorentini, G. and E. Sentana (2007). "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working Paper 0713.

Geweke, J. (1981). "The approximate slopes of econometric tests", *Econometrica*, 49, 1427-1442.

Gibbons, M., S. Ross, and J. Shanken (1989). "A test of the efficiency of a given portfolio", *Econometrica,* 57, 1121-1152.

Hansen, L. (1982). "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029–1054.

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

Hodgson, D., O. Linton and K. Vorkink (2002). "Testing the capital asset pricing model efficiently under elliptical symmetry: a semiparametric approach", *Journal of Applied Econometrics, 17, 617-639.*

Hodgson, D., O. Linton and K. Vorkink (2004). "Testing forward exchange rate unbiasedness efficiently: a semiparametric approach", *Journal of Applied Economics*, 7-I, 325-353.

Huberman, G. and S. Kandel (1987). "Mean-variance spanning", *Journal of Finance*, 42, 873-888.

Jobson, J.D. and B. Korkie (1982). "Estimation for Markowitz portfolios", *Journal of the American Statistical Association*, 75, 544-554.

Kan, R. and G. Zhou (2006). "Modeling non-normality using multivariate t: implications for asset pricing", mimeo Washington University in St.Louis.

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.

MacKinlay, A.C. and M. Richardson (1991). "Using generalized method of moments to test mean-variance efficiency", *Journal of Finance*, 46, 511–527.

Magnus, J.R. and H. Neudecker (1988). *Matrix differential calculus with applications on statistics and econometrics*, Wiley: Chichester.

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

Markowitz, H. (1952). "Portfolio selection", *Journal of Finance*, 8, pp. 77-91.

Masoom Ali, M. and S. Nadarajah (2007). "Information matrices for Normal and Laplace mixtures", *Information Sciences* 177, 947-955.

Meddahi, N. and E. Renault (1998). "Quadratic M-estimators for ARCH-type processes", CIRANO Working paper 98s-29.

Mencía, F.J. and E. Sentana (2008). "Distributional tests in multivariate dynamic models with Normal and Student t innovations", mimeo, CEMFI.

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

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

Peñaranda, F. and E. Sentana (2004). "Spanning tests in return and stochastic discount factor mean-variance frontiers: a unifying approach", CEMFI Working paper No. 0410.

Renault, E. and E. Sentana (2003). "GMM estimation of multivariate regression models with symmetric errors", mimeo, CEMFI.

Sentana, E. (2008). "The econometrics of mean-variance efficiency: a survey", mimeo, CEMFI.

White, H. (1980). "A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 45, 817-838.

Zhou, G. (1993). "Asset-pricing tests under alternative distributions", *Journal of Finance*, 48, 1927–1942.

Table 1

Table 1.a: Parameter estimates: $\mathbf{r}_t = \mathbf{a} + \mathbf{b} r_{Mt} + \mathbf{u}_t$

Industry portfolios

	GMM		HLV		t ML	
Category	a	b	a	b	a	b
Cnsmr	0.099	0.954	0.026	1.018	0.023	1.041
	(0.091)	(0.030)	(0.076)	(0.017)	(0.056)	(0.013)
Manuf	0.134	0.870	0.069	0.922	0.123	0.910
	(0.076)	(0.022)	(0.063)	(0.014)	(0.064)	(0.015)
SHiTec	−0.086	1.124	−0.064	1.054	−0.146	1.017
	(0.117)	(0.035)	(0.097)	(0.022)	(0.099)	(0.022)
Hlth	0.205	0.875	0.072	0.947	0.092	0.954
	(0.146)	(0.048)	(0.123)	(0.028)	(0.135)	(0.031)
Other	0.088	1.066	0.016	1.132	0.003	1.120
	(0.087)	(0.024)	(0.073)	(0.017)	(0.085)	(0.019)

Book-to-market sorted portfolios

	GMM		HLV		t ML	
Quintile	a	b	a	b	a	b
1	−0.108	1.072	−0.111	1.076	−0.121	1.066
	(0.062)	(0.018)	(0.053)	(0.012)	(0.035)	(0.008)
2	0.040	0.992	−0.068	1.027	0.018	1.024
	(0.060)	(0.019)	(0.050)	(0.012)	(0.049)	(0.011)
3	0.151	0.892	0.019	0.930	0.116	0.924
	(0.073)	(0.024)	(0.061)	(0.014)	(0.062)	(0.014)
4	0.328	0.848	0.105	0.913	0.215	0.909
	(0.087)	(0.031)	(0.073)	(0.017)	(0.064)	(0.015)
5	0.430	0.936	0.221	0.984	0.308	0.985
	(0.109)	(0.039)	(0.093)	(0.021)	(0.111)	(0.025)

Table 1.b: Mean-variance efficiency tests ($H_0 : \mathbf{a} = \mathbf{0}$)

Industry portfolios

	GMM	GMM robust	HLV	t ML	t ML robust
Statistic	12.056	10.911	2.194	15.226	14.365
p-value	0.034	0.053	0.822	0.009	0.013

Book-to-market sorted portfolios

	GMM	GMM robust	HLV	t ML	t ML robust
Statistic	21.350	21.417	12.837	21.846	19.772
p-value	0.001	0.001	0.025	0.001	0.001

Notes: Sample: July:1962-June:2007. Industry definitions: Cnsmr: Consumer Durables, Non-Durables, Wholesale, Retail, and Some Services (Laundries, Repair Shops). Manuf: Manufacturing, Energy, and Utilities. HiTec: Business Equipment, Telephone and Television Transmission. Hlth: Healthcare, Medical Equipment, and Drugs. Other: Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance.

Table 2: Student t tests

Industry portfolios

Test	Statistic	p-value
Kurtosis ($\tau_{mkT}(\hat{\boldsymbol{\phi}})$)	0.033	0.974
Skewness ($\tau_{aT}(\hat{\boldsymbol{\phi}})$)	10.003	0.075
Kurtosis & skewness ($\tau_{mgT}(\hat{\boldsymbol{\phi}})$)	10.036	0.123

Book-to-market sorted portfolios

Test	Statistic	p-value
Kurtosis ($\tau_{mkT}(\hat{\boldsymbol{\phi}})$)	0.037	0.970
Skewness ($\tau_{aT}(\hat{\boldsymbol{\phi}})$)	9.880	0.079
Kurtosis & skewness ($\tau_{mgT}(\hat{\boldsymbol{\phi}})$)	9.917	0.128

Notes: July:1962-June:2007. 'Kurtosis' is a two-sided test of the null hypothesis of Student t innovations versus the alternative hypothesis of symmetric Generalized Hyperbolic innovations; the test statistic is distributed as a $N(0,1)$ under the null. 'Skewness' refers to a test of the same null hypothesis versus asymmetric t innovations as alternative hypothesis; the test statistic is distributed as a χ^2_N under the null. 'Kurtosis & skewness' is a two-sided test of the null hypothesis of Student t innovations versus the alternative hypothesis of Generalized Hyperbolic innovations; the test statistic is distributed as a χ^2_{N+1} under the null. The test statistics are defined in Mencía and Sentana (2005).

Table 3: Conditional heteroskedasticity test

Industry portfolios

Category	Cnsmr	Manuf	HiTec	Hlth	Other
Statistic	45.026	9.633	14.635	48.257	4.866
p-value	0.000	0.008	0.001	0.000	0.088

Book-to-market sorted portfolios

Quintile	1	2	3	4	5
Statistic	24.070	11.748	27.480	41.262	62.098
p-value	0.000	0.003	0.000	0.000	0.000

Notes: July:1962-June:2007. Based on the statistical significance of $\delta_i = (\delta_{1i}, \delta_{2i})'$ in $\hat{\varepsilon}^2_{it} = c_i + \delta_{1i} r_{Mt} + \delta_{2i} r^2_{Mt} + v_{it}$, where $\hat{\varepsilon}_{it}$'s are the OLS residuals from a regression of r_{it} on a constant and r_{Mt}. The test statistic, nR^2 –where R^2 is the coefficient of determination of the regression–, is distributed as a χ^2_2 under the null hypothesis of conditional homoskedasticity.

Figure 1a: Standardized bivariate normal density

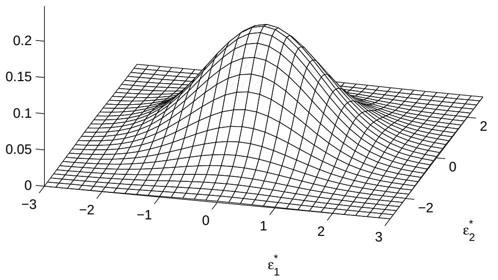


Figure 1b: Contours of a standardized bivariate normal density

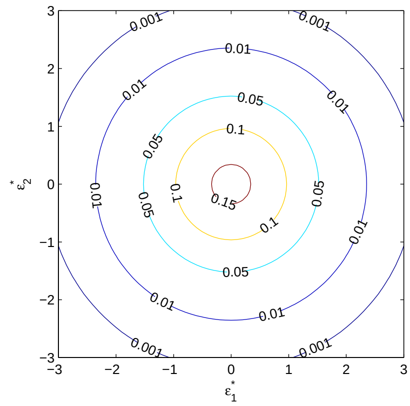


Figure 1c: Standardized bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

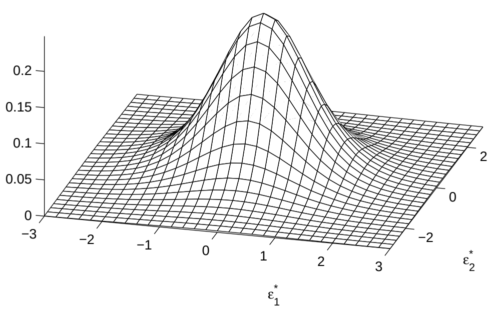


Figure 1d: Contours of a standardized bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)

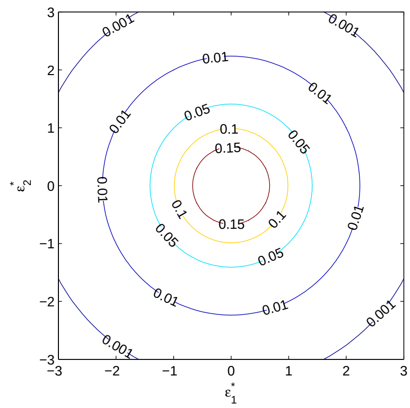


Figure 1e: Standardized bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

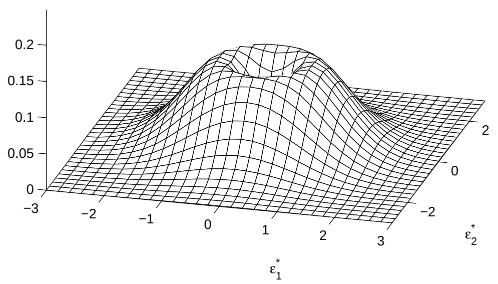


Figure 1f: Contours of a standardized bivariate Kotz density with multivariate excess kurtosis $\kappa = -0.15$

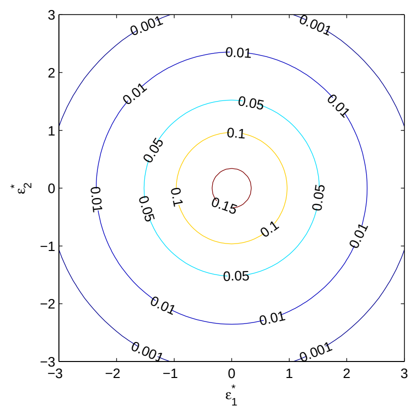


Figure 1g: Standardized bivariate Discrete scale mixture of normals density with multivariate excess kurtosis $\kappa = 0.125$ ($\pi = 0.5$)

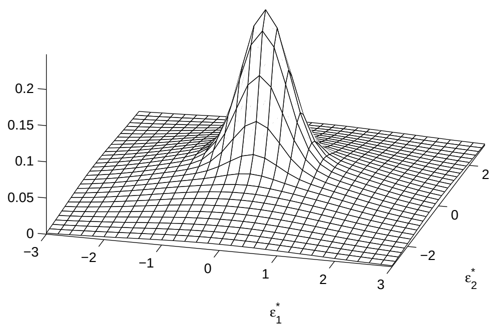


Figure 1h: Contours of a standardized bivariate Discrete scale mixture of normals density with multivariate excess kurtosis $\kappa = 0.125$ ($\pi = 0.5$)

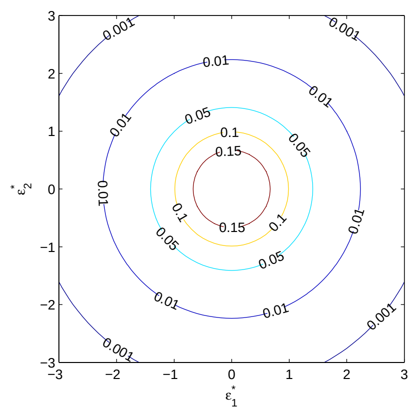
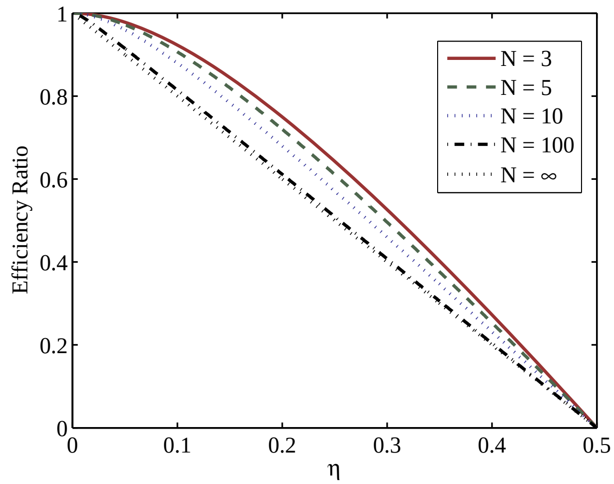
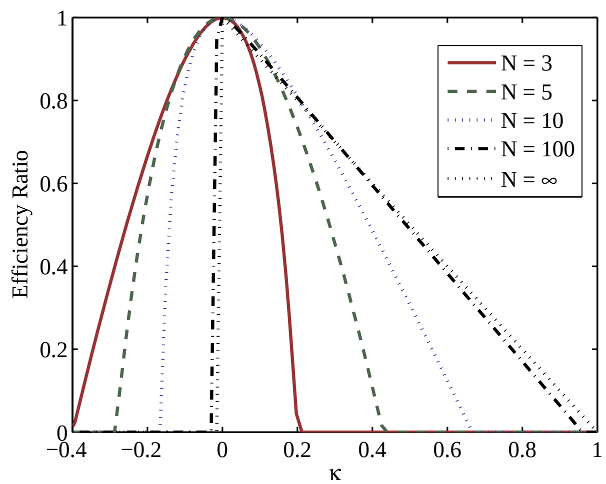


Figure 2: Relative efficiency ML/HLV vs Gaussian PML

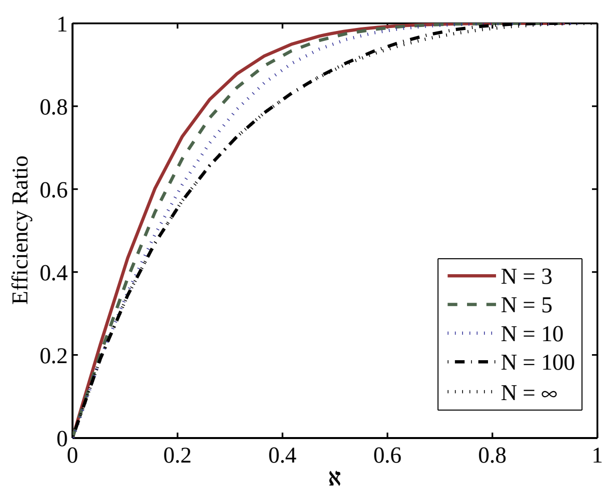
t innovations



Kotz innovations



Discrete mixture of normals innovations ($\pi = 0.5$)



Notes: The efficiency ratio is measured by $\mathrm{M}_{ll}^{-1}(\boldsymbol{\eta})$. For t innovations with ν degrees of freedom, $\eta = 1/\nu$. For Kotz innovations, κ denotes the coefficient of multivariate excess kurtosis. $\varkappa$ is the variance ratio of the two components in the Discrete scale mixture of normals.

Figure 3a: Power of the ML-based and Gaussian-PML-based Wald tests for elliptical innovations with $\mathrm{M}_{ll}(\boldsymbol{\eta})$ corresponding to t_8

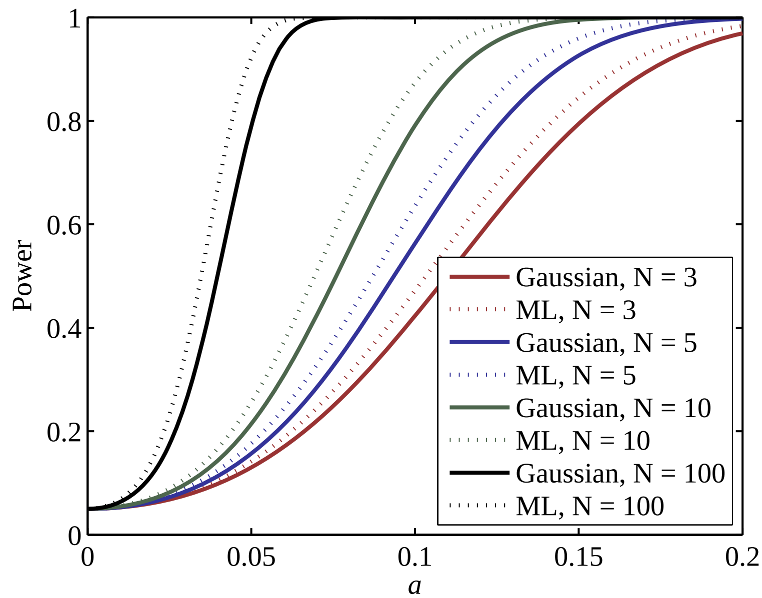
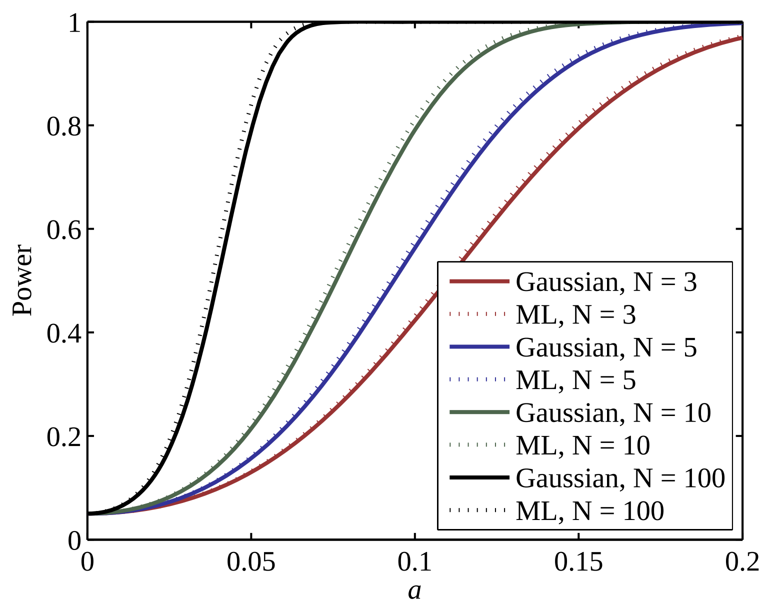


Figure 3b: Power of the ML-based and Gaussian-PML-based Wald tests for elliptical innovations with $\mathrm{M}_{ll}(\boldsymbol{\eta})$ corresponding to t_{20}



Notes: Results at the 5% level. $T = 500$, $\mathbf{\Omega} = \mathbf{I}_N$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, and $\mathbf{a} = a\ell_N$, with $\ell_N = (1, \ldots, 1)'$ and $a \in [0, .2]$.

Figure 4a: Relative efficiency of t PML vs Gaussian PML / ML for Gaussian innovations

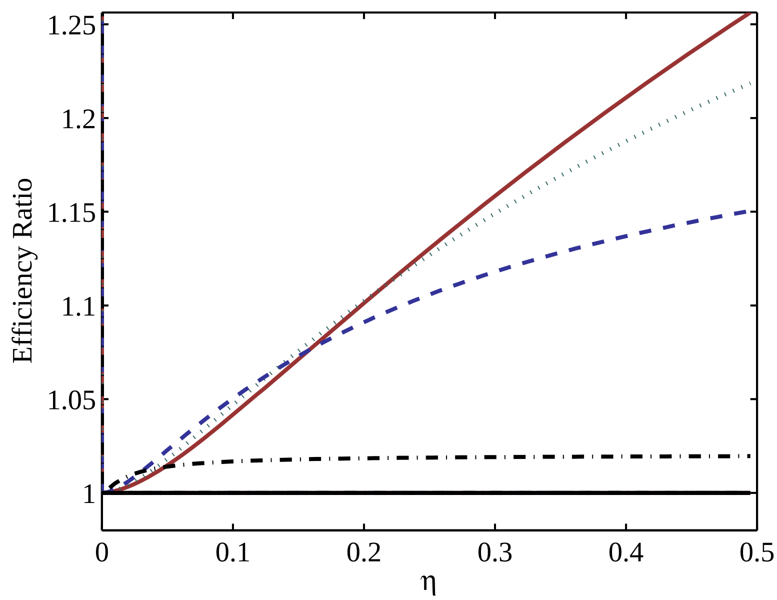
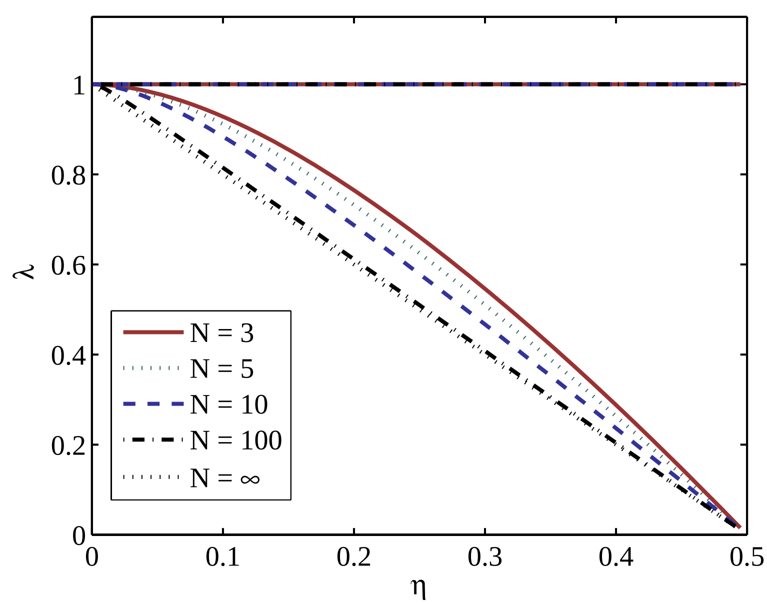


Figure 4b: Pseudo-true value of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$ for Gaussian innovations



Notes: The efficiency ratio is measured by $\mathrm{M}^{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}^{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^2\}$. Vertical lines in the top panel and horizontal lines in the bottom panel indicate the the pseudo-true values of η and λ when η is also estimated. Horizontal lines in the top panel indicate the relative efficiency of the ML estimator: $1/\mathrm{M}_{ll}(\boldsymbol{\eta}_0)$.

Figure 4c: Relative efficiency of ML / t PML vs Gaussian PML for Kotz innovations ($\kappa = 0.125$)

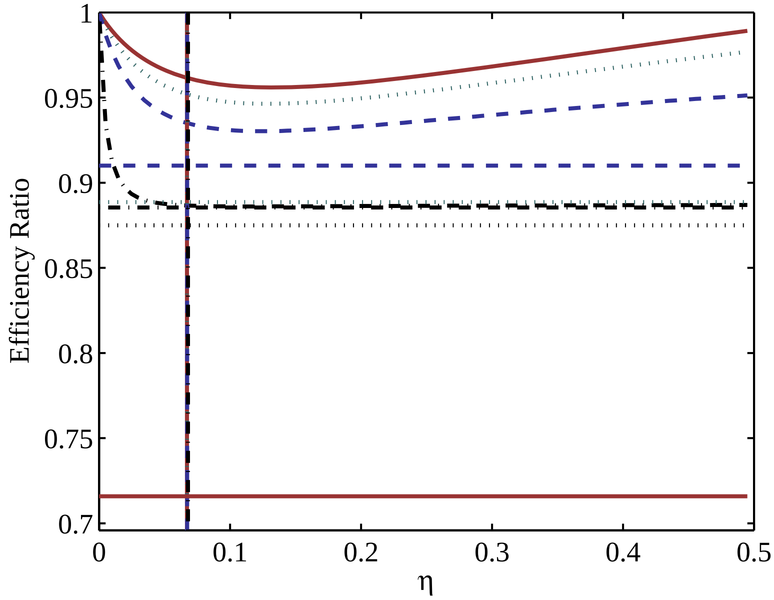
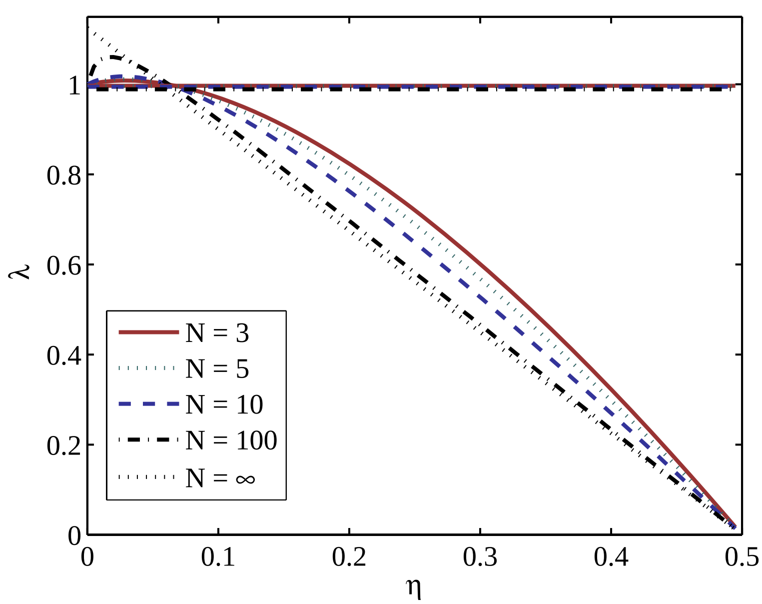


Figure 4d: Pseudo-true value of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$ for Kotz innovations ($\kappa = 0.125$)



Notes: The efficiency ratio is measured by $\mathrm{M}^{O}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}^{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)]^2\}$. Vertical lines in the top panel and horizontal lines in the bottom panel indicate the the pseudo-true values of η and λ when η is also estimated. Horizontal lines in the top panel indicate the relative efficiency of the ML estimator: $1/\mathrm{M}_{ll}(\boldsymbol{\eta}_0)$. κ is the coefficient of multivariate excess kurtosis.

Figure 4e: Relative efficiency of ML / t PML vs Gaussian PML for Discrete mixture of normals innovations ($\pi = 0.5$, $\kappa = 0.25$)

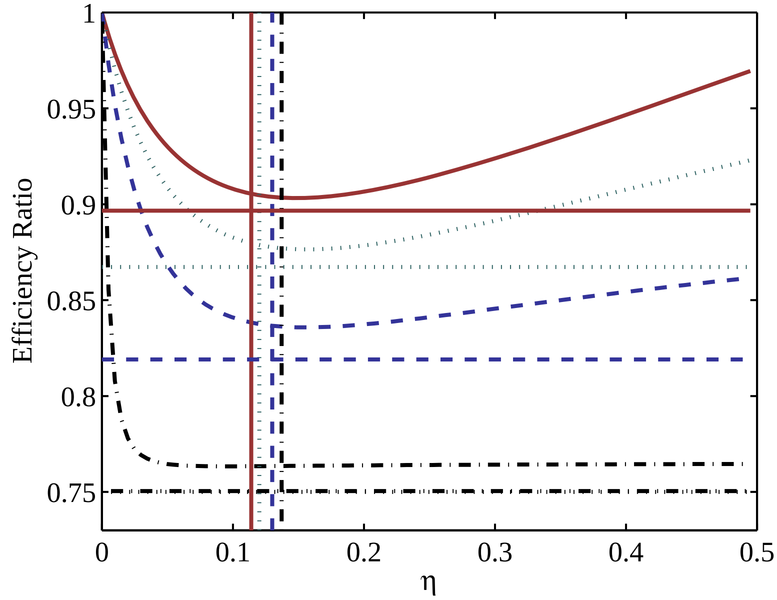
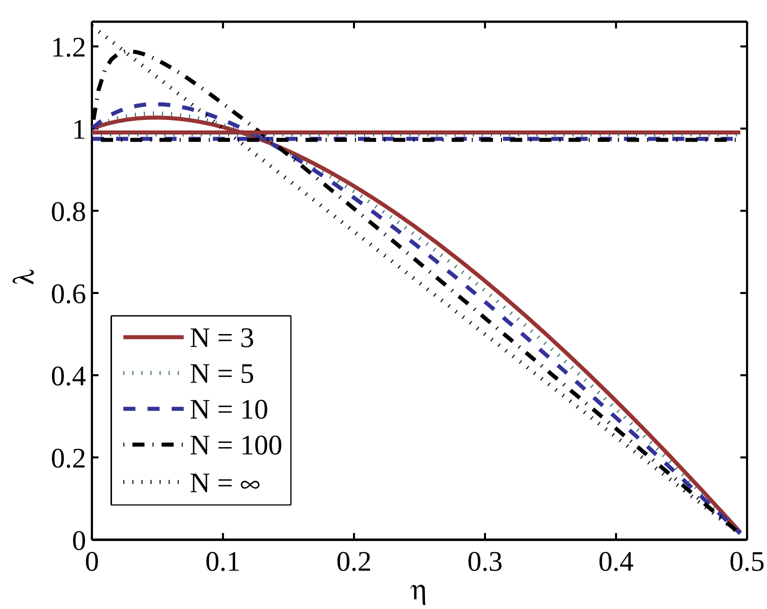


Figure 4f: Pseudo-true value of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$ for Discrete mixture of normals innovations ($\pi = 0.5$, $\kappa = 0.25$)



Notes: The efficiency ratio is measured by $\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^2\}$. Vertical lines in the top panel and horizontal lines in the bottom panel indicate the the pseudo-true values of η and λ when η is also estimated. Horizontal lines in the top panel indicate the relative efficiency of the ML estimator: $1/\mathrm{M}_{ll}(\boldsymbol{\eta}_0)$. κ is the coefficient of multivariate excess kurtosis and π characterizes the Bernoulli mixture variate.

Figure 5a: Relative efficiency of ML / t PML vs Gaussian PML for t returns ($\nu = 8$)

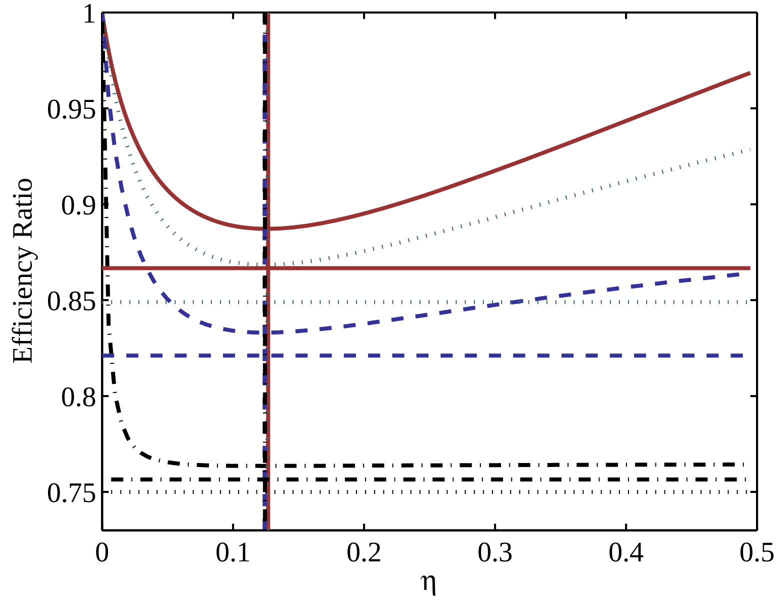
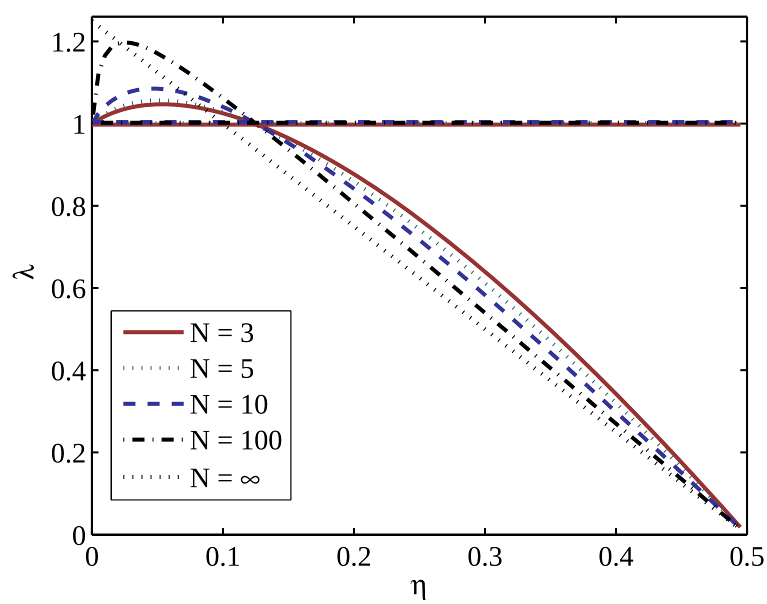
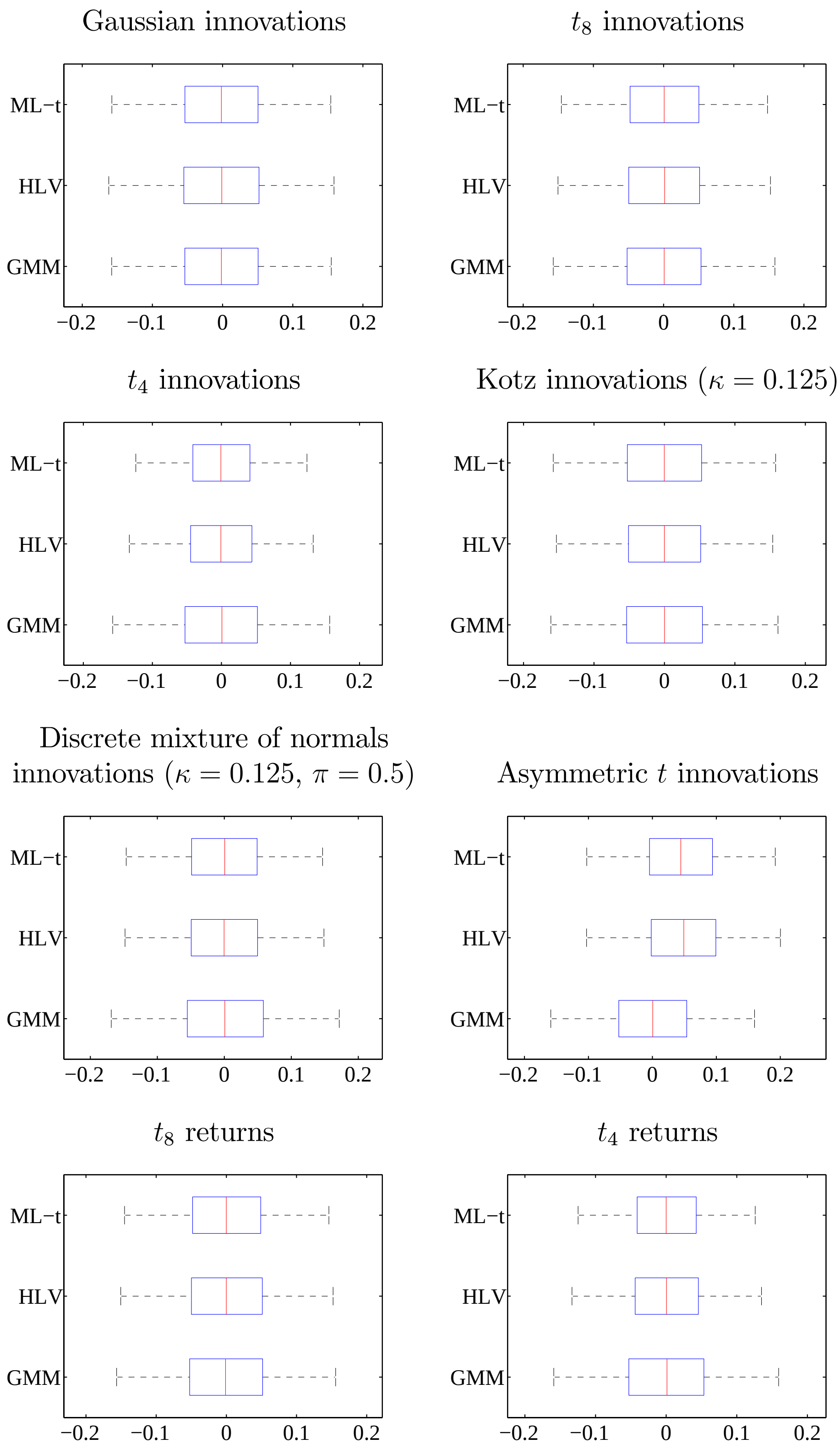


Figure 5b: Pseudo-true value of $\lambda_\infty(\bar{\eta}) = \tau_0/\tau_\infty(\bar{\eta})$ for t returns ($\nu = 8$)



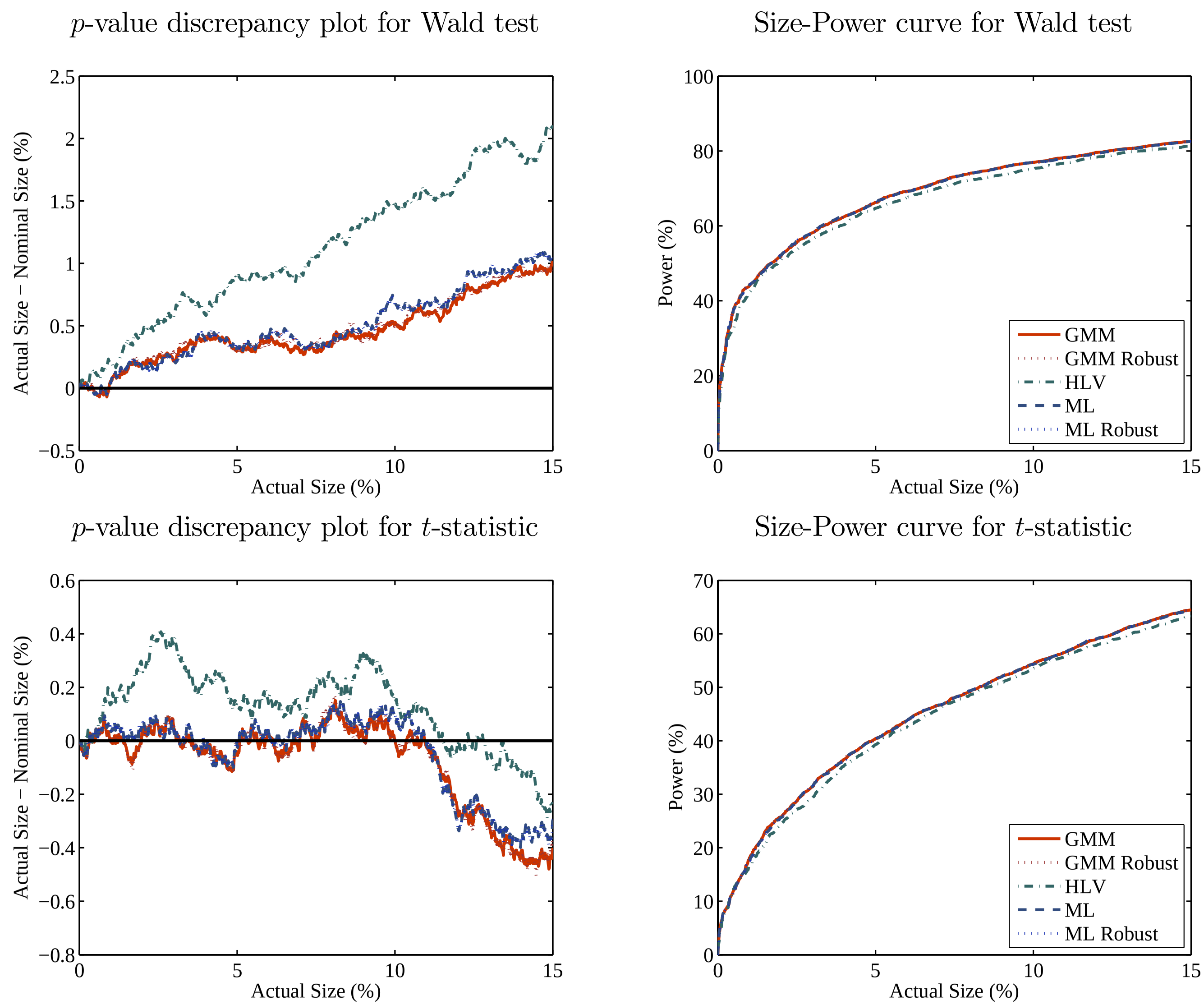
Notes: The efficiency ratio is measured by $\mathrm{M}^O_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}^H_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)]^2\}$. Vertical lines in the top panel and horizontal lines in the bottom panel indicate the the pseudo-true values of η and λ when η is also estimated. Horizontal lines in the top panel indicate the relative efficiency of the joint ML estimator described in Section 3.4.

Figure 6. Monte Carlo distributions of estimators of $\mathbf{a}$



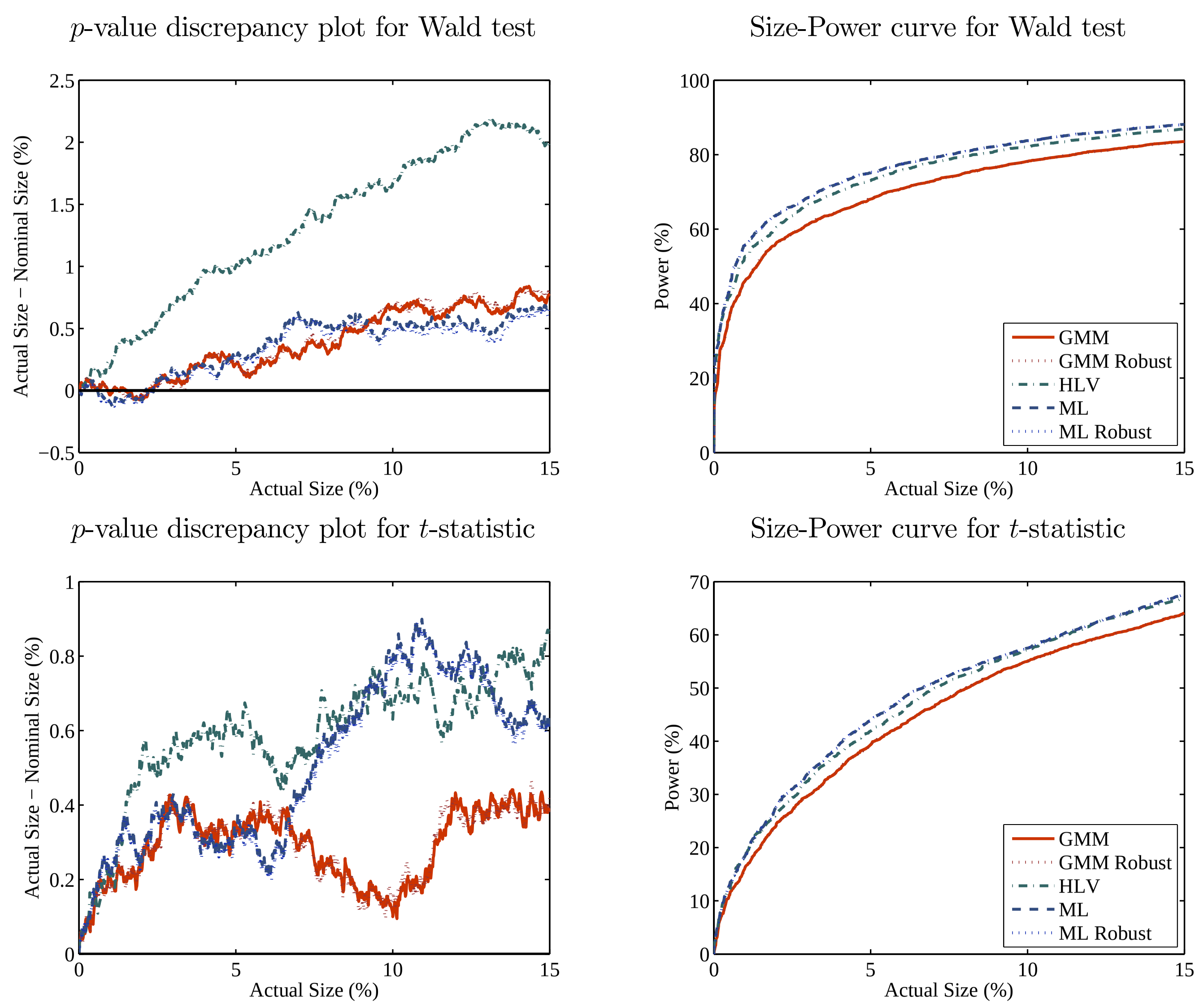
Notes: The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum lenght of the whiskers is one interquartile range. 10,000 replications. $T = 500$, $N = 5$, $\boldsymbol{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = \mathbf{0}$.

Figure 7a: p-value discrepancy and Size-Power plots for Gaussian innovations



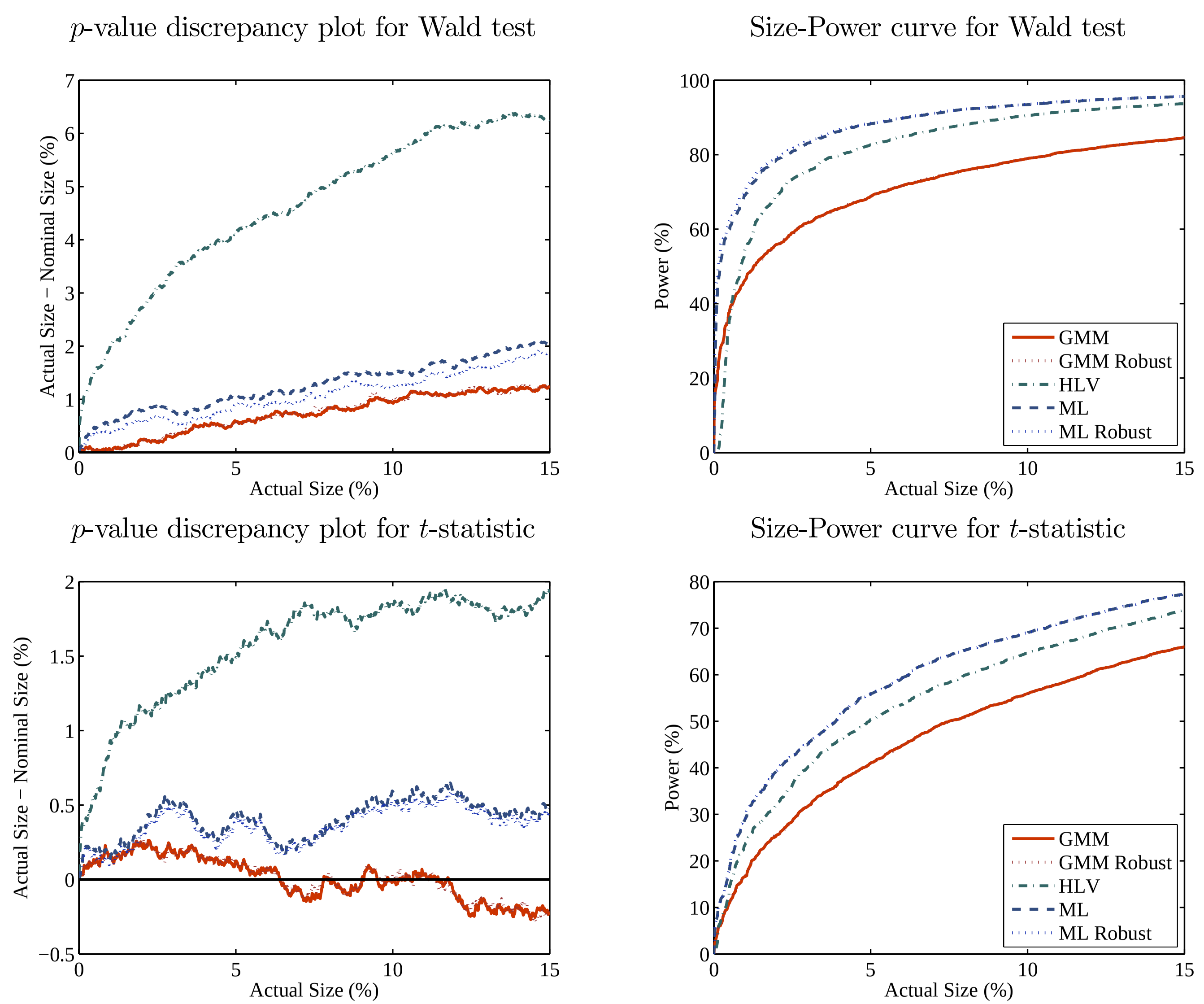
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis.

Figure 7b: p-value discrepancy and Size-Power plots for t_8 innovations



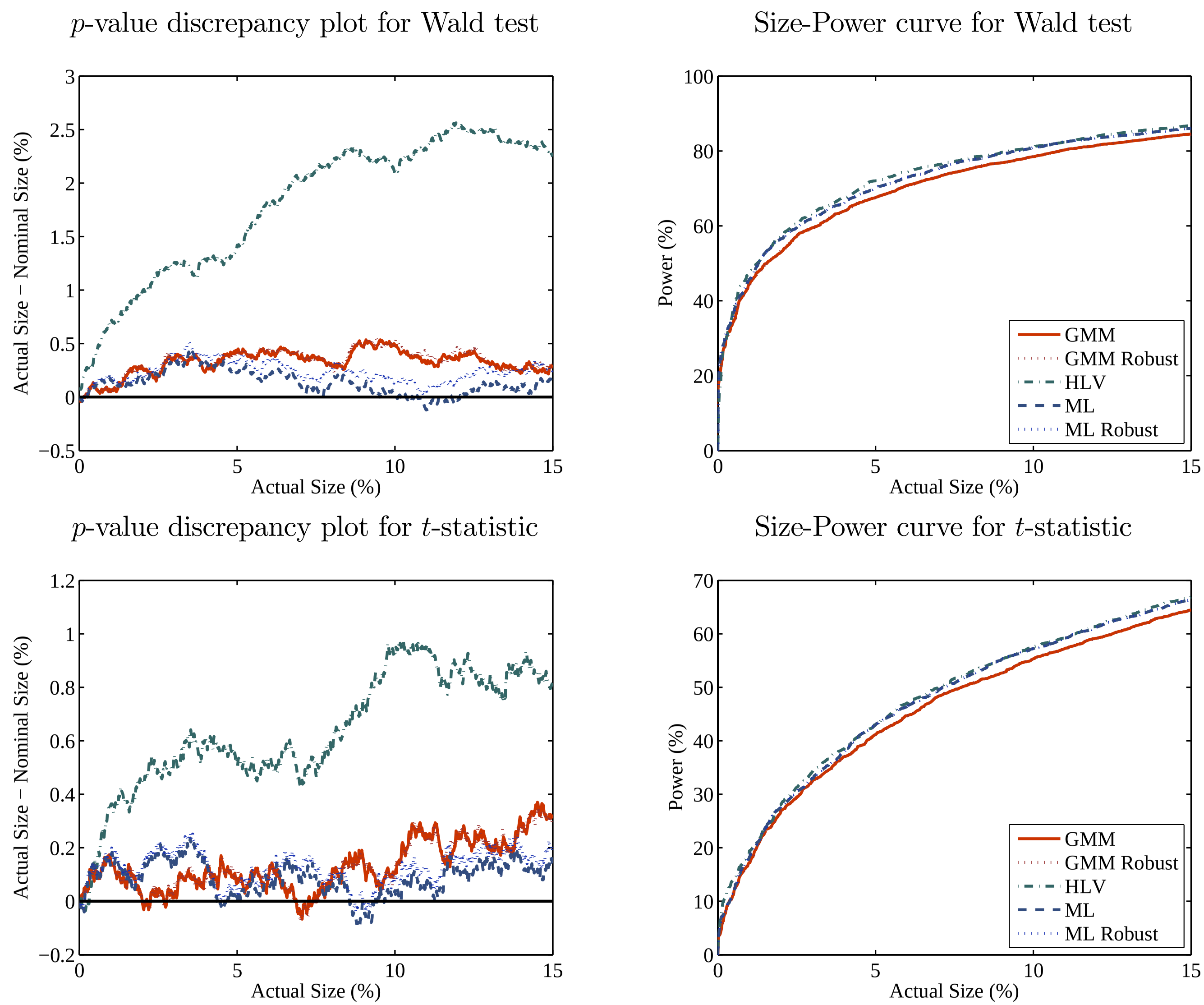
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis.

Figure 7c: p-value discrepancy and Size-Power plots for t_4 innovations



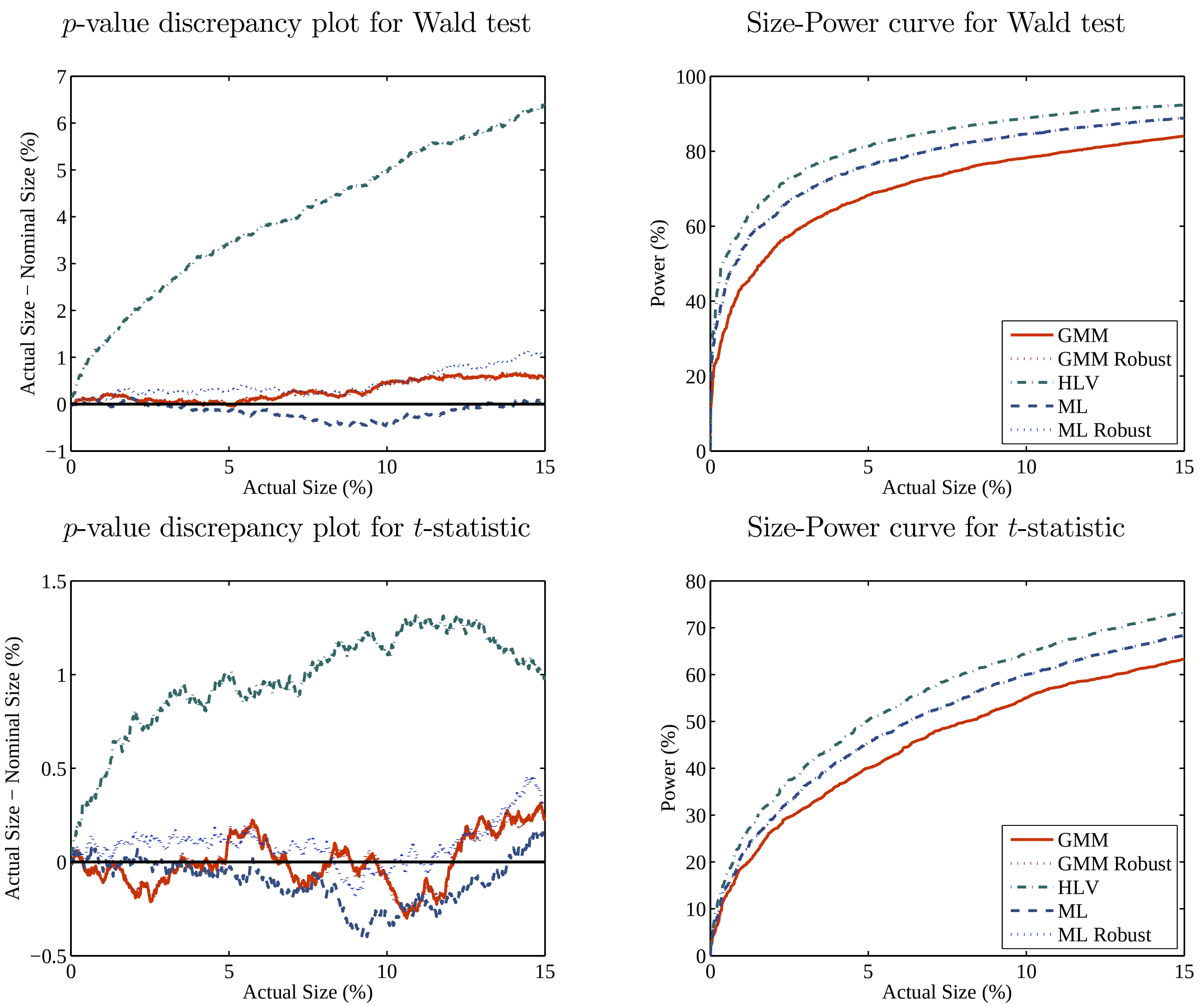
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis.

Figure 7d: p-value discrepancy and Size-Power plots for Kotz innovations ($\kappa = 0.125$)



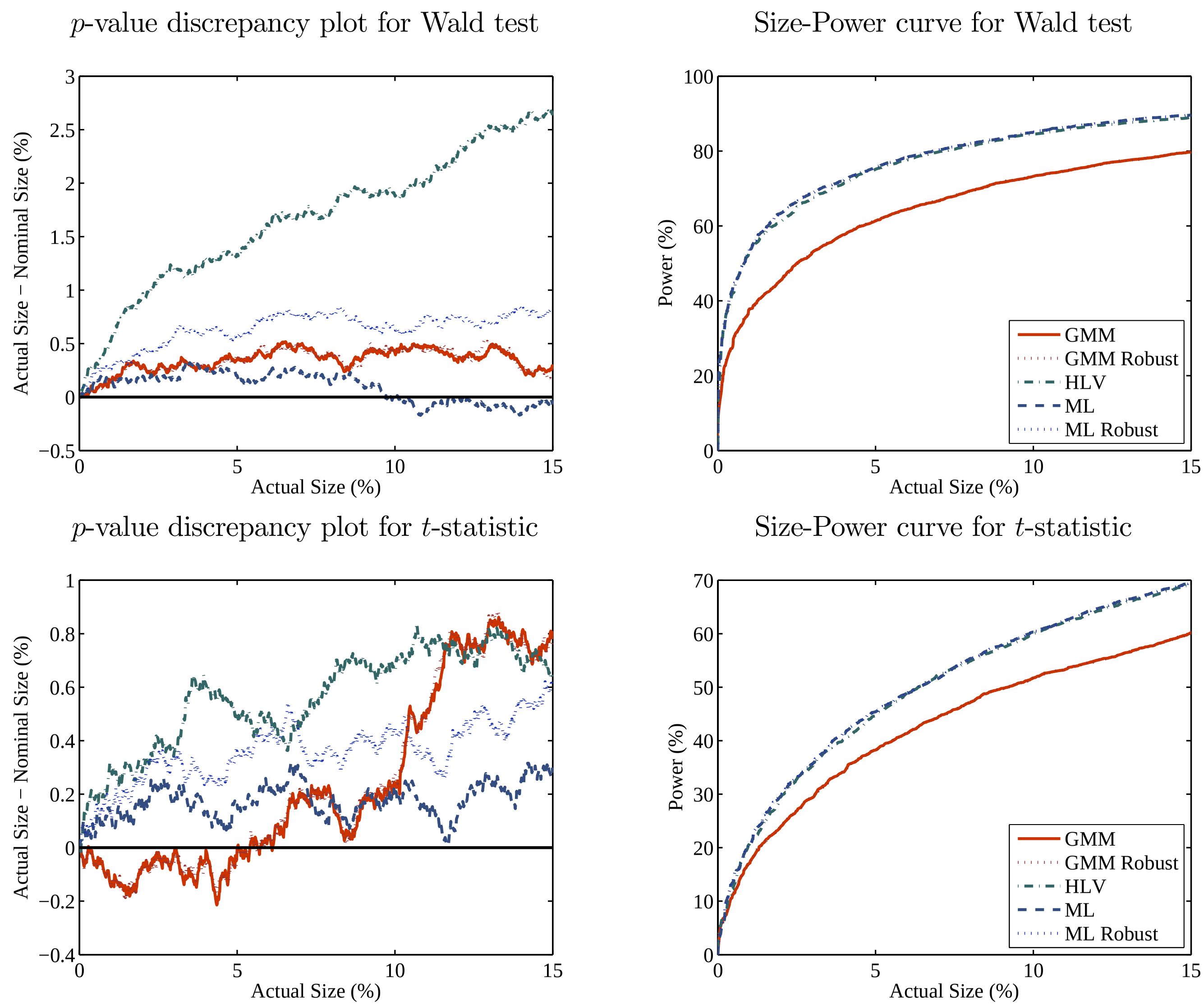
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis. κ denotes the coefficient of multivariate excess kurtosis.

Figure 7d: p-value discrepancy and Size-Power plots for Kotz innovations ($\kappa = 0.25$)



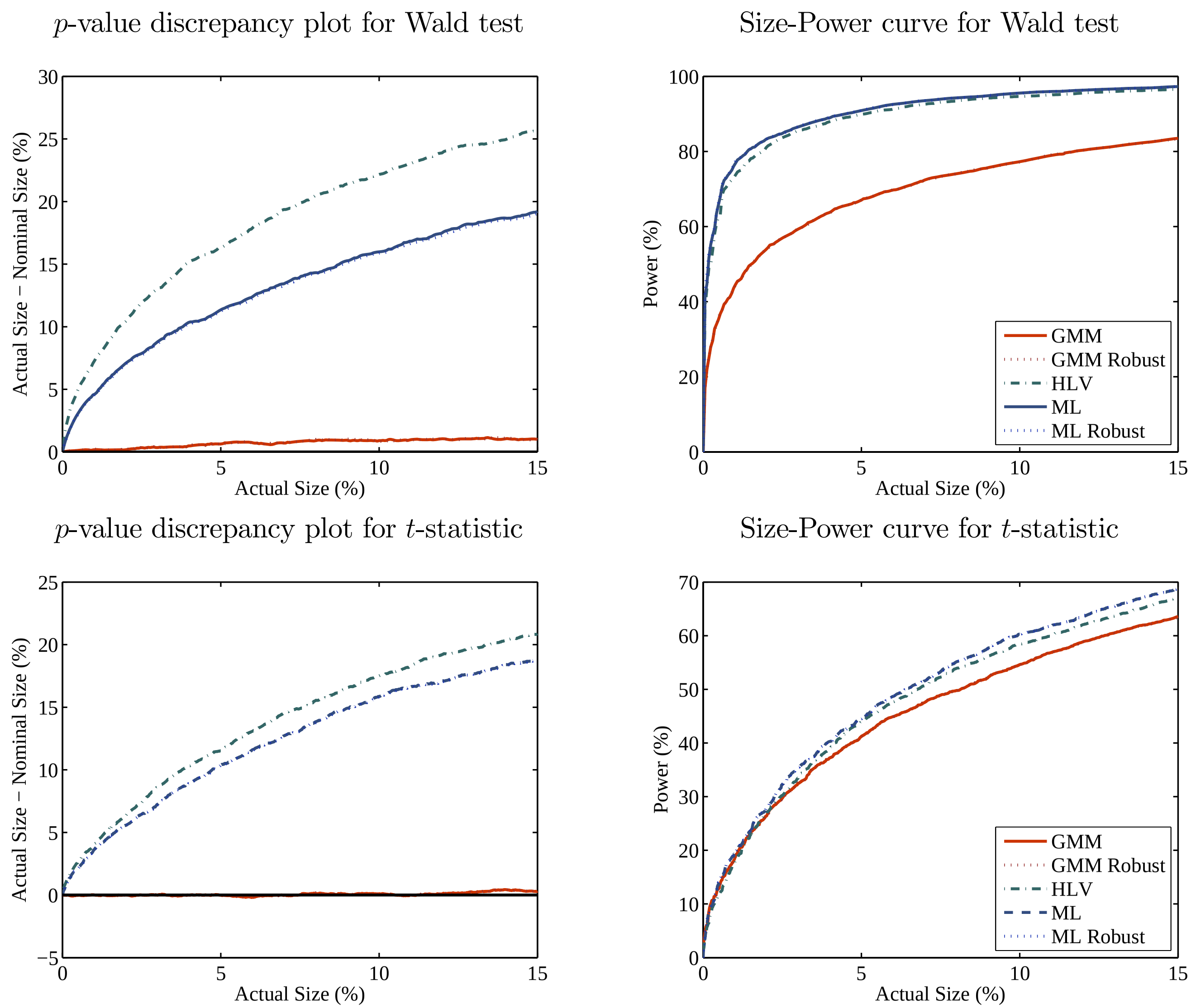
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis. κ denotes the coefficient of multivariate excess kurtosis.

Figure 7e: p-value discrepancy and Size-Power plots for Discrete mixture of normals innovations ($\kappa = 0.125$, $\pi = 0.5$)



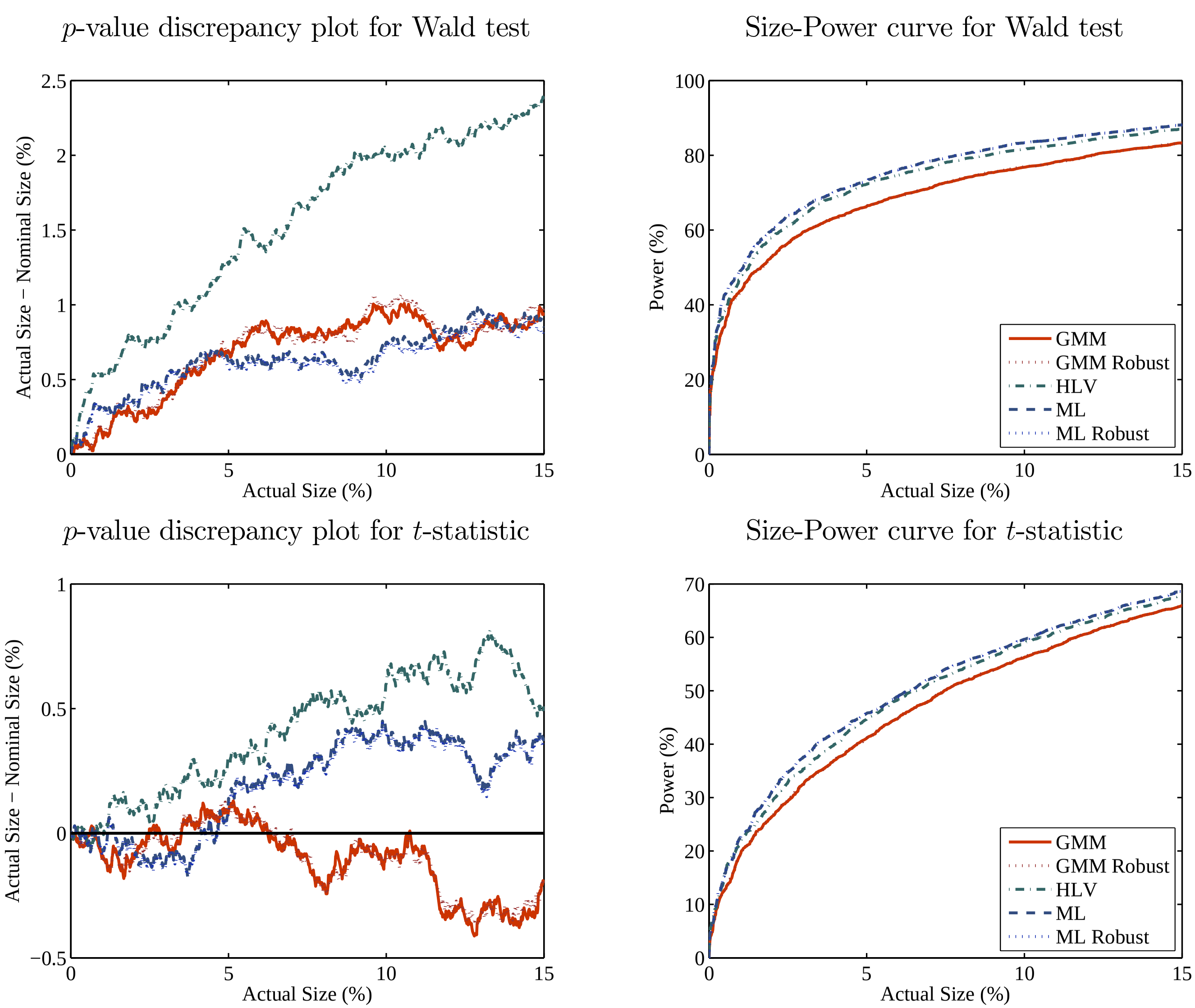
Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis. κ is the coefficient of multivariate excess kurtosis and π characterizes the Bernoulli mixture variate.

Figure 7f: p-value discrepancy and Size-Power plots for asymmetric t_8 innovations ($\boldsymbol{\beta} = -0.2 \times \ell_5$)



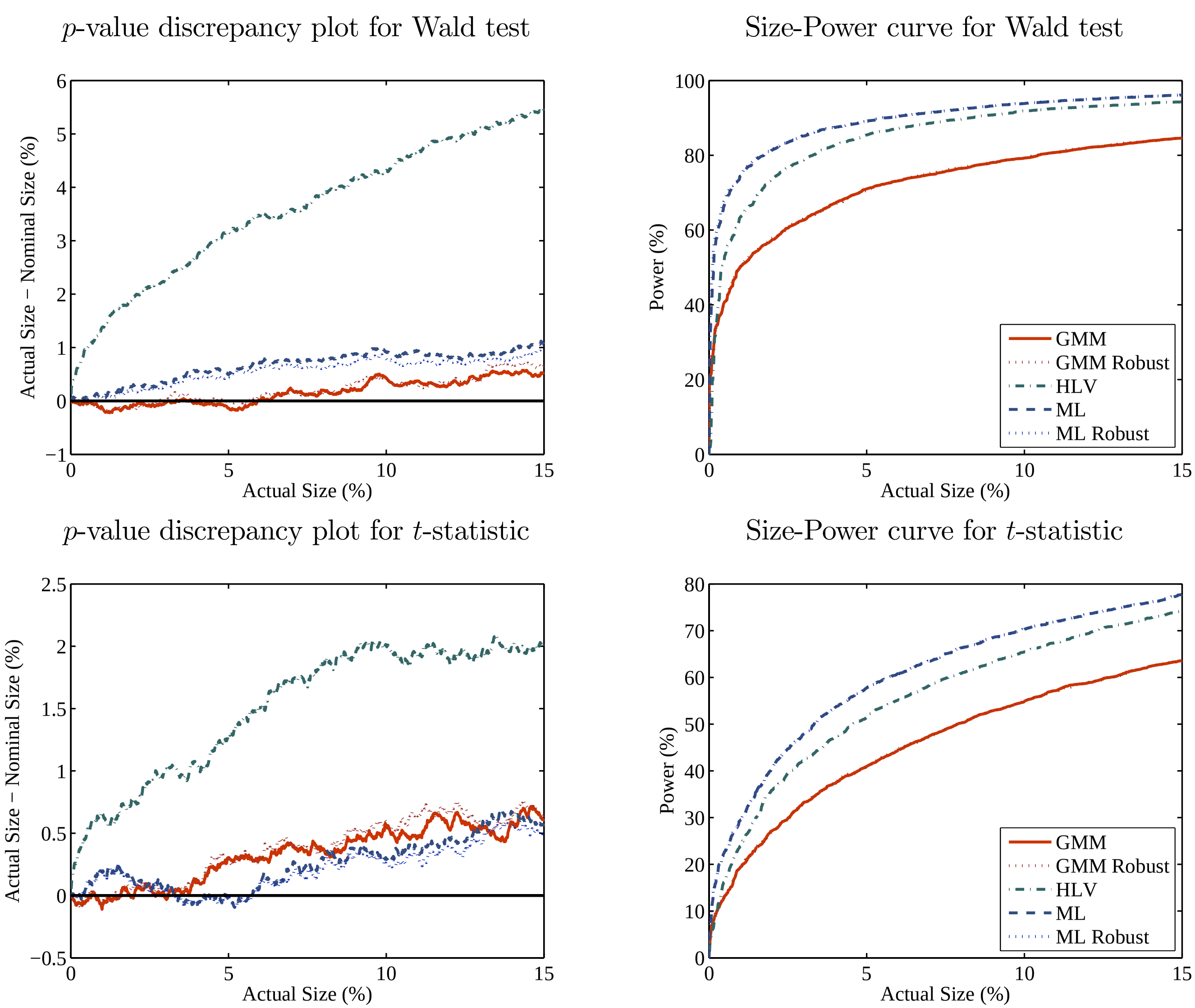
Notes: 10,000 replications. $T = 500$, $N = 5$, $\boldsymbol{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis. $\boldsymbol{\beta}$ introduces skewness in the distribution (see Section 3.3).

Figure 7g: p-value discrepancy and Size-Power plots for t_8 returns



Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis.

Figure 7h: p-value discrepancy and Size-Power plots for t_4 returns



Notes: 10,000 replications. $T = 500$, $N = 5$, $\mathbf{\Omega} = 4\sigma_M^2 \times \mathbf{I}_5$, $\sqrt{12}\mu_M/\sigma_M = \frac{1}{2}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = 4\mu_M \times \ell_5$ with $\ell_N = (1, \ldots, 1)'$ under the alternative hypothesis.

Supplemental Appendices for

A comparison of mean-variance efficiency tests

Dante Amengual
Department of Economics, Princeton University,
Fisher Hall, Princeton, NJ 08544-1021, USA
<amengual@princeton.edu>

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

April 2008

B Computation of the asymptotic efficiency of the t-based PML estimator when the true distribution of the innovations is elliptical

To compute the efficiency of the t-based ML estimator relative to the GMM estimator under ellipticity of the innovations, we first need to compute the pseudo-true values of the parameters. For a fixed value of $\eta > 0$, we know that $\mathbf{a}_\infty(\eta) = \mathbf{a}_0$, $\mathbf{b}_\infty(\eta) = \mathbf{b}_0$ and $\boldsymbol{\Omega}_\infty(\eta) = \lambda_\infty^{-1}(\eta)\boldsymbol{\Omega}_0$, where $\lambda_\infty(\eta)$ solves

$$E\left[\frac{N\eta+1}{1-2\eta+\eta\lambda_\infty(\eta)\varsigma}\frac{\lambda_\infty(\eta)\varsigma}{N}\right|\left.\boldsymbol{\phi}_0\right] = 1, \tag{B9}$$

with the expectation computed with respect to the true distribution of ς. This implicit equation is equivalent to the moment condition

$$E\left[\mathbf{s}_{\boldsymbol{\omega}t}(\mathbf{a}_0, \mathbf{b}_0, \lambda_\infty^{-1}(\eta)\boldsymbol{\omega}_0, \eta)\right|\left.\boldsymbol{\phi}_0\right] = \mathbf{0}$$

(see e.g. proof of Proposition 16 in Fiorentini and Sentana (2007)).

If η is not fixed, though, we will also have to compute the pseudo-true value of η, η_∞, say. If the innovations are distributed as a platykurtic elliptical random vector, then we know from Proposition 3 that $\eta_\infty = 0$ and $\lambda_\infty(0) = 1$. But when the innovations are drawn from a leptokurtic elliptical random vector instead, then under standard regularity conditions η_∞ can be understood as the value that makes

$$E\left[s_{\eta t}(\boldsymbol{\theta}_\infty, \eta_\infty)|\boldsymbol{\phi}_0\right] = 0, \tag{B10}$$

where

$$s_{\eta t}(\boldsymbol{\theta}, \eta) = \frac{\partial c(\eta)}{\partial \eta} + \frac{\partial g\left[\lambda_\infty \varsigma_t, \eta\right]}{\partial \eta}.$$

Fiorentini, Sentana and Calzolari (2003) show that for $\eta > 0$ this derivative is given by

$$\begin{aligned}
\frac{\partial c(\eta)}{\partial \eta} &= \frac{N}{2\eta(1-2\eta)} - \frac{1}{2\eta^2}\left[\psi\left(\frac{N\eta+1}{2\eta}\right) - \psi\left(\frac{1}{2\eta}\right)\right], \\
\frac{\partial g(\varsigma_t, \eta)}{\partial \eta} &= -\frac{N\eta+1}{2\eta(1-2\eta)}\frac{\varsigma_t}{1-2\eta+\eta\varsigma_t} + \frac{1}{2\eta^2}\log\left[1 + \frac{\eta}{1-2\eta}\varsigma_t\right],
\end{aligned}$$

where $\psi(.)$ is the di-gamma or Gauss' psi function (see Abramovich and Stegun (1964)).

In general, the presence of a log term implies that we must compute (B10) by numerical integration using recursive adaptive Simpson quadrature, where the required expectation is taken with respect to the true distribution of ς.

Unfortunately, both $\partial g(\varsigma_t, \eta)/\partial\eta$ and especially $\partial c(\eta)/\partial\eta$ are numerically unstable for η small, as documented by Fiorentini, Sentana and Calzolari (2003). For that reason, we follow their advice, and evaluate these expressions by means of the (directional) Taylor expansions around $\eta = 0$ in the following cases:

(i) if $\eta < 0.0008$, then use

$$\frac{\partial c_0(\eta)}{\partial\eta} = \frac{N(N+2)}{4} - \frac{N(N+2)(N-5)}{6}\eta + \frac{N(N+2)(N^2-6N+16)}{8}\eta^2$$

instead of $\partial c(\eta)/\partial\eta$, and

(ii) if $\eta < 0.03$ or $\eta\varsigma_t < 0.001$, then use

$$\begin{aligned}\frac{\partial g_0(\varsigma_t,\eta)}{\partial\eta} &= -\frac{N+2}{2}\varsigma_t + \frac{1}{4}\varsigma_t^2 \\ &+ \left[-2(N+2)\varsigma_t + \frac{N+4}{2}\varsigma_t^2 - \frac{1}{3}\varsigma_t^3\right]\eta \\ &+ \left[-12(N+2)\varsigma_t + 6(N+3)\varsigma_t^2 - (N+6)\varsigma_t^3 + \frac{1}{8}\varsigma_t^4\right]\frac{\eta^2}{2} \\ &+ \left[\begin{array}{c} -96(N+2)\varsigma_t + 24(3N+8)\varsigma_t^2 - 24(N+4)\varsigma_t^3 \\ +3(N+8)\varsigma_t^4 - \frac{12}{5}\varsigma_t^5 \end{array}\right]\frac{\eta^3}{6} \\ &+ \left[\begin{array}{c} -960(N+2)\varsigma_t + 600(2N+5)\varsigma_t^2 - 1440(3N+10)\varsigma_t^3 \\ +120(N+5)\varsigma_t^4 - 12(N+10)\varsigma_t^5 + 10\varsigma_t^6 \end{array}\right]\frac{\eta^4}{24}\end{aligned} \tag{B11}$$

instead of $\partial g(\varsigma_t, \eta)/\partial\eta$. Consequently, we evaluate (B10) as the weighted average of this expectation conditional on the complementary events $\varsigma_t < 0.001\eta_0$ and $\varsigma_t > 0.001\eta_0$ weighted by the corresponding probabilities. In many cases, both the expected value of (B11) conditional on $\varsigma_t < 0.001\eta_0$ and $P(\varsigma_t < 0.001\eta_0|\phi_0)$ can be computed analytically.

Having obtained the pseudo-true values, then we need to compute

$$\mathrm{M}_{II}^{H}[\eta, \lambda_\infty(\eta)] = E\left[\frac{N\eta+1}{1-2\eta+\eta\lambda_\infty(\eta)\varsigma_t}\left(1+\frac{2\eta}{1-2\eta+\eta\lambda_\infty(\eta)\varsigma_t}\frac{\lambda_\infty(\eta)\varsigma_t}{N}\right)\bigg|\,\boldsymbol{\phi}_0\right] \tag{B12}$$

and

$$\mathrm{M}_{II}^{O}[\eta, \lambda_\infty(\eta)] = E\left[\left(\frac{N\eta+1}{1-2\eta+\eta\lambda_\infty(\eta)\varsigma_t}\right)^2\frac{\lambda_\infty(\eta)\varsigma_t}{N}\bigg|\,\boldsymbol{\phi}_0\right]. \tag{B13}$$

It turns out that we can obtain analytical expressions for these expectations in the two examples that we consider in the paper.

B.1 Kotz innovations

As discussed in section 2.1, ς is Gamma distributed when the true innovations follow a Kotz distribution. Consequently, (B9), (B12) and (B13) can be decomposed in terms of the form

$$a \cdot E\left[\left(\frac{1}{b+dy}\right)^k y^h\right],$$

where $y = \alpha\varsigma/N$ is distributed as a standardized Gamma with parameter $\alpha = N[(N+2)\kappa + 2]^{-1}$, k and h are non-negative integers, and a, $b > 0$, and $d > 0$ are real constants. In fact we only need to find an analytical expression for $E\left[(1+cy)^{-k}\right]$ for $k = 1$ and $k = 2$, where $c = d/b > 0$, as

$$\frac{a}{b^k}E\left[\left(\frac{1}{1+cy}\right)^k y^h\right] = \frac{a}{b^k}\frac{\Gamma(\alpha+h)}{\Gamma(\alpha)}E\left[\frac{1}{(1+cy^*)^k}\right],$$

where $\Gamma(a)$ is the complete Gamma function and y^* a standardized Gamma with parameter $\alpha + h$.

To do so, we first compute the moment generating function of $1 + cy$, which is given by

$$M_{1+cy}(t) = E\left[e^{t(1+cy)}\right] = e^t E\left[e^{tcy}\right] = \frac{e^t}{(1-ct)^\alpha}$$

since $M_y(t) = E(e^{ty}) = (1-t)^{-\alpha}$. Then, we can exploit the result in equation (3) in Cressie, Davis, Folks and Policello (1981), which in our case yields

$$E\left[\frac{1}{(1+cy)^k}\right] = \frac{1}{\Gamma(k)}\int_0^\infty t^{k-1}M_{1+cy}(-t)dt$$

for any positive random variable y for which the above integral is well defined.

If we use the change of variable $s = t + c^{-1}$, so that $t = s - c^{-1}$, $cs = ct + 1$ and $ds = dc$, then we obtain that for $k = 1$,

$$E\left[\frac{1}{(1+cy)}\right] = \int_0^\infty \frac{e^{-t}}{(1+cy)^\alpha}dt = \frac{e^{c^{-1}}}{c^\alpha}\int_{c^{-1}}^\infty \frac{e^{-s}}{s^\alpha}ds = \frac{e^{c^{-1}}}{c^\alpha}\Gamma(1-\alpha, c^{-1}).$$

where $\Gamma(a, x)$ is the non-normalized incomplete Gamma function, which can be computed using standard software such as *Mathematica* or *Maple*. Similarly, for $k = 2$ we end up with

$$\begin{aligned}
E\left[\frac{1}{(1+cy)^2}\right] &= \int_0^\infty t\frac{e^{-t}}{(1+cy)^\alpha}dt \\
&= \int_{c^{-1}}^\infty (s-c^{-1})\frac{e^{-(s-c^{-1})}}{(cs)^\alpha}ds \\
&= \frac{e^{c^{-1}}}{c^\alpha}\left[\int_{c^{-1}}^\infty \frac{e^{-s}}{s^{\alpha-1}}ds - c^{-1}\int_{c^{-1}}^\infty \frac{e^{-s}}{s^\alpha}ds\right] \\
&= \frac{e^{c^{-1}}}{c^\alpha}\left[\Gamma(2-\alpha, c^{-1}) - c^{-1}\Gamma(1-\alpha, c^{-1})\right] \\
&= \frac{e^{c^{-1}}}{c^\alpha}\left\{\left[(1-\alpha) - c^{-1}\right]\Gamma(1-\alpha, c^{-1})\right\} + c^{-1}.
\end{aligned}$$

Finally, note that the terms $E[\varsigma^k|\varsigma < 0.001\eta_0^{-1};\boldsymbol{\phi}_0]$ that appear in the expectation of (**??**), together with $P[\varsigma < 0.001\eta_0^{-1}|\boldsymbol{\phi}_0]$ can be easily computed in terms of incomplete Gamma functions too.

B.2 Two-component scale mixture of normals

Since in this case ς is $Gamma(N/2, 1/2)$ conditional on the realization of the mixing variable s, we can use exactly the same formulas as in the case of the Kotz distribution, and then average across the two values of s. For instance,

$$\begin{aligned}\mathrm{M}_{II}^{H}[\eta,\lambda_\infty(\eta)] \equiv \pi E\left[\frac{N\eta+1}{1-2\eta+\eta\lambda_\infty(\eta)\varpi y}\left(1+\frac{2\eta}{1-2\eta+\eta\lambda_\infty(\eta)\varpi y}\frac{\lambda_\infty(\eta)\varpi y}{N}\right)\middle|\boldsymbol{\phi}_0, s=1\right] \\ +(1-\pi)E\left[\frac{N\eta+1}{1-2\eta+\eta\lambda_\infty(\eta)\varpi\varkappa y}\left(1+\frac{2\eta}{1-2\eta+\eta\lambda_\infty(\eta)\varpi\varkappa y}\frac{\lambda_\infty(\eta)\varpi\varkappa y}{N}\right)\middle|\boldsymbol{\phi}_0, s=0\right],\end{aligned}$$

where $\varpi\alpha y/N$ is distributed as a standardised Gamma with parameter $\alpha = N/2$.

C EM recursions for the multivariate t distribution

In this Appendix we specialise the expressions in Appendices B and D of Mencia and Sentana (2008) to the conditionally homoskedastic multivariate regression model with symmetric t innovations that we are considering. The rationale for using the EM algorithm comes from the fact that the model $\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*$, with $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1};\boldsymbol{\phi}_0 \sim i.i.d.\ t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ can be rewritten as

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \boldsymbol{\Omega}^{1/2}\sqrt{\frac{\nu_0-2}{\xi_t}}\boldsymbol{\varepsilon}_t^\circ$$

where $\boldsymbol{\varepsilon}_t^\circ|\xi_t, r_{Mt}, I_{t-1};\boldsymbol{\phi}_0 \sim N(\mathbf{0}, I_N)$ and $\xi_t|\boldsymbol{\phi}_0 \sim Gamma(\nu_0/2, 1/2)$.

Given that we know $f(\mathbf{r}_t|\xi_t,r_{Mt};\boldsymbol{\phi})$, $f(\xi_t|\boldsymbol{\phi})$ and $f(\mathbf{r}_t|r_{Mt};\boldsymbol{\phi})$, we can use Bayes theorem to obtain the distribution of ξ_t conditional on $\mathbf{r}_t$ and r_{Mt}. Specifically,

$$f(\xi_t|\mathbf{r}_t, r_{Mt};\boldsymbol{\phi}) = f(\mathbf{r}_t|\xi_t, r_{Mt};\boldsymbol{\phi})f(\xi_t|\boldsymbol{\phi})/f(\mathbf{r}_t|r_{Mt};\boldsymbol{\phi}) \propto f(\mathbf{r}_t|\xi_t, r_{Mt};\boldsymbol{\phi})f(\xi_t|\boldsymbol{\phi}).$$

Straightforward algebra shows that we can write

$$\begin{aligned} f(\xi_t|\mathbf{r}_t, r_{Mt};\boldsymbol{\phi}) &\propto \xi_t^{N/2}\exp\left[-\frac{\varsigma_t}{2}\frac{\eta}{1-2\eta}\xi_t\right]\xi_t^{\frac{1}{2\eta}-1}\exp\left(-\frac{\xi_t}{2}\right) \\ &\propto \xi_t^{\frac{N\eta+1}{2\eta}-1}\exp\left[-\frac{\xi_t}{2}\left(\frac{\eta\varsigma_t}{1-2\eta}+1\right)\right]\end{aligned}$$

where $\varsigma_t = (\mathbf{r}_t - \mathbf{a} - \mathbf{b}r_{Mt})'\boldsymbol{\Omega}^{-1}(\mathbf{r}_t - \mathbf{a} - \mathbf{b}r_{Mt})$, so that

$$\xi_t|\mathbf{r}_t, r_{Mt};\boldsymbol{\phi} \sim Gamma\left\{\frac{N\eta+1}{2\eta}, \frac{1}{2}\left[1+\frac{\eta\varsigma_t}{1-2\eta}\right]\right\}.$$

On this basis, we can show that the EM recursions with respect to $\mathbf{a}$, $\mathbf{b}$ and $\boldsymbol{\omega}$ will be given by

$$\begin{pmatrix} \mathbf{a}^{(i+1)} \\ \mathbf{b}^{(i+1)} \end{pmatrix} = \left\{ \left[\sum_{s=1}^{T} \xi_{s|s}^{(i)} \begin{pmatrix} 1 & r_{Ms} \\ r_{Ms} & r_{Ms}^2 \end{pmatrix} \right]^{-1} \otimes \mathbf{I}_N \right\} \sum_{t=1}^{T} \left\{ \left[\xi_{t|t}^{(i)} \begin{pmatrix} 1 \\ r_{Mt} \end{pmatrix} \right] \otimes \mathbf{r}_t \right\}$$

and

$$\boldsymbol{\omega}^{(i+1)} = vech \left[\frac{1}{T} \frac{\tilde{\eta}^{(i)}}{1-2\tilde{\eta}^{(i)}} \sum_{t=1}^{T} \xi_{t|t}^{(i)} (\mathbf{r}_t - \mathbf{a} - \mathbf{b} r_{Mt})(\mathbf{r}_t - \mathbf{a} - \mathbf{b} r_{Mt})' \right],$$

where

$$\xi_{t|t}^{(i)} = E[\xi_t | \mathbf{r}_t, r_{Mt}; \mathbf{a}^{(i)}, \mathbf{b}^{(i)}, \boldsymbol{\omega}^{(i)}, \tilde{\eta}^{(i)}] = \frac{N\tilde{\eta}^{(i)} + 1}{\tilde{\eta}^{(i)}} \left[\frac{\tilde{\eta}^{(i)} \varsigma_t}{1-2\tilde{\eta}^{(i)}} + 1 \right]^{-1}.$$

Although it is also possible to use the EM principle to update η, it involves numerical optimisation, so in practice it may be better to define $\tilde{\eta}^{(i+1)} = \arg\max L_T(\tilde{\boldsymbol{\theta}}^{(i+1)}, \eta)$ using $\tilde{\eta}^{(i)}$ as starting value. To initialise the EM recursions, we use the $\hat{\boldsymbol{\theta}}_{GMM}$ and the sequential ML estimator for η, $\hat{\eta}_{SML}$, which in turn we obtain using the MM estimator (22) as starting value.

D The information matrix for scale mixtures of normals

The density of ς when ε^* is a two-component scale mixture of normals is

$$h(\varsigma; \boldsymbol{\eta}) = \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \varsigma^{N/2-1} \left[\pi \exp\left(-\frac{1}{2\varpi}\varsigma \right) + (1-\pi)\varkappa^{-N/2} \exp\left(-\frac{1}{2\varpi\varkappa}\varsigma \right) \right],$$

where $\varpi = [\pi + \varkappa(1-\pi)]^{-1}$. If we combine $h(\varsigma; \boldsymbol{\eta})$ with expression (2.21) in Fang, Kotz and Ng (1990), then (5) follows. Hence,

$$\begin{aligned}
\mathrm{M}_{ll}(\boldsymbol{\eta}) &= E\left[\delta^2(\varsigma; \boldsymbol{\eta}) \frac{\varsigma}{N} \middle| \boldsymbol{\phi} \right] \\
&= \int_0^\infty \frac{1}{\varpi^2} \left\{ \pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma \right] \right\}^{-1} \\
&\quad \times \left\{ \pi^2 + 2\pi(1-\pi)\varkappa^{-(N/2+1)} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma \right] \right. \\
&\quad \left. + (1-\pi)^2 \varkappa^{-(N+2)} \exp\left[-\frac{1-\varkappa}{\varpi\varkappa}\varsigma \right] \right\} \\
&\quad \times \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{\varsigma^{N/2}}{N} \exp\left(-\frac{1}{2\varpi}\varsigma \right) d\varsigma \\
&= \mathrm{A}_1 + \mathrm{A}_2 + \mathrm{A}_3,
\end{aligned}$$

where

$$\mathrm{A}_1 = \frac{(2\varpi)^{-N/2}}{\varpi^2 \Gamma(N/2)} \pi^2 \int_0^\infty \left\{ \pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma \right] \right\}^{-1} \frac{\varsigma^{N/2}}{N} \exp\left(-\frac{1}{2\varpi}\varsigma \right) d\varsigma,$$

$$
\begin{aligned}
\mathrm{A}_2 &= \frac{(2\varpi)^{-N/2}}{\varpi^2\Gamma(N/2)} 2\pi(1-\pi) \int_0^\infty \left\{\pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\quad \times \varkappa^{-(N/2+1)} \frac{\varsigma^{N/2}}{N} \exp\left(-\frac{1}{2\varpi\varkappa}\varsigma\right) d\varsigma
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{A}_3 &= \frac{(2\varpi)^{-N/2}}{\varpi^2\Gamma(N/2)} (1-\pi)^2 \int_0^\infty \left\{\pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\quad \times \varkappa^{-(N+2)} \frac{\varsigma^{N/2}}{N} \exp\left[-\frac{2-x}{2\varpi\varkappa}\varsigma\right] d\varsigma.
\end{aligned}
$$

By analogy with Masoom and Nadarajah (2007), we can use the change of variable $v = \frac{1}{2\varpi\varkappa}(1-\varkappa)\varsigma$, so that $d\varsigma = 2\varpi\varkappa(1-\varkappa)^{-1}dv$, whence we get

$$
\begin{aligned}
\mathrm{A}_1 &= \frac{(2\varpi)^{-N/2}}{\varpi^2\Gamma(N/2)} \frac{1}{N} \pi \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp(-v)\right\}^{-1} v^{N/2} \exp\left(-\frac{\varkappa}{1-\varkappa}v\right) dv \\
&= \frac{1}{\varpi} \pi \left(\frac{\varkappa}{1-\varkappa}\right)^{N/2+1} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{\varkappa}{1-\varkappa}\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{A}_2 &= \frac{(2\varpi)^{-N/2}}{\varpi^2\Gamma(N/2)} 2(1-\pi) \frac{\varkappa^{-(N/2+1)}}{N} \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp(-v)\right\}^{-1} v^{N/2} \exp\left(-\frac{1}{1-\varkappa}v\right) dv \\
&= \frac{1}{\varpi} 2(1-\pi) \left(\frac{1}{1-\varkappa}\right)^{N/2+1} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{1}{1-\varkappa}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{A}_3 &= \frac{(2\varpi)^{-N/2}}{\varpi^2\Gamma(N/2)} \frac{(1-\pi)^2}{\pi} \frac{\varkappa^{-(N+2)}}{N} \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp(-v)\right\}^{-1} v^{N/2} \exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right] dv \\
&= \frac{1}{\varpi} \frac{(1-\pi)^2}{\pi} [\varkappa(1-\varkappa)]^{-(N/2+1)} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{2-\varkappa}{1-\varkappa}\right),
\end{aligned}
$$

where $F(z,s,r)$ denotes the Lerch function (see Erdelyi, 1981), which can be represented as

$$
F(z,s,r) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{v^{s-1}\exp(-rv)}{1-z\exp(-v)} dv.
$$

This function can be accurately computed using standard software such as *Mathematica*.

Therefore,

$$\begin{aligned}
\mathrm{M}_{ll}(\boldsymbol{\eta}) &= \frac{1}{\varpi}\pi\left(\frac{\varkappa}{1-\varkappa}\right)^{N/2+1} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{\varkappa}{1-\varkappa}\right) \\
&+\frac{2}{\varpi}(1-\pi)\left(\frac{1}{1-\varkappa}\right)^{N/2+1} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{1}{1-\varkappa}\right) \\
&+\frac{1}{\varpi}\frac{(1-\pi)^2}{\pi}[\varkappa(1-\varkappa)]^{-(N/2+1)} F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{2-\varkappa}{1-\varkappa}\right).
\end{aligned}$$

Similarly, we can use

$$\begin{aligned}
\frac{\partial\delta(\varsigma;\boldsymbol{\eta})}{\partial\varsigma} &= -\frac{1-\varkappa}{2\varpi^2\varkappa}\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\times(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right] \\
&+\frac{1-\varkappa}{2\varpi^2\varkappa}\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-2} \\
&\times\left\{\pi+(1-\pi)\varkappa^{-N/2+1}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\times(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]
\end{aligned}$$

to compute $\mathrm{M}_{ss}(\boldsymbol{\eta})$ from

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

with

$$\begin{aligned}
E\left[\frac{2\partial\delta[\varsigma;\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma^2}{N(N+2)}\right|\boldsymbol{\phi}\right] &= \int_0^\infty \frac{\varsigma^2}{N(N+2)}\left\{\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\right. \\
&\times\frac{(1-\varkappa)}{\varpi^2\varkappa}(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right] \\
&\times\left\{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\left.-\frac{(1-\varkappa)}{\varpi^2\varkappa}(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\times\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\varsigma^{N/2-1}\exp\left(-\frac{1}{2\varpi}\varsigma\right)d\varsigma \\
&= \mathrm{B}_1+\mathrm{B}_2+\mathrm{B}_3
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{B}_1 &= -\frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}(1-\pi)(1-\varkappa)\varkappa^{-(N/2+2)}\int_0^\infty \varsigma^{N/2+1}\exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}(1-\pi)(1-\varkappa)\varkappa^{-(N/2+2)}(2\varpi\varkappa)^{(N/2+2)}\Gamma\left(\frac{N}{2}+2\right) \\
&= -(1-\pi)(1-\varkappa)
\end{aligned}$$

$$
\begin{aligned}
\mathrm{B}_2 &= \frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}\pi(1-\pi)(1-\varkappa)\varkappa^{-(N/2+1)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2+1}\exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}(1-\pi)(1-\varkappa)\varkappa^{-(N/2+1)}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2+1}\exp\left[-\frac{1}{1-\varkappa}v\right]dv \\
&= (1-\pi)\varkappa\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{1}{1-\varkappa}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{B}_3 &= \frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}(1-\pi)^2(1-\varkappa)\varkappa^{-(N+2)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2+1}\exp\left[-\frac{2-\varkappa}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{(N+2)\Gamma(N/2)}\frac{1}{N\varpi^2}\frac{(1-\pi)^2}{\pi}(1-\varkappa)\varkappa^{-(N+2)}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2+1}\exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right]dv \\
&= \frac{(1-\pi)^2}{\pi}\varkappa^{-N/2}\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{2-\varkappa}{1-\varkappa}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{M}_{ss}(\boldsymbol{\eta}) &= -(1-\varkappa)(1-\pi) \\
&\quad+(1-\pi)\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{1}{1-\varkappa}\right) \\
&\quad+\frac{(1-\pi)^2}{\pi}\varkappa^{-N/2}\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{2-\varkappa}{1-\varkappa}\right).
\end{aligned}
$$

Finally, we can use

$$
\begin{aligned}
\frac{\partial\delta(\varsigma;\boldsymbol{\eta})}{\partial\pi} &= \varpi(1-\varkappa)\delta(\varsigma;\boldsymbol{\eta}) \\
&\quad+\frac{1}{\varpi}\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\quad\left\{1+\left[\frac{\varsigma}{2}(1-\pi)(1-\varkappa)^2\varkappa^{-(N/2+2)}-\varkappa^{-(N/2+1)}\right]\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\quad-\frac{1}{\varpi}\left\{1+\left[\frac{\varsigma}{2}(1-\pi)(1-\varkappa)^2\varkappa^{-(N/2+1)}-\varkappa^{-N/2}\right]\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\quad\times\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}^{-2} \\
&\quad\times\left\{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{(1-\varkappa)}{2\varpi\varkappa}\varsigma\right]\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \delta(\varsigma;\boldsymbol{\eta})}{\partial \varkappa} &= \varpi(1-\pi)\delta(\varsigma;\boldsymbol{\eta}) \\
&\quad -\left[\left(\frac{N}{2}+1\right)(1-\pi)\varkappa^{-(N/2+2)}+\frac{\varsigma}{2}\left[1-\pi(1-\varkappa^{-2})\right](1-\pi)\varkappa^{-(N/2+1)}\right] \\
&\quad \times\frac{1}{\varpi}\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right] \\
&\quad +\left[\frac{N}{2}(1-\pi)\varkappa^{-(N/2+1)}+\frac{\varsigma}{2}\left[1-\pi(1-\varkappa^{-2})\right](1-\pi)\varkappa^{-N/2}\right] \\
&\quad \times\frac{1}{\varpi}\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-2} \\
&\quad \times\left\{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]
\end{aligned}
$$

to compute

$$
\begin{aligned}
\mathrm{M}_{sr}(\boldsymbol{\eta}) &= -E\left[\frac{\varsigma_t(\boldsymbol{\theta})}{N}\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'}\middle|\boldsymbol{\phi}\right] \\
&= -E\left[\frac{\varsigma_t(\boldsymbol{\theta})}{N}\left(\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]}{\partial\pi},\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]}{\partial\varkappa}\right)\middle|\boldsymbol{\phi}\right].
\end{aligned}
$$

We then need

$$
\begin{aligned}
E\left[\frac{\varsigma}{N}\frac{\partial\delta(\varsigma,\boldsymbol{\eta})}{\partial\pi}\middle|\boldsymbol{\phi}\right] &= \int_0^\infty\frac{\varsigma}{N}\left\{(1-\varkappa)\left[\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right]\right. \\
&\quad +\frac{1}{\varpi}\left\{1+\left[\frac{\varsigma}{2}(1-\pi)(1-\varkappa)^2\varkappa^{-(N/2+2)}-\varkappa^{-(N/2+1)}\right]\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\quad -\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\quad \times\frac{1}{\varpi}\left\{1+\left[\frac{\varsigma}{2}(1-\pi)(1-\varkappa)^2\varkappa^{-(N/2+1)}-\varkappa^{-N/2}\right]\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\quad \left.\times\left[\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right]\right\} \\
&\quad \times\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\varsigma^{N/2-1}\exp\left(-\frac{1}{2\varpi}\varsigma\right)d\varsigma \\
&= \mathrm{C}_1+\mathrm{C}_2+\mathrm{C}_3+\mathrm{C}_4+\mathrm{C}_5+\mathrm{C}_6+\mathrm{C}_7+\mathrm{C}_8
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{C}_1 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{N}\left[(1-\varkappa)\pi+\frac{1}{\varpi}\right]\int_0^\infty\varsigma^{N/2}\exp\left(-\frac{1}{2\varpi}\varsigma\right)d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{(2\varpi)^{N/2+1}}{N}\left[(1-\varkappa)\pi+\frac{1}{\varpi}\right]\Gamma\left(\frac{N}{2}+1\right) \\
&= \varpi\pi(1-\varkappa)+1
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_2 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{N\varpi} \left[\varpi(1-\pi)(1-\varkappa)-1\right] \varkappa^{-(N/2+1)} \int_0^\infty \varsigma^{N/2} \exp\left(-\frac{1}{2\varpi\varkappa}\varsigma\right) d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{(2\varpi)^{N/2+1}}{N\varpi} \left[\varpi(1-\pi)(1-\varkappa)-1\right] \Gamma\left(\frac{N}{2}+1\right) \\
&= \varpi(1-\pi)(1-\varkappa)-1
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_3 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{2N\varpi} (1-\pi)(1-\varkappa)^2 \varkappa^{-(N/2+2)} \int_0^\infty \varsigma^{N/2+1} \exp\left(-\frac{1}{2\varpi\varkappa}\varsigma\right) d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{2N\varpi} (1-\pi)(1-\varkappa)^2 \, (2\varpi)^{N/2+2} \, \Gamma\left(\frac{N}{2}+2\right) \\
&= \varpi(1-\pi)(1-\varkappa)^2 \left(\frac{N}{2}+1\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_4 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{\pi}{N\varpi} \int_0^\infty \left\{\pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \varsigma^{N/2} \exp\left(-\frac{1}{2\varpi}\varsigma\right) d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{N\varpi} \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp\left[-v\right]\right\}^{-1} v^{N/2} \exp\left(-\frac{\varkappa}{1-\varkappa}v\right) dv \\
&= -\left(\frac{\varkappa}{1-\varkappa}\right)^{N/2+1} \mathcal{F}\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{\varkappa}{1-\varkappa}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_5 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{N\varpi} \left[(1-\pi)-\pi\varkappa\right] \varkappa^{-(N/2+1)} \\
&\quad \times \int_0^\infty \left\{\pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \varsigma^{N/2} \exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right] d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{N\varpi} \left[\frac{1-\pi}{\pi}-\varkappa\right] \varkappa^{-(N/2+1)} \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp\left[-v\right]\right\}^{-1} v^{N/2} \exp\left[-\frac{1}{1-\varkappa}v\right] dv \\
&= -\left[\frac{1-\pi}{\pi}-\varkappa\right] \left(\frac{1}{1-\varkappa}\right)^{N/2+1} \mathcal{F}\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+1, \frac{1}{1-\varkappa}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_6 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{N\varpi}\frac{\pi(1-\pi)}{2}\varkappa^{-(N/2+1)}(1-\varkappa)^2 \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2+1}\exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{N\varpi}\frac{(1-\pi)}{2}\varkappa^{-(N/2+1)}(1-\varkappa)^2\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2}\exp\left[-\frac{1}{1-\varkappa}v\right]dv \\
&= -\varpi(1-\pi)\varkappa\left(\frac{1}{1-\varkappa}\right)^{N/2}\left(\frac{N}{2}+1\right)\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{1}{1-\varkappa}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_7 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{N\varpi}(1-\pi)\varkappa^{-(N+1)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2}\exp\left[-\frac{2-\varkappa}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1-\pi}{\pi}\frac{\varkappa^{-(N+1)}}{N\varpi}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2}\exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right]dv \\
&= \frac{1-\pi}{\pi}\varkappa^{-N/2}\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+1,\frac{2-\varkappa}{1-\varkappa}\right);
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{C}_8 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}(1-\pi)^2(1-\varkappa)^2\varkappa^{-(N+2)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2+1}\exp\left[-\frac{2-\varkappa}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}\frac{(1-\pi)^2}{\pi}(1-\varkappa)^2\varkappa^{-(N+2)}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2+1}\exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right]dv \\
&= -\varpi\frac{(1-\pi)^2}{\pi}\varkappa^{-N/2}\left(\frac{1}{1-\varkappa}\right)^{N/2}\left(\frac{N}{2}+1\right)\digamma\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{2-\varkappa}{1-\varkappa}\right);
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\frac{\varsigma_t(\boldsymbol{\theta})}{N}\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]}{\partial\varkappa}\middle|\boldsymbol{\phi}\right] &= \int_0^\infty \frac{\varsigma}{N}\left\{(1-\pi)\left[\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right]\right. \\
&\quad -\left[\left(\frac{N}{2}+1\right)(1-\pi)+\frac{\varsigma}{2}\left[1-\pi(1-\varkappa^{-2})\right]\varkappa\right] \\
&\quad \times\frac{1}{\varpi}\varkappa^{-(N/2+2)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right] \\
&\quad +\left[\frac{N}{2}(1-\pi)+\frac{\varsigma}{2}\left[1-\pi(1-\varkappa^{-2})\right](1-\pi)\varkappa\right]\frac{\varkappa^{-(N/2+1)}}{\varpi} \\
&\quad \times\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \\
&\quad \left.\times\left\{\pi+(1-\pi)\varkappa^{-(N/2+1)}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\} \\
&\quad \times\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\varsigma^{N/2-1}\exp\left(-\frac{1}{2\varpi}\varsigma\right)d\varsigma \\
&= \mathrm{D}_1+\mathrm{D}_2+\mathrm{D}_3+\mathrm{D}_4+\mathrm{D}_5+\mathrm{D}_6+\mathrm{D}_7
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{D}_1 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{(1-\pi)\pi}{N}\int_0^\infty \varsigma^{N/2}\exp\left(-\frac{1}{2\varpi}\varsigma\right)d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{(1-\pi)\pi}{N}(2\varpi)^{N/2+1}\Gamma\left(\frac{N}{2}+1\right) \\
&= \varpi(1-\pi)\pi
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{D}_2 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{\varkappa^{-(N/2+2)}}{N}(1-\pi)\left[\frac{1}{\varpi}\left(\frac{N}{2}+1\right)-(1-\pi)\varkappa\right]\int_0^\infty \varsigma^{N/2}\exp\left(\frac{1}{2\varpi\varkappa}\varsigma\right)d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{\varkappa^{-(N/2+2)}}{N}(1-\pi)\left[\frac{1}{\varpi}\left(\frac{N}{2}+1\right)-(1-\pi)\varkappa\right](2\varpi\varkappa)^{N/2+1}\Gamma\left(\frac{N}{2}+1\right) \\
&= -(1-\pi)\frac{1}{\varkappa}\left[\left(\frac{N}{2}+1\right)-(1-\pi)\varkappa\varpi\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{D}_3 &= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}\left[1-\pi(1-\varkappa^{-2})\right](1-\pi)\varkappa^{-(N/2+1)}\int_0^\infty \varsigma^{N/2+1}\exp\left(\frac{1}{2\varpi\varkappa}\varsigma\right)d\varsigma \\
&= -\frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}\left[1-\pi(1-\varkappa^{-2})\right](1-\pi)\varkappa^{-(N/2+1)}(2\varpi\varkappa)^{N/2+2}\Gamma\left(\frac{N}{2}+2\right) \\
&= -\left(\frac{N}{2}+1\right)\varpi(1-\pi)\varkappa\left[1-\pi(1-\varkappa^{-2})\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{D}_4 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2\varpi}(1-\pi)\pi\varkappa^{-(N/2+1)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2}\exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2\varpi}(1-\pi)\varkappa^{-(N/2+1)}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2}\exp\left[-\frac{1}{1-\varkappa}v\right]dv \\
&= \frac{N}{2}(1-\pi)\left(\frac{1}{1-\varkappa}\right)^{N/2+1}\mathcal{F}\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+1,\frac{1}{1-\varkappa}\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{D}_5 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}\pi(1-\pi)\left[1-\pi(1-\varkappa^{-2})\right]\varkappa^{-N/2} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2+1}\exp\left[-\frac{1}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2N\varpi}(1-\pi)\left[1-\pi(1-\varkappa^{-2})\right]\varkappa^{-N/2}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2+1}\exp\left[-\frac{1}{1-\varkappa}v\right]dv \\
&= \varpi(1-\pi)\left[1-\pi(1-\varkappa^{-2})\right]\left(\frac{\varkappa}{1-\varkappa}\right)^2\left(\frac{1}{1-\varkappa}\right)^{N/2} \\
&\quad\times\left(\frac{N}{2}+1\right)\mathcal{F}\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+2,\frac{1}{1-\varkappa}\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{D}_6 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{\varpi}\frac{1}{2}(1-\pi)^2\varkappa^{-(N+2)} \\
&\quad\times\int_0^\infty\left\{\pi+(1-\pi)\varkappa^{-N/2}\exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1}\varsigma^{N/2}\exp\left[-\frac{2-\varkappa}{2\varpi\varkappa}\varsigma\right]d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)}\frac{1}{2\varpi}\frac{(1-\pi)^2}{\pi}\varkappa^{-(N+2)}\left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+1} \\
&\quad\times\int_0^\infty\left\{1+\frac{1-\pi}{\pi}\varkappa^{-N/2}\exp\left[-v\right]\right\}^{-1}v^{N/2}\exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right]dv \\
&= \frac{N}{2}\frac{(1-\pi)^2}{\pi}[\varkappa(1-\varkappa)]^{-(N/2+1)}\mathcal{F}\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2},\frac{N}{2}+1,\frac{2-\varkappa}{1-\varkappa}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{D}_7 &= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{2N\varpi} \left[1 - \pi(1 - \varkappa^{-2})\right] (1-\pi)^2 \varkappa^{-(N+1)} \\
&\quad \times \int_0^\infty \left\{\pi + (1-\pi)\varkappa^{-N/2} \exp\left[-\frac{1-\varkappa}{2\varpi\varkappa}\varsigma\right]\right\}^{-1} \varsigma^{N/2+1} \exp\left[-\frac{2-\varkappa}{2\varpi\varkappa}\varsigma\right] d\varsigma \\
&= \frac{(2\varpi)^{-N/2}}{\Gamma(N/2)} \frac{1}{2N\varpi} \frac{(1-\pi)^2}{\pi} \left[1 - \pi(1 - \varkappa^{-2})\right] \varkappa^{-(N+1)} \left(\frac{2\varpi\varkappa}{1-\varkappa}\right)^{N/2+2} \\
&\quad \times \int_0^\infty \left\{1 + \frac{1-\pi}{\pi}\varkappa^{-N/2} \exp\left[-v\right]\right\}^{-1} v^{N/2+1} \exp\left[-\frac{2-\varkappa}{1-\varkappa}v\right] dv \\
&= \left(\frac{N}{2}+1\right) \varpi \frac{(1-\pi)^2}{\pi} \left[1 - \pi(1 - \varkappa^{-2})\right] \varkappa^{-(N/2-1)} \left(\frac{1}{1-\varkappa}\right)^{N/2+2} \\
&\quad \times F\left(-\frac{1-\pi}{\pi}\varkappa^{-N/2}, \frac{N}{2}+2, \frac{2-\varkappa}{1-\varkappa}\right),
\end{aligned}
$$

CEMFI WORKING PAPERS

0601 *Beatriz Domínguez, Juan José Ganuza and Gerard Llobet*: “R&D in the pharmaceutical industry: a world of small innovations”.

0602 *Guillermo Caruana and Liran Einav*: “Production targets”.

0603 *Jose Ceron and Javier Suarez*: “Hot and cold housing markets: International evidence”.

0604 *Gerard Llobet and Michael Manove*: “Network size and network capture”.

0605 *Abel Elizalde*: “Credit risk models I: Default correlation in intensity models”.

0606 *Abel Elizalde*: “Credit risk models II: Structural models”.

0607 *Abel Elizalde*: “Credit risk models III: Reconciliation reduced – structural models”.

0608 *Abel Elizalde*: “Credit risk models IV: Understanding and pricing CDOs”.

0609 *Gema Zamarro*: “Accounting for heterogeneous returns in sequential schooling decisions”.

0610 *Max Bruche*: “Estimating structural models of corporate bond prices”.

0611 *Javier Díaz-Giménez and Josep Pijoan-Mas*: “Flat tax reforms in the U.S.: A boon for the income poor”.

0612 *Max Bruche and Carlos González-Aguado*: “Recovery rates, default probabilities and the credit cycle”.

0613 *Manuel Arellano and Jinyong Hahn*: “A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects”.

0614 *Manuel Arellano and Stéphane Bonhomme*: “Robust priors in nonlinear panel data models”.

0615 *Laura Crespo*: “Caring for parents and employment status of European mid-life women”.

0701 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: “Royalty stacking in high tech industries: separating myth from reality”.

0702 *Anne Layne-Farrar, A. Jorge Padilla and Richard Schmalensee*: “Pricing patents for licensing in standard setting organizations: Making sense of *FRAND* commitments”.

0703 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: “The *ex ante* auction model for the control of market power in standard setting organizations”.

0704 *Abel Elizalde*: “From Basel I to Basel II: An analysis of the three pillars”.

0705 *Claudio Michelacci and Josep Pijoan-Mas*: “The effects of labor market conditions on working time: the US-UE experience”.

0706 *Robert J. Aumann and Roberto Serrano*: “An economic index of *riskiness*”.

0707 *Roberto Serrano*: “El uso de sistemas dinámicos estocásticos en la Teoría de Juegos y la Economía”.

0708 *Antonio Cabrales and Roberto Serrano*: “Implementation in adaptive better-response dynamics”.

0709 *Roberto Serrano*: “Cooperative games: Core and Shapley value”.

0710 *Allan M. Feldman and Roberto Serrano*: “Arrow’s impossibility theorem: Preference diversity in a single-profile world”.

0711 *Victor Aguirregabiria and Pedro Mira*: “Dynamic discrete choice structural models: A Survey”.

0712 *Rene Saran and Roberto Serrano*: “The evolution of bidding behaviour in private-values auctions and double auctions”.

0713 *Gabriele Fiorentini and Enrique Sentana:* “On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models”.

0714 *Antonio Díaz de los Ríos and Enrique Sentana:* “Testing uncovered interest parity: A continuous-time approach”.

0715 *Francisco Peñaranda and Enrique Sentana:* “Duality in mean-variance frontiers with conditioning information”.

0716 *Stefano Gagliarducci, Tommaso Nannicini and Paolo Naticchioni:* “Electoral rules and politicians’ behavior: A micro test”.

0717 *Laura Hospido*: “Modelling heterogeneity and dynamics in the volatility of individual wages”.

0718 *Samuel Bentolila, Juan J. Dolado and Juan F. Jimeno*: “Does immigration affect the Phillips curve? Some evidence for Spain”.

0719 *Enrique Moral-Benito*: “Determinants of economic growth: A Bayesian panel data approach”.

0801 *David Martinez-Miera and Rafael Repullo*: “Does competition reduce the risk of bank failure?”.

0802 *Joan Llull*: “The impact of immigration on productivity”.

0803 *Cristina López-Mayán:* “Microeconometric analysis of residential water demand”.

0804 *Javier Mencía and Enrique Sentana:* “Distributional tests in multivariate dynamic models with Normal and Student *t* innovations”.

0805 *Javier Mencía and Enrique Sentana:* “Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation”.

0806 *Dante Amengual and Enrique Sentana:* “A comparison of mean-variance efficiency tests”.