Tests for serial dependence in static, non-Gaussian factor models*

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

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

October 2012
revised: January 2014

Abstract

We derive simple algebraic expressions for score tests of serial correlation in the levels and squares of common and idiosyncratic factors in static factor models with (semi) parametrically specified elliptical distributions even though one must generally compute the likelihood by simulation. We also robustify our Gaussian tests against non-normality. The orthogonality conditions resemble the orthogonality conditions of models with observed factors but the weighting matrices reflect their unobservability. Our Monte Carlo exercises assess the finite sample reliability and power of our proposed tests, and compare them to other existing procedures. Finally, we apply our methods to monthly US stock returns.

Keywords: ARCH, Financial returns, Kalman filter, LM tests, Non-Gaussian state space models, Predictability.

JEL: C32, C12, C13, C14, C38, C46, C58

*This is an abridged version of Fiorentini and Sentana (2012), which in turn was partly based on Fiorentini and Sentana (2009). We are grateful to Dante Amengual, Manuel Arellano, Javier Mencía, Javier Perote and Albert Satorra as well as to seminar audiences at Bologna, Cass Business School, CEMFI, Florence, GREQAM, Pablo de Olavide, Toulouse, Finance Forum 2008 (Madrid), ESEM 2009 (Barcelona), SAEe 2009 (Valencia), RCEF 2012 (Toronto) and the Oxford-Man Institute Time Series Econometrics Conference in honour of Andrew Harvey for helpful comments, discussions and suggestions. Detailed comments by Neil Shephard and two anonymous referees have also substantially improved the presentation. Of course, the usual caveat applies. Sentana gratefully acknowledges financial support from the Spanish Ministry of Science and Innovation through grants ECO 2008-00280 and 2011-26342, while Fiorentini acknowledges funding from MIUR PRIN MISURA - Multivariate models for risk assessment.

1 Introduction

There is a long tradition of factor or multi-index models in finance, where they were originally applied to simplify the computation of the covariance matrix of returns in a mean-variance portfolio allocation framework (see Connor, Goldberg and Korajczyk (2010) for a recent monograph). In this context, the common factors usually correspond to unobserved fundamental influences on returns, while the idiosyncratic factors reflect asset specific risks. In addition, the concept of factors plays a crucial role in two major asset pricing theories: the mutual fund separation theory (see e.g. Ross, 1978), of which the standard CAPM is a special case, and the Arbitrage Pricing Theory (see Ross (1976), and Connor (1984) for a unifying approach).

Factor models for low frequency financial returns are routinely estimated by Gaussian maximum likelihood under the assumption that the observations are serially independent using statistical factor analysis routines (see Lawley and Maxwell (1971)). In this context, the EM algorithm of Dempster, Laird and Rubin (1977) and Rubin and Thayer (1982) provides a cheap and reliable procedure for obtaining initial values as close to the optimum as desired, as illustrated by Lehmann and Modest (1988), who successfully employed this algorithm to handle a very large cross-sectional dataset of monthly returns on individual US stocks.

However, there are three empirical characteristics of assets returns which question the adequacy of this estimation procedure. First, there is some evidence of return predictability, which although far from uncontroversial, casts a doubt on the assumption of lack of serial correlation of common and idiosyncratic factors. Second, there is strong evidence that volatilities and correlations vary at high frequencies such as daily. Finally, many empirical studies with financial time series data indicate that the distribution of asset returns is rather leptokurtic, even after controlling for volatility clustering effects. In this context, the lack of normality implies that the Kalman filter prediction equations only provide the best linear least squares predictions and associated mean square errors, as opposed to the first two conditional moments (see Anderson and Moore (1979)), so that one cannot rely on standard results for Gaussian pseudo maximum likelihood estimators and tests, such as those in Bollerslev and Wooldridge (1992).

The objective of our paper is to provide joint diagnostic tests for serial dependence in the common and idiosyncratic factors that take into account the non-normality of asset returns. We will focus on Lagrange Multiplier (or score) tests, which only require estimation of the static model. As is well known, LM tests are asymptotically equivalent under the null and sequences of local alternatives to both Likelihood ratio and Wald tests, and therefore share their optimality properties. In this context, our main contribution is to derive simple algebraic expressions for the score tests of serial correlation in the levels and squares of common and idiosyncratic factors in static factor models when the distribution of the innovations in the latent variables is elliptically symmetric, which can be either parametrically or semiparametrically specified. Elliptical distributions are attractive in this context because they generalise the multivariate normal distribution while retaining its tractability irrespective of the number of assets. Importantly, our closed form tests are valid even though one must generally resort to simulation methods to approximate the log-likelihood function and its score in non-Gaussian state space models (see e.g. Durbin and Koopman (2000) and the references therein). In addition, we also explain how to robustify the Gaussian versions of our LM tests when the return distribution is not normal. Finally, we derive tests that focus on either the common factors or the specific factors, or indeed on some of their elements.

We proceed in steps. We initially derive (i) tests against AR/MA-type serial correlation in the latent factors under the maintained assumption that they are conditionally homoskedastic; (ii) tests against ARCH-type effects in those latent variables under the maintained assumption that they are serially uncorrelated; and (iii) joint tests of (i) and (ii) above. To keep the notation to a minimum, we focus on single factor models throughout, which suffice to illustrate our main results. Extensions to multiple factors are considered in Fiorentini and Sentana (2012). We complement our theoretical results with detailed Monte Carlo exercises to study the finite sample reliability and power of our proposed tests, and to compare them to other existing procedures. Finally, we also apply our methods to monthly stock returns on US broad industry portfolios.

The rest of the paper is organised as follows. First, we study the properties of likelihood-based estimators of the static factor model parameters under the null of serial independence. Then we derive tests against serial correlation in section 3, against conditional heteroskedasticity in section 4, and joint tests in section 5. A Monte Carlo evaluation of all the different tests can be found in section 6, followed by the empirical application to US sectorial stock returns in section 7. Finally, our conclusions, together with several interesting extensions, can be found in section 8. Proofs and auxiliary results are gathered in appendices.

2 Static factor models

Consider the following traditional (i.e. static, conditionally homoskedastic and exact) factor model:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}f_t + \boldsymbol{\Gamma}^{1/2}\mathbf{v}_t^*, \\ \begin{pmatrix} f_t \\ \mathbf{v}_t^* \end{pmatrix} | I_{t-1}; \boldsymbol{\phi}_s \sim s\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix}; \boldsymbol{\eta}\right] \end{array}\right\} \tag{1}$$

where $\mathbf{y}_t$ is an $N \times 1$ vector of observable variables with constant conditional mean $\boldsymbol{\pi}$, f_t is an unobserved common factor, whose constant variance, λ, we have normalised to 1 to avoid the usual scale indeterminacy,[1] $\mathbf{c}$ is the $N \times 1$ vector of factor loadings, $\mathbf{v}_t^*$ is a $N \times 1$ vector of standardised idiosyncratic noises, which are conditionally orthogonal to, but not necessarily independent from, f_t, $\mathbf{\Gamma}$ is a $N \times N$ diagonal positive semidefinite (p.s.d.) matrix of constant idiosyncratic variances, I_{t-1} is an information set that contains the values of $\mathbf{y}_t$ and f_t up to, and including time $t-1$, $\boldsymbol{\theta}_s = (\boldsymbol{\pi}',\mathbf{c}',\boldsymbol{\gamma}')'$, with $\boldsymbol{\gamma} = vecd(\mathbf{\Gamma})$ are the mean and variance parameters, $\boldsymbol{\eta}$ are some additional parameters that determine the shape of the conditional distribution of the spherical random vector $(f_t, \mathbf{v}_t^{*\prime})'$, which we assume has a well defined density, and $\boldsymbol{\phi}_s = (\boldsymbol{\theta}_s', \boldsymbol{\eta}')'$.

The most prominent example of spherical distribution is, of course, the standard normal distribution, which we assume corresponds to $\boldsymbol{\eta} = \mathbf{0}$. As in Bollerslev (1987) in a univariate context, and Harvey, Ruiz and Sentana (1992) in a multivariate one, followed by many others, we shall also consider in some detail a standardised multivariate Student t with ν degrees of freedom, or $i.i.d.$ $t(\mathbf{0}, \mathbf{I}_N, \nu)$ for short, which approaches the multivariate normal as $\nu \to \infty$, or η, its reciprocal, goes to 0. More flexible families of spherical distributions are discrete scale mixtures of normals and Laguerre polynomial expansions of the multivariate normal density (see Amengual, Fiorentini and Sentana (2013)), which could form the basis for a proper semiparametric procedure in which $\boldsymbol{\eta}$ would effectively be regarded as infinite dimensional.

Our assumptions trivially imply that

$$\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi}_s \sim s[\boldsymbol{\pi}, \mathbf{\Sigma}(\boldsymbol{\theta}_s), \boldsymbol{\eta}], \tag{2}$$

$$\mathbf{\Sigma}(\boldsymbol{\theta}_s) = \mathbf{c}\mathbf{c}' + \mathbf{\Gamma}, \tag{3}$$

where we have exploited the fact that linear combinations of elliptical random variables are elliptical (see thm. 2.16 in Fang, Kotz and Ng (1990)). As a result, if we define the standardised innovations

$$\boldsymbol{\varepsilon}_t^* = \mathbf{\Sigma}^{-1/2}(\boldsymbol{\theta}_{s0})(\mathbf{y}_t - \boldsymbol{\pi}_0) \tag{4}$$

as an $N-$dimensional martingale difference sequence that satisfies $E(\boldsymbol{\varepsilon}_t^* | I_{t-1}; \boldsymbol{\phi}_{s0}) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^* | I_{t-1}; \boldsymbol{\phi}_{s0}) = \mathbf{I}_N$, then $\boldsymbol{\eta}$ fully determines the shape of the conditional density of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$.

The most distinctive feature of these models is that they provide a parsimonious specification of the cross-sectional dependence in the observed variables,[2] which results in a significant reduction in the number of parameters, and allows the estimation of these models with a large number of series (see e.g. Lehmann and Modest (1988)). For these reasons, model (1) continues to be rather popular in empirical finance applications such as portfolio allocation, asset pricing tests,

[1] To free up the variance of the common factor, we can impose alternative restrictions as, for instance, $c_1 = 1$ or $\mathbf{c}'\mathbf{c} = 1$.

[2] See Sentana (2000) for a random field interpretation of factor models, and their time-series and cross-sectional state-space representations.

hedging and portfolio performance evaluation (see Connor, Goldberg and Korajczyk (2010) for details).

The parameters of interest are usually estimated jointly from the log-likelihood function of the observed variables. The ellipticity assumption and the serial independence of the variables involved imply that a modified version of the Kalman filter can still be used to estimate the underlying latent variables even though the innovations are not Gaussian. In particular, we can prove that:

$$E\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\middle| \mathbf{Y}_t;\boldsymbol{\phi}_s\right) = \left[\begin{array}{c} \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi}) \\ \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi}) \end{array}\right] = \left[\begin{array}{c} f_{kt}(\boldsymbol{\theta}_s) \\ \mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{array}\right], \tag{5}$$

and

$$V\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\middle| \mathbf{Y}_t;\boldsymbol{\phi}_s\right) = \left[\begin{array}{cc} \upsilon_{kt}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) & -\mathbf{c}'\upsilon_{kt}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) \\ -\mathbf{c}\upsilon_{kt}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) & \mathbf{c}\mathbf{c}'\upsilon_{kt}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) \end{array}\right] = \mathfrak{h}[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\cdot V\left(\begin{array}{c} f_t - f_{kt}(\boldsymbol{\theta}_s) \\ \mathbf{v}_t-\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{array}\middle| \boldsymbol{\phi}_s\right),$$

where $\mathbf{Y}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots\}$,

$$\mathbf{v}_t = \mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}f_t = \boldsymbol{\Gamma}^{1/2}\mathbf{v}_t^*,$$

$$\varsigma_t(\boldsymbol{\theta}_s) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_s)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_s) = (\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})$$

and $\mathfrak{h}(\varsigma_t;\boldsymbol{\eta})$ is a scalar factor of proportionality that reflects the non-linear dependence between the elements of a spherical random vector. For example, for the Student t

$$\mathfrak{h}(\varsigma_t;\boldsymbol{\eta}) = \frac{\nu-2}{\nu+N-2}\left(1+\frac{\varsigma_t}{\nu-2}\right),$$

which reduces to 1 under normality (see Harvey, Ruiz and Sentana (1992)). This scaling factor, whose unconditional mean is 1, multiplies the matrix of unconditional mean square errors

$$\begin{aligned} V\left(\begin{array}{c} f_t - f_{kt}(\boldsymbol{\theta}_s) \\ \mathbf{v}_t-\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{array}\middle| \boldsymbol{\phi}_s\right) &= \left[\begin{array}{cc} 1-\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & -\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \\ -\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}-\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \end{array}\right] \\ &= \left[\begin{array}{cc} \omega_k(\boldsymbol{\theta}_s) & -\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) \\ -\mathbf{c}\omega_k(\boldsymbol{\theta}_s) & \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) \end{array}\right], \end{aligned} \tag{6}$$

which has rank one because we are trying to infer $N+1$ latent variables from N observed ones. The elements of $f_{kt}(\boldsymbol{\theta}_s)$ and $\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$ are known as the "regression scores" in the factor analysis literature because the weights in (5) coincide with the coefficients in the theoretical regression of each unobserved variable onto the observed series, while (6) coincides with the unconditional residual covariance matrix from those regressions. As explained in Sentana (2004), the MSE criterion can be given an intuitive justification in terms of a mean-variance investor, since it corresponds to the so-called "tracking error" variability in the finance literature. In that sense, $f_{kt}(\boldsymbol{\theta}_s)$ are the excess returns to the portfolio that best "tracks" f_t, while $\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$ are the excess returns to the original vector of asset returns after we have hedged them against the common source of risk. As we shall see, $f_{kt}(\boldsymbol{\theta}_s)$, $\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$, $\omega_k(\boldsymbol{\theta}_s)$ and $\upsilon_{kt}(\boldsymbol{\theta}_s, \boldsymbol{\eta})$ constitute the basic ingredients of our tests.

In this context, we can formally characterise the asymptotic distribution of three likelihood-based estimators of the static model parameters: the usual maximum likelihood estimator that simultaneously estimates $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$; the elliptically symmetric semiparametric estimator of $\boldsymbol{\theta}$ considered by Hodgson and Vorkink (2003), Hafner and Rombouts (2007) and others, which restricts $\boldsymbol{\varepsilon}_t^*$ to have an *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N; \boldsymbol{\eta})$ conditional distribution but does not impose any structure on the distribution of ς_t;[3] and the Gaussian pseudo maximum likelihood estimator of $\boldsymbol{\theta}$, which sets $\boldsymbol{\eta} = \mathbf{0}$ even though the true conditional distribution of $\boldsymbol{\varepsilon}_t^*$ may well be non-normal.

Proposition 1 *Assume that (i) $V(\mathbf{y}_t)$ in (3) can be uniquely decomposed into $\mathbf{cc}'$ and $\boldsymbol{\Gamma}$, and (ii) the matrix*

$$[\boldsymbol{\Gamma} - \mathbf{c}(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\mathbf{c}'] \odot [\boldsymbol{\Gamma} - \mathbf{c}(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\mathbf{c}']$$

has full rank, where $\odot$ denotes the Hadamard product of two matrices of equal orders. Then:

1. *The asymptotic distribution of the maximum likelihood estimators $\hat{\boldsymbol{\theta}}_s$ and $\hat{\boldsymbol{\eta}}$ will be*

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_s - \boldsymbol{\phi}_{s0}) \to N[\mathbf{0}, \mathcal{I}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}(\boldsymbol{\phi}_{s0})],$$

where the information matrix $\mathcal{I}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}(\boldsymbol{\phi}_s)$ will be block diagonal between the elements corresponding to $\boldsymbol{\pi}$ and the elements corresponding to $(\mathbf{c}, \boldsymbol{\gamma}, \boldsymbol{\eta})$, with the first block given by $\mathrm{M}_{ll}(\boldsymbol{\eta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)$ and the second block by

$$\left[\begin{array}{c} \mathrm{M}_{ss}(\boldsymbol{\eta})\{[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{cc}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{cc}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) \\ \mathrm{M}_{ss}(\boldsymbol{\eta})\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] + \frac{1}{2}[\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\,vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \\ \mathrm{M}_{sr}'(\boldsymbol{\eta})\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) \end{array}\right.$$

$$\begin{array}{c} \mathrm{M}_{ss}(\boldsymbol{\eta})[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\mathbf{E}_N + \frac{1}{2}[\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) \\ \frac{1}{2}\mathrm{M}_{ss}(\boldsymbol{\eta})[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) \odot \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] + \frac{1}{4}[\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \\ \frac{1}{2}\mathrm{M}_{sr}'(\boldsymbol{\eta})vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \end{array}$$

$$\left.\begin{array}{c} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\mathrm{M}_{sr}(\boldsymbol{\eta}) \\ \frac{1}{2}vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\mathrm{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{array}\right],$$

where $\mathbf{E}_n$ is the unique $n^2 \times n$ "diagonalisation" matrix which transforms $vec(\mathbf{A})$ into

[3] The main advantage of this estimator over traditional semiparametric estimators is that one can obtain an estimate of the joint density of $\boldsymbol{\varepsilon}_t^*$ from a nonparametric estimate of the univariate density of ς_t, thereby avoiding the curse of dimensionality (see e.g. appendix B1 in Fiorentini and Sentana (2010a) for details).

$vecd(\mathbf{A})$ as $vecd(\mathbf{A})=\mathbf{E}_n'vec(\mathbf{A})$,

$$\mathrm{M}_{ll}(\boldsymbol{\eta})=E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}\middle|\boldsymbol{\phi}\right\}=E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t(\boldsymbol{\theta})}{N}+\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\middle|\boldsymbol{\phi}\right\}, \quad (7)$$

$$\mathrm{M}_{ss}(\boldsymbol{\eta})=\frac{N}{N+2}\left[1+V\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t}{N}\middle|\boldsymbol{\phi}\right\}\right]=E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)}\middle|\boldsymbol{\phi}\right\}+1, \quad (8)$$

$$\mathrm{M}_{sr}(\boldsymbol{\eta})=E\left[\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}-1\right\}\mathbf{e}_{rt}'(\boldsymbol{\phi})\middle|\boldsymbol{\phi}\right]=-E\left\{\frac{\varsigma_t(\boldsymbol{\theta})}{N}\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'}\middle|\boldsymbol{\phi}\right\}, \quad (9)$$

$$\begin{aligned}\mathcal{M}_{rr}(\boldsymbol{\eta})&=V[\,\partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}+\partial g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\,/\partial\boldsymbol{\eta}|\,\boldsymbol{\phi}]\\&=-E[\partial^2c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'+\partial^2g\,[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\,/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'|\,\boldsymbol{\phi}], \quad (10)\end{aligned}$$

$$\delta(\varsigma_t,\boldsymbol{\eta})=-2\partial g(\varsigma_t,\boldsymbol{\eta})/\partial\varsigma, \quad (11)$$

$c(\boldsymbol{\eta})$ is the constant of integration of the assumed elliptical density and $g(\varsigma_t,\boldsymbol{\eta})$ its kernel.

2. *Assuming that the population coefficient of multivariate excess kurtosis*

$$\kappa=E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)]-1 \quad (12)$$

is such that $-2/(N+2)<\kappa_0<\infty$, the efficiency bound associated to the elliptically symmetric semiparametric estimator $\mathring{\boldsymbol{\theta}}_s$ will be block diagonal between $\boldsymbol{\pi}$ and $(\mathbf{c},\boldsymbol{\gamma})$, where the first block coincides with the first block of the information matrix, and the second one with the corresponding block of the information matrix minus

$$\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]-\frac{4}{N\left[(N+2)\kappa+2\right]}\right\}$$
$$\times\begin{bmatrix}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\\ \frac{1}{2}vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) & vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\end{bmatrix}.$$

3. *If $\kappa_0<\infty$, the asymptotic distribution of the Gaussian pseudo maximum likelihood estimator $\bar{\boldsymbol{\theta}}_s$ will be*

$$\sqrt{T}(\bar{\boldsymbol{\theta}}_s-\bar{\boldsymbol{\phi}}_{s0})\to N[\mathbf{0},\mathcal{A}^{-1}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_{s0})\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi}_{s0})\mathcal{A}^{-1}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_{s0})],$$

where

$$\mathcal{A}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_s)=\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_s,\mathbf{0}),$$

$$\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_s,\mathbf{0})=\begin{bmatrix}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)+\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\mathbf{E}_N\\ \mathbf{0} & \mathbf{E}_N'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\odot\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\end{bmatrix}$$

and $\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi})$ has the same expression as $\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_s,\boldsymbol{\eta})$ but with $\mathrm{M}_{ll}(\boldsymbol{\eta})$ and $\mathrm{M}_{ss}(\boldsymbol{\eta})$ replaced by 1 and $(\kappa+1)$, respectively.

In the multivariate standardised Student t case, in particular, Fiorentini, Sentana and Calzolari (2003) show that:

$$\mathrm{M}_{ll}(\eta) = \frac{\nu(N+\nu)}{(\nu-2)(N+\nu+2)},\ \mathrm{M}_{ss}(\eta) = \frac{(N+\nu)}{(N+\nu+2)},\ \mathrm{M}_{sr}(\eta) = -\frac{2(N+2)\nu^2}{(\nu-2)(N+\nu)(N+\nu+2)},$$
$$\mathrm{M}_{rr}(\eta) = \frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{N+\nu}{2}\right)\right] - \frac{N\nu^4\left[\nu^2+N(\nu-4)-8\right]}{2(\nu-2)^2(N+\nu)(N+\nu+2)},$$

where $\psi(.)$ is the di-gamma function, which under normality reduce to 1, 1, 0 and $N(N+2)/2$, respectively.

Finally, it is worth mentioning that if we reparametrised the covariance matrix $\mathbf{\Sigma}(\boldsymbol{\theta}_s)$ as $\vartheta_2\mathbf{\Sigma}^\circ(\boldsymbol{\vartheta}_1)$, where ϑ_2 is a scalar and

$$\mathbf{\Sigma}^\circ(\boldsymbol{\vartheta}_1) = \mathbf{c}^*\mathbf{c}^{*\prime} + \mathbf{\Gamma}^*,$$

with $\mathbf{\Gamma} = \vartheta_2\mathbf{\Gamma}^*$ and $\mathbf{c} = \sqrt{\vartheta_2}\mathbf{c}^*$, Proposition 8 in Fiorentini and Sentana (2010a) implies that the maximum likelihood estimator and the elliptically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_1$ would be adaptive (i.e. as efficiently estimated as if we knew $\boldsymbol{\eta}$). If we further eliminated the resulting scale indeterminacy by forcing

$$|\mathbf{\Sigma}^\circ(\boldsymbol{\vartheta}_1)| = \left(\prod_{i=1}^N \gamma_i^*\right)\left(1+\sum_{j=1}^N c_j^{*2}/\gamma_j^*\right)$$

to be 1 (or any other fixed value),[4] the same proposition implies that the asymptotic covariance matrices of the three estimators of $\boldsymbol{\vartheta}_1$ and ϑ_2 considered in Proposition 1 would be block diagonal. Moreover, the ML estimator of ϑ_2 could only achieve the asymptotic efficiency of its Gaussian pseudo maximum likelihood estimator, which would be given by the expression:

$$\begin{aligned}
\vartheta_2(\boldsymbol{\vartheta}_1) &= \frac{1}{N}\frac{1}{T}\sum_{t=1}^T \varsigma_t^\circ(\boldsymbol{\vartheta}_1), \\
\varsigma_t^\circ(\boldsymbol{\vartheta}_1) &= (\mathbf{y}_t-\boldsymbol{\pi})'\mathbf{\Sigma}^{\circ-1}(\boldsymbol{\vartheta}_1)(\mathbf{y}_t-\boldsymbol{\pi}),
\end{aligned}$$

evaluated at the Gaussian PML estimator $\bar{\boldsymbol{\vartheta}}_1$.

[4]We can solve the resulting determinantal equation for one of the $c^{*\prime}s$, which yields

$$c_i^* = \pm\gamma_i^*\left(\frac{1}{\prod_{i=1}^N\gamma_i^*} - 1 - \sum_{j\neq i}^N c_j^{*2}/\gamma_j^*\right),$$

or for one of the $\gamma^{*\prime}s$, yielding

$$\gamma_j^* = \left(1 - c_j^2\prod_{i\neq j}\gamma_i^*\right)\Big/\left[\prod_{i\neq j}\gamma_i^*\left(1+\sum_{i\neq j}c_i^{*2}/\gamma_i^*\right)\right].$$

3 Serial correlation tests for common and idiosyncratic factors

3.1 Baseline case

The most natural way of introducing serial correlation in model (1) would be to assume that

$$\left.\begin{array}{r}\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t \\ x_t = \rho x_{t-1} + f_t \\ \mathbf{u}_t = diag(\boldsymbol{\rho}^*)\mathbf{u}_{t-1} + \boldsymbol{\Gamma}^{1/2}\mathbf{v}_t^*\end{array}\right\} \tag{13}$$

and

$$\begin{pmatrix} f_t \\ \mathbf{v}_t^* \end{pmatrix} | I_{t-1}, \boldsymbol{\phi} \sim s\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}, \boldsymbol{\eta}\right], \tag{14}$$

where the parameters of interest become $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_s', \boldsymbol{\rho}^{\dagger\prime})'$, with $\boldsymbol{\rho}^\dagger = (\rho, \boldsymbol{\rho}^{*\prime})'$, as this reduces to our baseline specification (1) under $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}$.

The problem with formulation (13) is that unless the true conditional distribution of the latent variables is Gaussian, the conditional distribution of the observed variables given their past values alone is unknown when $\boldsymbol{\rho}^\dagger \neq \mathbf{0}$ and $\boldsymbol{\Gamma}$ has full rank, and can only be approximated by simulation (see e.g. Durbin and Koopman (2000)). While it is true that the Kalman filter continues to produce the best linear least squares predictions of the underlying state variables in those circumstances (see Anderson and Moore (1979)), its prediction equations do not generally provide the conditional mean vector and covariance matrix of $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ (and the parameter values). As a result, we cannot rely on standard results for Gaussian pseudo maximum likelihood estimators and tests, such as those in Bollerslev and Wooldridge (1992). For that reason, in the rest of this section we assume that the mean vector and covariance matrix of $\mathbf{y}_t$ conditional on $\mathbf{Y}_{t-1}$ are given by the usual Kalman filter recursions (see appendix B in Fiorentini and Sentana(2012)), but the conditional distribution is elliptically symmetric. We will revisit this assumption in subsection 3.2.1.

Gaussian versions of dynamic factor models such as (13) have become increasingly popular in macroeconomic applications (see e.g. Bai and Ng (2008) and the references therein), but they are not widely used for stock returns (see Dungey, Martin and Pagan (2000) or Jegadeesh and Pennacchi (1996) for applications to bonds).

Assuming the stationarity conditions $|\rho| < 1$ and $|\rho_i^*| < 1 \; \forall i$ hold, the autocovariance matrices of the observed series will be:

$$\mathbf{G}_{\mathbf{y}}(j) = \mathbf{c}\mathbf{c}' G_x(j) + \mathbf{G}_{\mathbf{u}}(j). \tag{15}$$

The factor structure applies in particular to $\boldsymbol{\Sigma}$, the unconditional covariance matrix of $\mathbf{y}_t$, even though x_t or $\mathbf{u}_t$ are serially correlated (see Doz and Lenglart (1999)). It is also easy to see that the autocovariance structure in (15) corresponds to a special case of a VARMA(2,1) model since

$$(1-\rho L)[\mathbf{I}_N - diag(\boldsymbol{\rho}^*)L](\mathbf{y}_t - \boldsymbol{\pi}) = [\mathbf{I}_N - diag(\boldsymbol{\rho}^*)L]\mathbf{c}f_t + (1-\rho L)\mathbf{v}_t,$$

whose right hand side has the autocovariance structure of a VMA(1).[5]

As the next proposition shows, however, testing the null of multivariate white noise against such a complex VARMA(2,1) specification is extremely easy. Importantly, we shall distinguish between the optimal score test obtained by exploiting the non-normality of the conditional distribution, and the Gaussian pseudo LM test, which although uses the Gaussian scores, has been robustified against possible non-normality:

Proposition 2 *Let*

$$\bar{G}_f(j;\boldsymbol{\eta}) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}] f_{kt}(\boldsymbol{\theta}_s) f_{kt-j}(\boldsymbol{\theta}_s)$$

denote the sample cross moment of $\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}(\boldsymbol{\theta}_s)$ and $f_{kt-j}(\boldsymbol{\theta}_s)$, where $\delta(\varsigma_t,\boldsymbol{\eta})$ is defined in (11) and $f_{kt}(\boldsymbol{\theta}_s)$ is obtained from the updating equations (5) of the static factor model (1). Similarly, let

$$\bar{\mathbf{G}}_{\mathbf{v}}(j;\boldsymbol{\eta}) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}] \mathbf{v}_{kt}(\boldsymbol{\theta}_s) \mathbf{v}'_{kt-j}(\boldsymbol{\theta}_s)$$

denote the analogous sample cross moments for the specific factors.

1. *Under the null hypothesis $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}$, the score test statistic $LM_{AR(1)}(\boldsymbol{\eta}_0)$ given by T times*

$$\left(\bar{G}_f(1;\boldsymbol{\eta}_0), vecd'[\boldsymbol{\Gamma}_0^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(1;\boldsymbol{\eta}_0)\boldsymbol{\Gamma}_0^{-1/2}]\right) \mathcal{I}^{-1}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_{s0},\mathbf{0};\boldsymbol{\eta}_0) \left(\bar{G}_f(1;\boldsymbol{\eta}_0), vecd'[\boldsymbol{\Gamma}_0^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(1;\boldsymbol{\eta}_0)\boldsymbol{\Gamma}_0^{-1/2}]\right)',$$

with

$$\mathcal{I}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) = \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) \odot \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}),$$

where

$$\begin{aligned}
\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) &= V\begin{bmatrix} \delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}(\boldsymbol{\theta}_s) \\ \delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{bmatrix} = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}), \\
\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) &= V\begin{bmatrix} f_{kt}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \end{bmatrix} = \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}),
\end{aligned}$$

[5] When $\boldsymbol{\rho}^* = \rho\mathbf{I}_N$, though, the reduced form process becomes a VAR(1) with a scalar companion matrix. As a result, any linear combination of $\mathbf{y}_t$ will have the autocorrelation structure of an AR(1) process with autoregressive coefficient ρ.

will be distributed as a χ^2 with $N+1$ degrees of freedom for N fixed as T goes to infinity. Moreover, this asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_{s0}$ and $\boldsymbol{\eta}_0$ by their feasible maximum likelihood estimators in the first part of Proposition 1.

2. *It also remains valid if we replace $\boldsymbol{\theta}_{s0}$ by its elliptically symmetric semiparametric estimator in the second part of Proposition 1, which requires the nonparametric estimation of the density of $\varsigma_t(\boldsymbol{\theta}_s)$.*

3. *Under the same null hypothesis, the Gaussian pseudo score test statistic $LM_{AR(1)}(\mathbf{0})$ given by T times*

$$\left(\bar{G}_f(1;\mathbf{0}), vecd'[\boldsymbol{\Gamma}_0^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(1;\mathbf{0})\boldsymbol{\Gamma}_0^{-1/2}]\right)\mathcal{I}^{-1}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_{s0},\mathbf{0};\mathbf{0})\left(\bar{G}_f(1;\mathbf{0}), vecd'[\boldsymbol{\Gamma}_0^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(1;\mathbf{0})\boldsymbol{\Gamma}_0^{-1/2}]\right)',$$

will be distributed as a χ^2 with $N+1$ degrees of freedom for N fixed as T goes to infinity irrespective of the normality of the conditional distribution. This result continues to hold if we replace $\boldsymbol{\theta}_{s0}$ by its Gaussian pseudo maximum likelihood estimator $\bar{\boldsymbol{\theta}}_s$ in the third part of Proposition 1.

Researchers may sometimes be interested in tests that separately assess the serial correlation of either the common factor or the specific factors. In principle, they might even like to focus on a particular v_{it}. By combining the relevant elements of $\bar{G}_f(j;\boldsymbol{\eta})$ and $\bar{\mathbf{G}}_{\mathbf{v}}(1;\boldsymbol{\eta})$ with the corresponding blocks of the information matrix, $\mathcal{I}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_{s0},\mathbf{0};\boldsymbol{\eta})$, we can easily exploit the results in Proposition 2 to derive the required test statistics for those subcomponents under the maintained hypothesis of serial independence.[6] Intuitively, the reason is that we can interpret $LM_{AR(1)}(\boldsymbol{\eta})$ as a test based on the $N+1$ orthogonality conditions:

$$E\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}(\boldsymbol{\theta}_s)f_{kt-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\} = 0, \tag{16}$$

$$E\{\gamma_i^{-1}\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]v_{kit}(\boldsymbol{\theta}_s)v_{kit-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\} = 0 \qquad (i=1,\ldots,N). \tag{17}$$

Similarly, $LM_{AR(1)}(\mathbf{0})$ is based on

$$E\left[f_{kt}(\boldsymbol{\theta}_s)f_{kt-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\right] = 0, \tag{18}$$

$$E\left[\gamma_i^{-1}v_{kit}(\boldsymbol{\theta}_s)v_{kit-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\right] = 0 \qquad (i=1,\ldots,N), \tag{19}$$

which are the conditions that we would use to test for first order serial correlation if we treated $f_{kt}(\boldsymbol{\theta}_s)$ and $v_{kit}(\boldsymbol{\theta}_s)$ as the series of interest in the Gaussian case (see Breusch and Pagan (1980) or Godfrey (1988)). In that sense, the factor $\delta(\varsigma_t,\boldsymbol{\eta})$, which is equal to 1 under Gaussianity and

[6] See Bera and Yoon (1993) for a possible way of orthogonalising those individual LM test under alternatives local to $H_0:\boldsymbol{\rho}^\dagger=\mathbf{0}$.

to $(N\eta+1)/(1-2\eta+\eta\varsigma_t)$ for the Student t, can be regarded as the type of damping factor for big observations used in the robust estimation literature (see e.g. Maronna, Martin and Yohai (2006)) because it is a decreasing function of ς_t for fixed $\eta > 0$, the more so the higher η is (see Fiorentini and Sentana (2010b) for a closely related discussion for univariate models).

Given that we have fixed the variance of the innovations in the common factor to 1, the moment conditions (18) and (19) closely resemble

$$\begin{aligned} E(f_t f_{t-1}|\boldsymbol{\theta}_s, \mathbf{0}) &= 0, \\ E(\gamma_i^{-1} v_{it} v_{it-1}|\boldsymbol{\theta}_s, \mathbf{0}) &= 0 \qquad (i = 1, \ldots, N), \end{aligned}$$

which are the Gaussian-based orthogonality conditions that we could use to test for first order serial correlation if we could observe all the latent variables.

The similarity between these two sets of moment conditions becomes even stronger if we consider individual tests for serial correlation in each latent variable. Let us start with a test of $H_0 : \rho = 0$ under the maintained assumption that $\boldsymbol{\rho}^* = \mathbf{0}$. Part 3 of Proposition 2 implies that the asymptotic variance of $\bar{G}_f(1; \mathbf{0})$ is simply $[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2$. But we can use (6) to interpret $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$ as the R^2 in the theoretical least squares projection of f_t on a constant and $\mathbf{y}_t$. Therefore, the higher the degree of observability of the common factor, the closer the asymptotic variance of $\bar{G}_f(1; \mathbf{0})$ will be to 1, which is the asymptotic variance of the first sample autocorrelation of f_t. Intuitively, this convergence result simply reflects the fact that the common factor becomes observable in the limit, which implies that our Gaussian test of $H_0 : \rho = 0$ will become arbitrarily close to a Gaussian first order serial correlation test for the common factor as the "signal to noise" ratio $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$ approaches 1. Before the limit, though, our test takes into account the unobservability of f_t. A particularly interesting situation arises if we consider models in which N is large. Since $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} = (\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})/[1 + (\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})]$ under the assumption that $\boldsymbol{\Gamma}$ has full rank, the aforementioned R^2 converges to 1 as N grows because $(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) \to \infty$ in those circumstances due to the pervasive nature of the common factor (see e.g. Sentana (2004)).

Likewise, part 3 of Proposition 2 implies that the asymptotic variance of $\bar{G}_{v_i}(1; \mathbf{0})$ is $[\gamma_i \sigma^{ii}(\boldsymbol{\theta}_s)]^2$, where $\sigma^{ii}(\boldsymbol{\theta}_s)$ denotes the i^{th} diagonal element of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)$. But we can again use (6) to interpret $\gamma_i \sigma^{ii}(\theta_s)$ as the R^2 in the theoretical least squares projection of v_{it} on a constant and $\mathbf{y}_t$. Therefore, we can apply a similar line of reasoning to a Gaussian test of $H_0 : \rho_i^* = 0$ under the maintained assumption that both ρ and the remaining elements of $\boldsymbol{\rho}^*$ are 0. In this respect, note that $\sigma^{ii}(\boldsymbol{\theta}_s) = \gamma_i^{-1} - \gamma_i^{-2} c_i^2/[1 + (\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})]$ when $\boldsymbol{\Gamma}$ has full rank, which means that $\gamma_i \sigma^{ii}(\boldsymbol{\theta}_s)$ also converges to 1 as N increases for fixed c_i and γ_i.

Nevertheless, it is important to emphasise that our joint tests take into account the covariance between the Kalman filter estimators of common and specific factors, even though the latent variables themselves are uncorrelated. In fact, $\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta},\boldsymbol{\eta};\boldsymbol{\eta})$ has rank N instead of $N+1$ because of the negative relationship $\mathbf{v}_{kt}(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c} f_{kt}(\boldsymbol{\theta})$, which rules out the direct application of the multivariate serial correlation test discussed in section 3.3 to the vector process $[f_{kt}(\boldsymbol{\theta}_s), \mathbf{v}'_{kt}(\boldsymbol{\theta}_s)]'$.

Part 3 of Proposition 2 also implies that the asymptotic distribution of the Gaussian tests does not depend on normality or indeed ellipticity. Effectively, this result mimics the fact that under conditional homoskedasticity, standard score tests for serial correlation in observed series are also robust to non-normality. In fact, we can strengthen this intuition as follows. Since $V[f_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}] = \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$, we can obtain an asymptotically equivalent test of $H_0 : \rho = 0$ by computing the F test of the regression of $f_{kt}(\boldsymbol{\theta}_s)$ on a constant and $f_{kt-1}(\boldsymbol{\theta}_s)$, whose asymptotic null distribution does not depend on Gaussianity.

Finally, it is worth mentioning that the orthogonality conditions (16) and (17) remain valid when $\mathbf{y}_t$ is serially uncorrelated irrespective of $V(\mathbf{y}_t)$ having an exact single factor structure. Therefore, one could also use them to derive a standard moment test (see e.g. Newey and McFadden (1994), Newey (1985) and Tauchen (1985)), which will continue to have non-trivial power even though it will no longer be an LM test (see Sentana and Shah (1994) for an interpretation of $\boldsymbol{\theta}_s$ when $\boldsymbol{\Sigma}(\boldsymbol{\theta}_s)$ is misspecified). Naturally, the same applies to (18) and (19).

3.2 Extensions

3.2.1 Unobservable conditional means

The assumption that the distribution of $\mathbf{y}_t$ conditional on its past alone is elliptically symmetric but with a mean vector and covariance matrix given by the usual Gaussian Kalman filter recursions may be regarded as a way of constructing a convenient auxiliary model that coincides with the model of interest for $\boldsymbol{\rho}^\dagger = \mathbf{0}$ or $\boldsymbol{\eta} = \mathbf{0}$, but whose log-likelihood function and score we can obtain in closed form for every possible value of $\boldsymbol{\rho}^\dagger$ when $\boldsymbol{\eta} \neq \mathbf{0}$. In this regard, it is important to bear in mind that the fact that we can compute the true log-likelihood function of $\mathbf{y}_t$ under the null of $\boldsymbol{\rho}^\dagger = \mathbf{0}$ is not sufficient to compute its derivative with respect to $\boldsymbol{\rho}^\dagger$. Nevertheless, it is possible to use the EM principle to obtain this score. Remarkably, it turns out that the score of the potentially non-Gaussian state-space model (13) and the approximating model used in the previous section coincide under the null, even though the Kalman filter prediction equations do not provide the true conditional mean and covariance matrix under the alternative. As a result, the test statistics we have derived in Proposition 2 remain valid for model (13) too. The following proposition formalises our claim for the multivariate Student t, but we conjecture it applies to most other elliptical distributions:

Proposition 3 *Let* $\mathbf{s}_t(\boldsymbol{\phi}) = \partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ *denote the log-likelihood score of the conditionally elliptical model for* $\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi}$ *in section 3.1. Similarly, let* $\mathbf{q}_t(\boldsymbol{\phi}) = \partial p(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ *denote the exact log-likelihood score of model (13). If (14) is a (standardised) multivariate Student t with* $0 \le \eta < .5$ *then* $l_t(\boldsymbol{\phi}) = p(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi})$ *and* $\mathbf{s}_t(\boldsymbol{\phi}) = \mathbf{q}_t(\boldsymbol{\phi})$ *when evaluated at* $\boldsymbol{\rho}^\dagger = \mathbf{0}$.

In other words, the approximating model "smoothly embeds" (in the sense used by Gallant and Tauchen (1996) in their Theorem 2) the original model in those circumstances.

3.2.2 Moving average processes

Specification (13) assumes that common and specific factors follow AR(1) processes. However, recent macroeconomic applications of dynamic factor models have often considered moving average processes instead, sometimes treating the lagged latent variables as additional factors (see again Bai and Ng (2008)). Thus, we could alternatively assume that

$$\begin{aligned} x_t &= f_t + \varphi f_{t-1}, \\ \mathbf{u}_t &= \mathbf{v}_t + diag(\boldsymbol{\varphi}^*)\mathbf{v}_{t-1}. \end{aligned} \tag{20}$$

In this case the autocorrelation structure of $\mathbf{y}_t$ corresponds to a restricted VMA(1) process. Although the Kalman filter recursions for this dynamic model change, we can show that the scores corresponding to $\boldsymbol{\varphi}^\dagger = (\varphi, \boldsymbol{\varphi}^{*\prime})'$ evaluated at $\boldsymbol{\varphi}^\dagger = \mathbf{0}$ numerically coincide with the scores corresponding to $\boldsymbol{\rho}^\dagger$ in model (13) evaluated at $\boldsymbol{\rho}^\dagger = \mathbf{0}$. Hence, we can also interpret $LM_{AR(1)}(\boldsymbol{\eta})$ in Proposition 2 as the LM test of $H_0 : \boldsymbol{\varphi}^\dagger = \mathbf{0}$. This result mimics the well known fact that MA(1) and AR(1) processes provide locally equivalent alternatives in univariate tests for serial correlation (see e.g. Godfrey (1988)).

3.2.3 Higher order processes

Consider now the following alternative:

$$\begin{aligned} x_t &= \textstyle\sum_{l=1}^{h} \rho_l x_{t-l} + f_t, \\ u_{it} &= \textstyle\sum_{l=1}^{h_i^*} \rho_{il}^* u_{it-l} + v_{it}, \qquad (i = 1, \ldots, N). \end{aligned}$$

In view of the discussion in section 3.1, it is perhaps not surprising that the score test of $\rho_l = 0$ will be based on a modified version of (18) with $f_{kt-l}(\boldsymbol{\theta}_s)$ replacing $f_{kt-1}(\boldsymbol{\theta}_s)$, while the test of $\rho_{il}^* = 0$ will be based on the analogue version of (19). Given that $\mathbf{y}_t$ is *i.i.d.* under the null, it is not difficult to show that the joint test for higher order dynamics will be given by T times the sum of terms of the form

$$\left(\begin{array}{cc} \bar{G}_f(l;\boldsymbol{\eta}) & vecd'[\boldsymbol{\Gamma}^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(l;\boldsymbol{\eta})\boldsymbol{\Gamma}^{-1/2} \end{array} \right) \mathcal{I}^{-1}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\eta}) \left(\begin{array}{cc} \bar{G}_f(l;\boldsymbol{\eta}) & vecd'[\boldsymbol{\Gamma}^{-1/2}\bar{\mathbf{G}}_{\mathbf{v}}(l;\boldsymbol{\eta})\boldsymbol{\Gamma}^{-1/2} \end{array} \right)'.$$

As expected, these statistics are also LM tests against MA(h) structures in the factors. And if for some reason we wanted to test for different orders of serial correlation in different latent variables, then we should eliminate the irrelevant autocovariances from the above expression.

Similarly, we could be interested either in models in which the autoregressive structure of the latent variable follows some restricted distributed lag, or in panel data type structures in which $\rho^*_{il} = \rho^*_l \; \forall i, l$ to alleviate the incidental parameter problems for large N. In those cases, we can use the usual chain rule to obtain the relevant moment conditions and their asymptotic covariance matrix. For example, if we assume that $\rho_l = \rho \; \forall l$, the relevant orthogonality condition of the Gaussian tests will become

$$E\left[f_{kt}(\boldsymbol{\theta}_s)\sum\nolimits_{l=1}^{h} f_{kt-l}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}\right] = 0,$$

with $h \cdot [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2$ being the corresponding asymptotic variance. Interestingly, this expression is entirely analogous to the so-called Hodrick (1992) standard errors used in LM tests for long run return predictability in univariate regressions with overlapping observations.

3.3 The relative power of AR tests in multivariate contexts

Although we investigate the finite sample properties of our proposed tests of serial correlation in common and specific factors in section 6, it is illustrative to theoretically compare their power to other possibilities, such as the multivariate generalisation of the Box and Pierce (1970) test proposed by Hosking (1981),[7] a standard univariate AR(1) test applied to the Equally Weighted Portfolio (EWP), and a joint test of univariate first-order autocorrelation in all N series ($H_0 : vecd[\mathbf{G_y}(1)] = \mathbf{0}$), which takes into account that the $y'_{it}s$ are contemporaneously correlated even when they are serially uncorrelated.[8] We consider a single factor model of the form:

$$\begin{aligned} y_{it} &= \pi_i + c_i x_t + u_{it} \qquad (i = 1, \dots, 5) \\ x_t &= \rho x_{t-1} + \sqrt{1-\rho^2} f_t \\ u_{it} &= \rho^*_i u_{it-1} + \sqrt{1-\rho^{*2}_i} v_{it} \end{aligned}$$

where $\boldsymbol{\pi} = (.5, .4, .5, .4, .5)$, $\mathbf{c} = (5, 4, 5, 4, 5)$, $\boldsymbol{\gamma} = (5, 9, 5, 9, 5)$ and $\rho^*_i = \rho^* \; \forall i$. Such a design is motivated by the empirical application in section 7 and our desire to avoid exchangeable models, in which unusual simplifications occur. We evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\boldsymbol{\rho}^{\dagger}_{0T} = \bar{\boldsymbol{\rho}}^{\dagger}/\sqrt{T}$ (see appendix B for details). To avoid penalising Hosking's test, in this section we only consider the Gaussian versions of our tests. In any case, all the Gaussian tests that we compare will be robust to the presence of non-normality.

[7] In the first order case, one can reinterpret his proposal as a test of the null hypothesis of lack of serial correlation against an unrestricted VAR(1) model, as in Hendry (1971), Gulkey (1974) and Harvey (1982).

[8] Given the single factor structure of $\boldsymbol{\Sigma}$, this test differs from Test 2 in Harvey (1982), which tests the null hypothesis $H_0 : vecd(\mathbf{G}_y(1)) = \mathbf{0}$ under the maintained assumption that $\boldsymbol{\Sigma}$ is diagonal.

In view of the discussion following Proposition 2, it is worth looking at the first two unconditional moments of $\mathbf{y}_t$. In this sense, note that by construction $E(x_t) = 0$, $V(x_t) = 1$, $E(u_{it}) = 0$, $V(u_{it}) = \gamma_i$ and $cov(x_t, u_{it}) = 0$ both under the null and the different alternatives, which implies that $E(\mathbf{y}_t) = \boldsymbol{\pi}$ and $V(\mathbf{y}_t) = \mathbf{cc}' + \boldsymbol{\Gamma}$. Thus, the unconditional standard deviations will be $\sqrt{30}$ for the first, third and fifth series, and 5 for the second and fourth ones, while the unconditional correlations will be $.8\mathring{3}$ (odd with odd), .73 (odd with even) or .64 (even with even). Finally, the "signal to noise" ratio $\mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c}$, which coincides with the R^2 in the theoretical least squares projection of f_t on a constant and $\mathbf{y}_t$, is .95.[9] As for the means, note that we have implicitly imposed that linear factor pricing holds because $\boldsymbol{\pi} = .1\mathbf{c}$. Although this restriction is inconsequential for our econometric results, it implies an a priorily realistic unconditional mean-variance frontier, with a maximum Sharpe ratio of .34 on an annual basis.[10]

Figure 1a shows that when $\rho^* = 1.5\rho$ our proposed test of $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}$ is the most powerful at the usual 5% significance level, closely followed by the test of $H_0 : \boldsymbol{\rho}^* = \mathbf{0}$. Next, we find the pormentau test of $H_0 : vec[\mathbf{G}_\mathbf{y}(1)] = \mathbf{0}$ and the univariate test applied to EWP, which is barely distinguishable from the test of serial correlation in the common factor and very close to the "diagonal" serial correlation test of $H_0 : vecd[\mathbf{G}_\mathbf{y}(1)] = \mathbf{0}$. However, the results depend on the "signal to noise" ratio $\mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c}$. Figure 1b shows the equivalent picture when we multiply all the elements of $\boldsymbol{\gamma}$ by 10, so that the R^2 in the regression of f_t on $\mathbf{y}_t$ reduces to .65. In this case, the power of our test of serial correlation in f_t, the univariate test on EWP and especially the diagonal test increases substantially. In contrast, Figure 1c illustrates the effects of dividing the elements of $\boldsymbol{\gamma}$ by 5, so that the aforementioned R^2 reaches .99. In this context, the diagonal test becomes the least powerful.

The other crucial determinant of the power of the different tests is the relative magnitudes of ρ and ρ^*. Figure 2a shows the effect of setting $\rho^* = 0$ for our baseline signal to noise ratio, while Figure 2b illustrates the effects of $\rho = 0$. In the first case, the test of serial correlation in the common factor becomes the most powerful, with the test of serial correlation in the specific factors having power virtually equal to size, while exactly the opposite happens in the second case.[11]

[9] A more common measure of the importance of commonalities is the R^2 in the theoretical regression of each series on the common factor, which is $.8\mathring{3}$ for the odd numbered series and .64 for the even numbered ones.

[10] The ex-ante optimal mean-variance portfolio % weights are (25.7,11.4,25.7,11.4,25.7).

[11] Although the test of $H_0 : \rho = 0$ has non-trivial local power when $\rho = 0$ but $\boldsymbol{\rho}^* \neq \mathbf{0}$, and the same is true of the test of $H_0 : \boldsymbol{\rho}^* = \mathbf{0}$ when $\boldsymbol{\rho}^* = \mathbf{0}$ but $\rho \neq 0$, a much larger horizontal axis would be necessary to appreciate those effects.

3.4 The relative power of the normality tests

Let us now assess the gains that accrue from exploiting the non-normality in the distribution of returns. It is not difficult to show that the ratio of non-centrality parameters of the normality test $LM_{AR(1)}(\mathbf{0})$ and the elliptical likelihood test $LM_{AR(1)}(\boldsymbol{\eta})$ is $\mathrm{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$, which reflects the fact that the non-centrality parameter of the Gaussian tests is invariant to the true conditional distribution of the data. In the multivariate Student t case with $\nu_0 > 4$, in particular, this asymptotic efficiency ratio becomes

$$\frac{(\nu_0 - 2)(\nu_0 + N + 2)}{\nu_0(\nu_0 + N)}. \tag{21}$$

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 $\nu_0 \to 2^+$. For instance, for $N = 1$ it takes the values of $.9\mathring{3}$ and $.8$ for $\nu_0 = 9$ and $\nu_0 = 5$, respectively. At the same time, this ratio is decreasing in N for a given ν_0, which reflects the fact that Fisher's information for the mean is "increasing" in N in the Student t case (see Fiorentini and Sentana (2010a)). For $N = 3$ and $\nu_0 = 9$, for instance, it takes the value of .907, while for $\nu_0 = 5$, its value is only .75.

It is also straightforward to map those efficiency ratios into power gains by considering sequences of local alternatives. For illustrative purposes, we look at the baseline design in section 3.3 under the assumption that the true conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is a multivariate t_6. Figure 2c shows that the power gains that accrue to our proposed serial correlation tests by exploiting the leptokurtosis of the Student t distribution are far from trivial.

4 Tests for ARCH effects in common and idiosyncratic factors

4.1 Baseline case

The alternative that we consider next is the following conditionally heteroskedastic factor model:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c} f_t + \mathbf{v}_t \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\eta} \sim s\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}, \boldsymbol{\eta}\right] \end{array}\right\}, \tag{22}$$

with

$$\left.\begin{array}{c} \lambda_t(\boldsymbol{\theta}) = 1 + \alpha[E(f_{t-1}^2|\mathbf{Y}_{t-1}; \boldsymbol{\theta}, \mathbf{0}) - 1], \\ \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha_i^*[E(v_{it-1}^2|\mathbf{Y}_{t-1}; \boldsymbol{\theta}, \mathbf{0}) - \gamma_i], \qquad (i = 1, \ldots, N) \end{array}\right\} \tag{23}$$

where $E(f_{t-1}^2|\mathbf{Y}_{t-1}; \boldsymbol{\theta}, \mathbf{0})$ and $E(v_{it-1}^2|\mathbf{Y}_{t-1}; \boldsymbol{\theta}, \mathbf{0})$ are the Kalman filter estimators of the squares of the underlying common and idiosyncratic factors obtained from this model (see appendix B in Fiorentini and Sentana (2012) for details). In this case, the parameters of interest become $\boldsymbol{\phi} = (\boldsymbol{\theta}_s', \boldsymbol{\eta}')'$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_s', \boldsymbol{\alpha}^{\dagger\prime})'$, where $\boldsymbol{\alpha}^\dagger = (\alpha, \boldsymbol{\alpha}^*)$ and $\boldsymbol{\alpha}^* = (\alpha_1, \ldots, \alpha_N)$. Although it is in principle very important to distinguish between $I_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{f}_{t-1},\ \mathbf{y}_{t-2}, \mathbf{f}_{t-2}, \ldots\}$, and the econometrician's information set $\mathbf{Y}_{t-1}$, which only includes lagged values of $\mathbf{y}_t$, (see Harvey, Ruiz and Sentana (1992)), for ease of exposition we postpone the discussion of those cases in which $\lambda_t(\boldsymbol{\theta}) \notin \mathbf{Y}_{t-1}$ until section 4.2.1.

Given (22) and (23), the distribution of $\mathbf{y}_t$ conditional on $\mathbf{Y}_{t-1}$ is $N(\mathbf{0}, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\Sigma}_t = \mathbf{cc}'\lambda_t + \boldsymbol{\Gamma}_t$ has the usual exact factor structure. For this reason, we shall refer to the data generation process specified by (22) as a multivariate conditionally heteroskedastic exact factor model, which reduces to our baseline specification (1) under the null hypothesis that $H_0 : \boldsymbol{\alpha}^\dagger = \mathbf{0}$. But even if f_t or $\mathbf{v}_t$ are conditionally heteroskedastic, provided that they are covariance stationary, model (22) also implies an unconditional exact factor structure for $\mathbf{y}_t$. That is, the unconditional covariance matrix, $\boldsymbol{\Sigma}$, can be written as:

$$\boldsymbol{\Sigma} = E(\boldsymbol{\Sigma}_t|\boldsymbol{\theta}) = \mathbf{cc}' + \boldsymbol{\Gamma}, \tag{24}$$

because we have set the unconditional variance of the common factor to 1 to eliminate the usual scale indeterminacy.[12]

The above model has very interesting implications for correlations. A stylised fact that has been noted before is that periods when markets are increasingly correlated are also times when markets are volatile (see King, Sentana and Wadhwani (1994)). Since the empirical evidence typically suggests that changes in the unobservable factor lead to individual stocks moving in the same direction, model (22) implies that periods when the volatility of the unobservable factor rises are also those when, ceteris paribus, individual stocks appear to exhibit greater inter-correlation. Specifically, the conditional correlation coefficient between any two elements of $\mathbf{y}_t$ is given by

$$\rho_{12t} = \frac{c_1 c_2 \lambda_t}{\sqrt{c_1^2\lambda_t + \gamma_{1t}}\sqrt{c_2^2\lambda_t + \gamma_{2t}}}.$$

Hence, ρ_{12t} will be increasing in λ_t if $c_1 c_2 > 0$ and decreasing in γ_{1t} and γ_{2t}.

A more precise way to characterise the serial dependence structure implied by model (22) is to consider the autocovariance structure of

$$vec[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'] = (\mathbf{c} \otimes \mathbf{c})f_t^2 + vec(\mathbf{v}_t\mathbf{v}_t') + (\mathbf{I}_{N_2} + \mathbf{K}_{NN})(\mathbf{c} \otimes \mathbf{I}_N)vec(f_t\mathbf{v}_t),$$

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see Magnus and Neudecker (1988)). Given that $vec(f_t\mathbf{v}_t)$ is a martingale difference sequence, $\mathbf{y}_t$ follows a weak ARCH model (see Nijman and Sentana (1996)) which shares the factor structure in (15) not for the levels but for the squares and cross-products of the observed variables $\mathbf{y}_t$ (see appendix B for further details).

[12] See Fiorentini, Sentana and Shephard (2004) for symmetric scaling assumptions for integrated ARCH models.

In this sense, another empirically appealing feature of (22) is that all linear combinations of $\mathbf{y}_t$ will follow weak ARCH processes as long as α and $\boldsymbol{\alpha}^*$ are strictly positive.

Sentana and Fiorentini (2001) develop tests of the null hypothesis $H_0 : \alpha = 0$ under the maintained hypotheses that $\boldsymbol{\alpha}^* = \mathbf{0}$ and the conditional distribution is Gaussian. The following proposition extends their results to joint tests of ARCH effects in common and specific factors in elliptical contexts.

Proposition 4 *Let*

$$\bar{S}_f(j;\boldsymbol{\eta}) = \frac{1}{T}\sum\nolimits_{t=1}^{T}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\}[f_{kt-j}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]$$

denote the sample cross moment of $\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1$ *and* $E(f_{t-j}^2|\mathbf{Y}_{t-j};\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}) = f_{kt-j}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1$*, where* $\delta(\varsigma_t,\boldsymbol{\eta})$ *is defined in (11) and* $f_{kt}(\boldsymbol{\theta}_s)$ *and* $\omega_k(\boldsymbol{\theta}_s)$ *are obtained from the updating equations (5) of the static factor model (1). Similarly, let*

$$\begin{aligned}\bar{\mathbf{S}}_{\mathbf{v}}(j;\boldsymbol{\eta}) &= \frac{1}{T}\sum\nolimits_{t=1}^{T} vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}\}\\ &\quad\times vecd[\mathbf{v}_{kt-j}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-j}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\end{aligned}$$

denote the analogous sample cross moments for the specific factors.

1. *Under the null hypothesis* $H_0 : \boldsymbol{\alpha}^\dagger = \mathbf{0}$*, the score test statistic* $LM_{ARCH(1)}(\boldsymbol{\eta})$ *given by*

$$\tfrac{T}{4}\left(\bar{S}_f(1;\boldsymbol{\eta}_0), vecd'[\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(1;\boldsymbol{\eta})\boldsymbol{\Gamma}_0^{-1}]\right)\mathcal{I}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}^{-1}(\boldsymbol{\theta}_{s0},\mathbf{0};\boldsymbol{\eta}_0)\left(\bar{S}_f(1;\boldsymbol{\eta}_0), vecd'[\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(1;\boldsymbol{\eta}_0)\boldsymbol{\Gamma}_0^{-1}]\right)',$$

is distributed as a χ^2 *with* $N+1$ *degrees of freedom for* N *fixed as* T *goes to infinity, where*

$$\mathcal{I}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) = \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})\odot\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}),$$

$$\begin{aligned}\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) &= V\begin{bmatrix}\frac{1}{\sqrt{2}}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\}\\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}\}\end{bmatrix}\\ &= \mathrm{M}_{ss}(\boldsymbol{\eta})\begin{bmatrix}[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\end{bmatrix}\\ &\quad+\frac{[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]}{2}\begin{bmatrix}[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}]\\ [\mathbf{c}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] & vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}]\end{bmatrix}.\end{aligned}$$

and $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$ *mimics* $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta})$ *after replacing* $\mathrm{M}_{ss}(\boldsymbol{\eta})$ *by* $\kappa+1$*. Moreover, this asymptotic null distribution is unaffected if we replace* $\boldsymbol{\theta}_{s0}$ *and* $\boldsymbol{\eta}_0$ *by their feasible maximum likelihood estimators in Proposition 1.*

2. *It also remains valid if we replace* $\boldsymbol{\theta}_{s0}$ *by its elliptically symmetric semiparametric estimator in Proposition 1, which requires the nonparametric estimation of the density of* $\varsigma_t(\boldsymbol{\theta}_s)$.

3. *Under the same null hypothesis, the Gaussian pseudo score test statistic* $LM_{ARCH(1)}(\mathbf{0})$ *given by*

$$\tfrac{T}{4}\left(\bar{S}_f(1;\mathbf{0}), vecd'[\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(1;\mathbf{0})\boldsymbol{\Gamma}_0^{-1}]\right)\mathcal{B}^{-1}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\phi}_0)\left(\bar{S}_f(1;\mathbf{0}), vecd'[\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(1;\mathbf{0})\boldsymbol{\Gamma}_0^{-1}]\right)',$$

with

$$\mathcal{B}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\phi}) = \mathcal{V}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) \odot \mathcal{V}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}), \tag{25}$$

will be distributed as a χ^2 *with* $N+1$ *degrees of freedom for* N *fixed as* T *goes to infinity irrespective of whether the elliptical conditional distribution is normal. This result continues to hold if we replace* $\boldsymbol{\theta}_{s0}$ *by its Gaussian pseudo maximum likelihood estimator* $\bar{\boldsymbol{\theta}}_s$ *in Proposition 1.*

Researchers may once more be interested in tests that separately assess the conditional heteroskedasticity of either the common factor or the specific factors. Indeed, they might even like to focus on a particular v_{it}. By combining the relevant elements of $\bar{S}_f(j;\boldsymbol{\eta})$ and $\bar{\mathbf{S}}_{\mathbf{v}}(1;\boldsymbol{\eta})$ with the corresponding blocks of the information matrix, $\mathcal{I}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$, we can easily exploit the results in Proposition 4 to derive the required test statistics for those subcomponents under the maintained hypothesis of serial independence. Intuitively, the reason is that we can interpret $LM_{ARCH(1)}(\boldsymbol{\eta})$ as a test based on the $N+1$ orthogonality conditions:

$$E\left\{\begin{array}{c}\frac{1}{2}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\}\\ \cdot[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\end{array}\right\} = 0, \tag{26}$$

$$E\left\{\begin{array}{c}\frac{1}{2}\gamma_i^{-2}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]v_{kit}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i\}\\ \cdot[v_{kit-1}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\end{array}\right\} = 0 \qquad (i=1,\ldots,N). \tag{27}$$

Similarly, $LM_{ARCH(1)}(\mathbf{0})$ is based on

$$E\left\{\begin{array}{c}\frac{1}{2}[f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]\\ \cdot[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\end{array}\right\} = 0, \tag{28}$$

$$E\left\{\begin{array}{c}\frac{1}{2}\gamma_i^{-2}[v_{kit}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]\\ \cdot[v_{kit-1}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}\end{array}\right\} = 0 \qquad (i=1,\ldots,N). \tag{29}$$

As in the serial correlation tests, $\delta(\varsigma_t,\boldsymbol{\eta})$ acts as a damping factor for big observations (see Fiorentini and Sentana (2010b) for a closely related discussion for univariate models).[13]

[13] This factor also plays an important role in the beta-t-ARCH models proposed by Harvey and Chakravarty (2008), although if one derived an LM test for conditional homoskedasticity against their models, $\delta(\varsigma_t,\boldsymbol{\eta})$ would appear not only in the regressand but also in the regressor.

Once again, given that we normalise $V(f_t)$ to 1, the moment conditions (28) and (29) closely resemble

$$\begin{aligned} E[(f_t^2-1)(f_{t-1}^2-1)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}] &= 0, \\ E[\gamma_i^{-2}(v_{it}^2-\gamma_i)(v_{it-1}^2-\gamma_i)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}] &= 0 \qquad (i=1,\ldots,N), \end{aligned}$$

which are the Gaussian-based orthogonality conditions that we would use to test for first order ARCH effects if we could observe the latent variables (see e.g. Engle (1982)).

The similarity between these two sets of moment conditions becomes even stronger if we consider individual tests for ARCH in each latent variable. Let us start with a test of H_0 : $\alpha = 0$ under the maintained assumption that $\boldsymbol{\alpha}^* = \mathbf{0}$. Part 3 of Proposition 4 implies that the asymptotic variance of $\bar{S}_f(1;\mathbf{0})$ is simply $\frac{1}{2}(3\kappa+2)^2[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^4$. But as we saw in section 3.1, we can interpret $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$ as the R^2 in the theoretical least squares projection of f_t on a constant and $\mathbf{y}_t$. Therefore, the higher the degree of observability of the common factor, the closer the asymptotic variance of $\bar{S}_f(1;\mathbf{0})$ will be to $\frac{1}{2}(3\kappa+2)^2$, which is the asymptotic variance of the first sample autocovariance of f_t^2 under normality. Intuitively, this convergence result simply reflects the fact that the common factor becomes observable in the limit, which implies that our test of $H_0 : \alpha = 0$ will become arbitrarily close to a first order ARCH test for the common factor as the "signal to noise" ratio $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$ approaches 1. Before the limit, though, our test takes into account the unobservability of f_t.

Likewise, part 3 of Proposition 4 implies that the asymptotic variance of $\bar{S}_{v_{ki}v_{ki}}(1,\mathbf{0})$ is $\frac{1}{2}(3\kappa+2)^2[\gamma_i\sigma^{ii}(\boldsymbol{\theta}_s)]^4$, where $\sigma^{ii}(\boldsymbol{\theta}_s)$ denotes the i^{th} diagonal element of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)$. But since we can again interpret $\gamma_i\sigma^{ii}(\theta_s)$ as the R^2 in the theoretical least squares projection of v_{it} on a constant and $\mathbf{y}_t$, we can apply a similar line of reasoning to a test of $H_0 : \alpha_i^* = 0$ under the maintained assumption that $\alpha = 0$ and the remaining elements of $\boldsymbol{\alpha}^*$ are 0. Once again, though, it is important to emphasise that our joint tests take into account the covariance between the Kalman filter estimators of the underlying factors, even though the latent variables themselves are uncorrelated.

Part 3 of Proposition 4 also implies that the asymptotic distribution of the Gaussian tests does not depend on normality, although if the conditional distribution of $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ were not elliptical, then one would have to replace $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$ in (25) by the joint unconditional covariance matrix of $\frac{1}{\sqrt{2}}[f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]$ and $\frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}'(\boldsymbol{\theta}_s)+\mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]$ under the null of $H_0 : \boldsymbol{\alpha}^\dagger = \mathbf{0}$. The advantage of using the theoretical expressions in the elliptical case is that they should improve the finite sample reliability of the Gaussian tests.

Interestingly, such robust versions of the test for ARCH effects in common and idiosyncratic factors can be regarded as the factor analytic analogues to the suggestion that Koenker (1981) made to robustify tests of conditional homoskedasticity based on Gaussian scores, such as the original univariate ARCH test in Engle (1982), whose information matrix version is only valid under conditional normality. In fact, we can obtain an asymptotically equivalent test of H_0 : $\alpha = 0$ by computing the F test of the regression of $f_{kt}^2(\boldsymbol{\theta}_s)$ on a constant and $f_{kt-1}^2(\boldsymbol{\theta}_s)$, whose asymptotic null distribution remains valid irrespective of the normality of $f_{kt}(\boldsymbol{\theta}_s)$ because it is effectively using a consistent estimator of $V[f_{kt}^2(\boldsymbol{\theta}_s)]$ as the residual variance of the regression under the null. But if we impose that the residual variance is $2[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2$ instead, which is its value under normality because $V[f_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}] = \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}$, then our F test will be incorrectly sized when the conditional distribution is not Gaussian.

Again, it would be straightforward to adapt Proposition 4 to handle large N panel data restrictions such as $\alpha_i^* = \alpha^*\ \forall i$, as in Sentana, Calzolari and Fiorentini (2008). Further, given that the orthogonality conditions (26) and (27) remain valid when $\mathbf{y}_t$ is serially independent irrespective of $V(\mathbf{y}_t)$ having an exact single factor structure, one could also use them to derive a standard moment test that will still have non-trivial power even though it will no longer be an LM test.

4.2 Extensions

4.2.1 Unobservable conditional variances

Following the discussion at the beginning of section 5 in Harvey, Ruiz and Sentana (1992), specification (22) assumes that the conditional variances of common and specific factors are a function of lagged observable variables. But it may seem more natural to assume that those variances are in fact functions of the lagged latent variables. Specifically,

$$\left.\begin{array}{r} \lambda_t(\boldsymbol{\theta}) = 1 + \alpha(f_{t-j}^2 - 1), \\ \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha_i^*(v_{it-1}^2 - \gamma_i), \qquad (i = 1, \ldots, N). \end{array}\right\} \tag{30}$$

The problem with this formulation is that even in the Gaussian case the log-likelihood function can no longer be written in closed form except when $\boldsymbol{\alpha}^\dagger = \mathbf{0}$, and one has to resort to simulation methods, such as the MCMC procedures put forward by Fiorentini, Sentana and Shephard (2004). As explained by Sentana, Calzolari and Fiorentini (2008), the combination of (22) with (23) may be regarded as a convenient auxiliary model that coincides with the model of interest for $\boldsymbol{\alpha}^\dagger = \mathbf{0}$, but whose log-likelihood function and score we can obtain in closed form for every possible value of $\boldsymbol{\alpha}^\dagger$. In this regard, it is important to bear in mind that the fact that we can compute the true log-likelihood function of $\mathbf{y}_t$ under the null of $\boldsymbol{\alpha}^\dagger = \mathbf{0}$ is not sufficient to compute its derivative with respect to $\boldsymbol{\alpha}^\dagger$. Fortunately, it is once again possible to use the EM principle to obtain this score. Remarkably, it turns out that the score of the model with latent variances (30) is virtually identical to the score of the approximating model under the null of conditionally homoskedasticity despite both the non-measurability of λ_t and $\mathbf{\Gamma}_t$ and the potential non-normality of the conditional distribution. In fact, they would coincide if we had followed section 5.2 of Harvey, Ruiz and Sentana (1992) instead, and replaced the conditional variances of common and specific factors in (23) by

$$\left.\begin{array}{r}\lambda_t(\boldsymbol{\theta}) = 1 + \alpha[E(f_{t-1}^2|\mathbf{Y}_{t-1};\boldsymbol{\theta},\boldsymbol{\eta}) - 1], \\ \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha_i^*[E(v_{it-1}^2|\mathbf{Y}_{t-1};\boldsymbol{\theta},\boldsymbol{\eta}) - \gamma_i], \qquad (i=1,\ldots,N)\end{array}\right\} \tag{31}$$

where

$$\begin{aligned} E(f_{t-1}^2|\mathbf{Y}_{t-1};\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) &= f_{kt-1}^2(\boldsymbol{\theta}_s) + \upsilon_{kt-1}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) - 1, \\ E(v_{it-1}^2|\mathbf{Y}_{t-1};\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) &= v_{ikt-1}^2(\boldsymbol{\theta}_s) + c_i^2\upsilon_{kt-1}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) - \gamma_i, \end{aligned}$$

with $f_{kt}(\boldsymbol{\theta}_s)$, $\omega_k(\boldsymbol{\theta}_s)$ and $\upsilon_{kt-1}(\boldsymbol{\theta}_s,\boldsymbol{\eta})$ defined in (5). The following result, which generalises Proposition 1 in Sentana, Calzolari and Fiorentini (2008), formalises our claim for the multivariate Student t, but we conjecture it applies to most other elliptical distributions:

Proposition 5 *Let* $\mathbf{s}_t(\boldsymbol{\phi}) = \partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ *denote the log-likelihood score of the conditionally heteroskedastic model for* $\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi}$ *in (22) when the variances of the latent factors are given by (31). Similarly, let* $\mathbf{q}_t(\boldsymbol{\phi}) = \partial p(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ *denote the exact log-likelihood score of the same model when the variances of the latent factors are given by (30) instead. If the conditional distribution is a (standardised) multivariate Student t with* $0 \leq \eta < .5$ *then* $l_t(\boldsymbol{\phi}) = p(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\phi})$ *and* $\mathbf{s}_t(\boldsymbol{\phi}) = \mathbf{q}_t(\boldsymbol{\phi})$ *when evaluated at* $\boldsymbol{\alpha}^\dagger = \mathbf{0}$.

Therefore, the approximating model that uses (31) "smoothly embeds" the original model in those circumstances.

4.2.2 Higher order processes

Consider the following alternative specification:

$$\begin{aligned} \lambda_t(\boldsymbol{\theta}) &= 1 + \sum\nolimits_{j=1}^{q} \alpha_j[E(f_{t-j}^2|Y_{t-j};\boldsymbol{\theta},\mathbf{0}) - 1], \\ \gamma_{it}(\boldsymbol{\theta}) &= \gamma_i + \sum\nolimits_{j=1}^{q_i^*} \alpha_{ij}^*[E(v_{it-j}^2|Y_{t-j};\boldsymbol{\theta},\mathbf{0}) - \gamma_i]. \qquad (i=1,\ldots,N), \end{aligned}$$

In view of the discussion in section 4.1, it is perhaps not surprising that the score test of $\alpha_j = 0$ will be based on a modified version of (28) with $f_{kt-j}^2(\boldsymbol{\theta}_s)$ replacing $f_{kt-1}^2(\boldsymbol{\theta}_s)$, while the test of $\alpha_{ij}^* = 0$ will be based on the analogue version of (29). Given that $\mathbf{y}_t$ is *i.i.d.* under the null hypothesis, it is not difficult to show that the joint test for higher order dynamics will be given by $\frac{1}{4}T$ times the sum of terms of the form

$$\left(\bar{S}_f(j;\boldsymbol{\eta}), vecd'[\mathbf{\Gamma}^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(j;\boldsymbol{\eta})\mathbf{\Gamma}^{-1}]\right)\mathcal{I}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}^{-1}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})\left(\bar{S}_f(j;\boldsymbol{\eta}), vecd'[\mathbf{\Gamma}^{-1}\bar{\mathbf{S}}_{\mathbf{v}}(j;\boldsymbol{\eta})\mathbf{\Gamma}^{-1}]\right)'.$$

Once again, we could eliminate the irrelevant autocovariances from the above expression to test for different orders of serial correlation in the squares of different latent variables.

The univariate empirical evidence, though, suggests that $\mathrm{G{\scriptsize ARCH}}(1,1)$ specifications of the form

$$\begin{aligned}
\lambda_t(\boldsymbol{\theta}) &= 1-\alpha-\beta+\alpha E(f_{t-j}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})+\beta\lambda_{t-1}(\boldsymbol{\theta}) \\
&= 1+\alpha\textstyle\sum_{j=1}^{\infty}\beta^{j-1}[E(f_{t-j}^2|Y_{t-j};\boldsymbol{\theta},\mathbf{0})-1], \\
\gamma_{it}(\boldsymbol{\theta}) &= \gamma_i(1-\alpha_i^*-\beta_i^*)+\alpha_i^*E(v_{it-j}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})+\beta_i^*\gamma_{it-1}(\boldsymbol{\theta}) \\
&= \gamma_i+\alpha_i^*\textstyle\sum_{j=1}^{\infty}(\beta_i^*)^{j-1}[E(v_{it-j}^2|Y_{t-j};\boldsymbol{\theta},\mathbf{0})-\gamma_i]
\end{aligned}$$

should be more realistic than unrestricted $\mathrm{A{\scriptsize RCH}}(q)$ ones. As Bollerslev (1986) noted in a univariate context, however, one cannot derive a score test for conditional homoskedasticity versus these $\mathrm{G{\scriptsize ARCH}}(1,1)$ specifications in the usual way, because β and β_i^* are only identified under the alternative. A possible solution to testing situations such as this one involves computing the test statistic for many values of β and β_i^* in the range [0,1), which are then combined to construct an overall statistic, as initially suggested by Davies (1977, 1987). Andrews (2001) discusses ways of obtaining critical values for such tests by regarding the different LM statistics as continuous stochastic processes indexed with respect to the parameters β and β_i^* $(i=1,\ldots,N)$. Unfortunately, his procedure is difficult to apply in our context because $\dim(\boldsymbol{\beta}^{\dagger})=N+1$. An alternative solution involves choosing arbitrary values of the underidentified parameters to carry out a score test of $\alpha=0$ and $\boldsymbol{\alpha}^*=\mathbf{0}$ based on the moment conditions

$$E\left\{[\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]\sum_{l=1}^{\infty}\beta^{l-1}[f_{kt-l}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}\right\}=0,$$

$$E\left\{[\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]v_{kit}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]\sum_{l=1}^{\infty}(\beta_i^*)^{l-1}[v_{kit-l}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}\right\}=0,$$

whose asymptotic covariance matrix would be

$$\sum_{l=0}^{\infty}diag^l[\beta,\boldsymbol{\beta}^{*\prime}]\mathcal{I}_{\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}^{\dagger}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})diag^l[\beta,\boldsymbol{\beta}^{*\prime}],$$

which can be obtained in closed form. The values of β and $\boldsymbol{\beta}^*$ influence the small sample power of these tests, achieving maximum power when they coincide with their true values (see Demos and Sentana (1998)), but the advantage is that the resulting tests have standard distributions under H_0. An attractive possibility is to set β and $\boldsymbol{\beta}^*$ to the decay factor recommended by RiskMetrics (1996) to obtain exponentially weighted volatility estimates for f_{kt} and v_{ikt}.

4.3 The relative power of ARCH tests in multivariate contexts

We compare the power of our LM tests, Hosking's test applied to $vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})']$ as in Duchesne and Lalancette (2003), a standard univariate $\mathrm{A{\scriptsize RCH}}(1)$ test applied to the EW portfolio, a joint test of univariate first-order autocorrelation in all $N(N+1)/2$ squares and cross-products of the (demeaned) observed series, and an analogous test that only focuses on their squares. Note that our joint LM test can also be understood as test of univariate first-order autocorrelation in the squares of $[f_{kt}(\boldsymbol{\theta}_s), \mathbf{v}'_{kt}(\boldsymbol{\theta}_s)]$.[14] We consider another non-exchangeable single factor model of the form:

$$\begin{aligned} y_{it} &= \pi_i + c_i f_t + v_{it} \qquad (i = 1, \ldots, 5) \\ \lambda_t &= (1-\alpha) + \alpha f_{t-1}^2 \\ \gamma_{it} &= \gamma_i(1-\alpha_i^*) + \alpha_i^* v_{it-1}^2 \end{aligned}$$

where $\boldsymbol{\pi} = (.5, .4, .5, .4, .5)$, $\mathbf{c} = (5, 4, 5, 4, 5)$, $\boldsymbol{\gamma} \propto (5, 9, 5, 9, 5)$ and $\alpha_i^* = \alpha^* \ \forall i$, whose first two unconditional moments are also empirically motivated, as they coincide with those of the model considered in section 3.3. We evaluate power against *compatible* sequences of local alternatives of the form $\boldsymbol{\alpha}_{0T}^\dagger = \bar{\boldsymbol{\alpha}}^\dagger/\sqrt{T}$ (see appendix B for details). To avoid penalising Hosking's test, in this section we only consider the Gaussian versions of our tests.[15]

For the baseline case in which $\boldsymbol{\gamma} = (5, 9, 5, 9, 5)$, and $\alpha^* = \alpha$, Figure 3a shows that our proposed test of $H_0 : \boldsymbol{\alpha}^\dagger = \mathbf{0}$ is the most powerful at the usual 5% significance level, followed by our test of $H_0 : \boldsymbol{\alpha}^* = \mathbf{0}$. Next we find our test of ARCH effects in the common factor and the univariate ARCH test applied to EWP, the diagonal serial correlation tests of $vecd[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})']$ and $vech[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})']$, and finally the pormanteau test of unrestricted first-order serial dependence, which suffers from having a very large number of degrees of freedom. Once again, though, this ranking crucially depends on the "signal to noise" ratio $\mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c}$. Figure 3b shows the equivalent picture when we multiply all the elements of $\boldsymbol{\gamma}$ by 10, so that the R^2 in the regression of f_t on $\mathbf{y}_t$ reduces to .65. In this case, the power of the two univariate tests decreases substantially, while the power of the diagonal tests increases. In contrast, Figure 3c illustrates the effects of dividing the elements of $\boldsymbol{\gamma}$ by 5, so that the aforementioned R^2 reaches .99. Not surprisingly, the power of the two univariate tests almost coincides because EWP and $f_{kt}(\boldsymbol{\theta}_0)$ become very highly correlated.

The other crucial determinant of the power of the different tests is the relative magnitudes of α and α^*. Figure 4a shows the effect of setting $\alpha^* = 0$ for our baseline signal to noise ratio, while Figure 4b illustrates the effects of $\alpha = 0$. In the first case, the test of conditional homoskedasticity in the common factor becomes the most powerful, with the specific factors test having power virtually equal to size, while exactly the opposite happens in the second case.

[14] Another implication of the single factor structure of $\boldsymbol{\Sigma}$ is that our proposed LM test differs from the multivariate ARCH test considered by Dufour, Khalaf and Beaulieu (2010), who apply Hosking's test to the vech of the outer product of standardised values of $\mathbf{y}_t$ obtained from a Cholesky decomposition of $\bar{\boldsymbol{\Sigma}}$.

[15] See footnote 12 in Fiorentini and Sentana (2012) for ways of making Hosking's tests for squares and cross-products robust to non-normality.

4.4 The relative power of the normality tests

To keep the algebra simple, we shall initially compare the individual tests of $H_0 : \alpha = 0$ under the maintained assumption that all the remaining ARCH parameters in $\boldsymbol{\alpha}^*$ are 0. In this context, we can show that the ratio of non-centrality parameters of the Gaussian test and the elliptical test is $4/\{[3\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1](3\kappa_0 + 2)\}$. In the multivariate Student t case with $\nu_0 > 4$, in particular, this asymptotic efficiency ratio becomes

$$\frac{(\nu_0 + N + 2)\,(\nu_0 - 4)}{(\nu_0 - 1)(\nu_0 + N - 1)}.$$

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 $\nu_0 \to 4^+$. For instance, for $N = 1$, it takes the values of $.8\mathring{3}$ and .4 for $\nu_0 = 9$ and $\nu_0 = 5$, respectively. At the same time, this ratio is decreasing in N for a given ν_0. For $N = 3$ and $\nu_0 = 9$, for instance, it takes the value of .795, while for $\nu_0 = 5$, its value is only .75. Exactly the same results apply to tests of $H_0 : \alpha_i^* = 0$.

More generally, we can combine the asymptotic distribution of the different estimators of $\boldsymbol{\alpha}^\dagger$ under the null derived in Proposition 4 with the expressions in appendix B to obtain the non-centrality parameters of joint tests of $\boldsymbol{\alpha}^* = \mathbf{0}$ or $\boldsymbol{\alpha}^\dagger = \mathbf{0}$. Unlike in the case of the mean parameters, though, the asymptotic relative efficiency of the different tests depends on the values of the static factor analysis parameters $\boldsymbol{\theta}_s$. In any case, it is straightforward to map those efficiency ratios into power gains by considering sequences of local alternatives. For illustrative purposes, we look at the baseline design in section 4.3 under the assumption that the true conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is a multivariate t_6. Figure 4c shows that the power gains are even bigger for our proposed ARCH tests, which is in line with the asymptotic relative efficiency results derived above.

5 Joint tests for serial dependence

In this section we shall consider joint tests of AR(1)-ARCH(1) effects in common and specific factors. Therefore, our alternative will be a single factor version of a dynamic, conditionally heteroskedastic exact factor model in which both common and idiosyncratic factors follow covariance stationary AR(1)-ARCH(1) type processes. Specifically,

$$\left.\begin{array}{c}\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t \\ x_t = \rho x_{t-l} + f_t \\ \mathbf{u}_t = diag(\boldsymbol{\rho}^*)\mathbf{u}_{t-1} + \mathbf{v}_t \\ \left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\right) |I_{t-1};\boldsymbol{\theta} \sim s\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{array}\right), \boldsymbol{\eta}\right], \\ V(f_t|I_{t-1};\boldsymbol{\theta}) = \lambda_t(\boldsymbol{\theta}) = 1 + \alpha[E(f_{t-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0}) - 1], \\ V(v_{it}|I_{t-1};\boldsymbol{\theta}) = \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha_i^*[E(v_{it-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0}) - \gamma_i], \qquad (i = 1, \ldots, N) \end{array}\right\}. \tag{32}$$

When the conditional variances of the common and idiosyncratic factors are constant (i.e., $\alpha = \mathbf{0}$ and $\boldsymbol{\alpha}^* = \mathbf{0}$), the above formulation reduces to (13). Similarly, when the levels of the latent variables are unpredictable (i.e., $\rho = \mathbf{0}$ and $\boldsymbol{\rho}^* = \mathbf{0}$), the above model simplifies to (22). Finally, under the null hypothesis of lack of predictability in mean ($\boldsymbol{\rho}^\dagger = \mathbf{0}$) and variance ($\boldsymbol{\alpha}^\dagger = \mathbf{0}$), model (32) reduces to the traditional (static) factor model (1), which is our baseline specification.

It turns out that the joint tests of Ar(1)-Arch(1) in Propositions 2 and 4 is simply the sum of the separate tests:

Proposition 6 *1. Under the joint null hypothesis $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}, \boldsymbol{\alpha}^\dagger = \mathbf{0}$ the score test statistic*

$$LM_{AR(1)-ARCH(1)}(\boldsymbol{\eta}_0) = LM_{AR(1)}(\boldsymbol{\eta}_0) + LM_{ARCH(1)}(\boldsymbol{\eta}_0),$$

will be distributed as a χ^2 with $2(N+1)$ degrees of freedom for N fixed as T goes to infinity. This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}_0$ by their joint maximum likelihood estimators in Proposition 1.

2. *It also remains valid if we replace $\boldsymbol{\theta}_{s0}$ by its elliptically symmetric semiparametric estimator, which requires the nonparametric estimation of the density of $\varsigma_t(\boldsymbol{\theta}_s)$.*

3. *Under the same null hypothesis*

$$LM_{AR(1)-ARCH(1)}(\mathbf{0}) = LM_{AR(1)}(\mathbf{0}) + LM_{ARCH(1)}(\mathbf{0})$$

will also be distributed as a χ^2 with $2(N+1)$ degrees of freedom for N fixed as T goes to infinity irrespective of whether the elliptical conditional distribution is normal. This result continues to hold if we replace $\boldsymbol{\theta}_{s0}$ by its Gaussian pseudo maximum likelihood estimator $\overline{\boldsymbol{\theta}}_s$ in Proposition 1.

Intuitively, the reason is that the serial correlation orthogonality conditions (16)-(17) are asymptotically orthogonal to the Arch orthogonality conditions (26)-(27) because all odd order moments of multivariate spherical distributions are 0, which means that the joint test is simply the sum of its two components.

This additivity, though, no longer holds for non-spherical distributions, in which case one could robustify the Gaussian tests by using as weighting matrix

$$\begin{bmatrix} \mathcal{B}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}) & \mathcal{B}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) \\ \mathcal{B}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) & \mathcal{B}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \\ \mathcal{V}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \end{bmatrix} \odot \begin{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \\ \mathcal{V}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \end{bmatrix},$$

where

$$\begin{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \\ \mathcal{V}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho}) \end{bmatrix} = V \begin{bmatrix} f_{kt}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \\ \frac{1}{\sqrt{2}}[f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1} vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}] \end{bmatrix} \tag{33}$$

has to be computed taking into account the third and fourth multivariate moments of the distribution of $\mathbf{y}_t$, except for $\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho})$, whose Gaussian expression remains valid.

6 Monte Carlo analysis

6.1 Design

We assess the finite sample performance of the different testing procedures discussed above by means of an extensive Monte Carlo exercise, with an experimental design that nests those in sections 3.3 and 4.3, and is thereby adapted to the empirical application in section 7. For that reason, we only report the results for samples of 720 observations each (plus another 100 for initialisation) in which the cross-sectional dimension is $N = 5$. This sample size corresponds to 60 years of monthly data, roughly the same as in our empirical analysis. In this sense, the main reason for looking at a small cross-sectional dimension is to handicap our proposed tests relative to the existing multivariate serial dependence tests, which in the case of Hosking's test applied to $vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})']$ already involves 784 degrees of freedom for $N = 7$. We carry out 20,000 replications for the purposes of estimating actual sizes and powers with high precision.[16] All the examples of the DGP in (32) considered can be written as nonexchangeable single factor models of the form:

$$\begin{aligned}
y_{it} &= \pi_i + c_i x_t + u_{it} \qquad (i = 1,\ldots,5) \\
x_t &= \rho x_{t-1} + f_t \\
u_{it} &= \rho_i^* u_{it-1} + v_{it} \qquad (i = 1,\ldots,5) \\
\lambda_t &= (1-\alpha-\beta)(1-\rho^2) + \alpha(f_{kt-1}^2 + \omega_k - 1) + \beta\lambda_{t-1} \\
\gamma_{it} &= \gamma_i(1-\alpha_i^*-\beta_i^*)(1-\rho_i^*)^2 + \alpha_i^*(v_{it-1}^2 + c_i^2\omega_k - \gamma_i) + \rho_i^*\gamma_{it-1} \qquad (i = 1,\ldots,5)
\end{aligned}$$

[16] For instance, the 95% confidence interval for a nominal size of 5% would be (4.7%,5.3%).

with $\boldsymbol{\pi} = (.5, .4, .5, .4, .5)$, $\mathbf{c} = (5, 4, 5, 4, 5)$, $\boldsymbol{\gamma} = (5, 9, 5, 9, 5)$, $\rho_i^* = \rho^*$, $\alpha_i^* = \alpha^*$ and $\beta_i^* = \beta^*$ $\forall i$. Thus, the values of ρ, ρ^*, α, α^*, β, β^* fully explain the differences between our designs.

We generate samples from a Gaussian distribution, a Student t with 6 degrees of freedom, a discrete scale mixture of normals (DSMN) with the same kurtosis but finite higher order moments, and an asymmetric Student t such that the marginal distribution of an equally-weighted portfolio of $\mathbf{y}_t$ has the maximum negative skewness compatible with the kurtosis of a univariate t_6 (see Mencía and Sentana (2009, 2012) for details). These distributions allow us to assess the reliability of the robust Gaussian tests, and to shed some light on the "efficiency-consistency" trade-offs of those tests that exploit the leptokurtosis of financial returns.

We draw spherical Gaussian random vectors using the NAG library Fortran G05FDF routine after initialisation by G05CBF. To sample standardised Student t vectors, we simply divide those Gaussian random vectors by the square root of an independent univariate Gamma(3,2) random variable, and scale the result by 2. Similarly, we generate a standardised version of a two-component scale mixture of multivariate normals as

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

where $\boldsymbol{\varepsilon}_t^\circ$ is a spherical multivariate normal, $\varkappa$ the variance ratio of the two components, and s_t is an independent Bernoulli variate with $P(s_t = 1) = \pi$, which we draw by comparing π with a uniform from G05CAC. Specifically, we choose $\pi = .05$ and $\varkappa = .1438$. Finally, following Mencía and Sentana (2012), we generate a standardised asymmetric multivariate t by choosing

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\beta}\left[\xi_t^{-1} - c(\boldsymbol{\beta},\eta)\right] + \sqrt{\frac{\zeta_t}{\xi_t}}\boldsymbol{\Xi}^{1/2}\boldsymbol{\varepsilon}_t^\circ, \tag{34}$$

where 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 the $N \times N$ positive definite matrix

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

with

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

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 $\boldsymbol{\beta} = \mathbf{0}$. In the asymmetric t case, though, we use $\boldsymbol{\beta} = -10^6\boldsymbol{\iota}_N$.

Importantly, we use the same underlying pseudo-random numbers in all designs to minimise experimental error. In particular, we make sure that the standard Gaussian random vectors are the same for all four distributions. Given that the usual routines for simulating gamma random variables involve some degree of rejection, which unfortunately can change for different values of

η, we use the slower but smooth inversion method based on the NAG G01FFF gamma quantile function so that we can keep the underlying uniform variates fixed across simulations. Those uniform random variables are also used to generate the DSMN random vectors.

Finally, we combine the underlying random numbers with the vector of conditional means $\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$ provided by the relevant Kalman filter recursions, which we describe in appendix B of fiorentini and Sentana (2012).[17] We start up the recursions by exploiting covariance stationarity with $x_{-100|-100} = u_{i,-100|-100} = 0$, $\lambda_{-100} = 1-\rho^2$, $\gamma_{i,-100} = (1-\rho_i^{*2})\gamma_i$, $\boldsymbol{\Omega}_{11,-100|-100} = diag(1,\boldsymbol{\gamma}')$ and $\boldsymbol{\Omega}_{12,-100|-100} = \boldsymbol{\Omega}_{22,-100|-100} = diag(1-\rho^2, 1-\rho^{*2}\boldsymbol{\iota}_5)$.

For each Monte Carlo sample thus generated, our ML estimation procedure employs the following numerical strategy. First, we estimate the static mean and variance parameters $\boldsymbol{\theta}_s$ under normality with a scoring algorithm that combines the E04LBF routine with the analytical expressions for the score and the $\mathcal{A}(\boldsymbol{\phi}_0)$ matrix appearing in the proof of Proposition 1. For this purpose, the EM algorithm of Rubin and Thayer (1982) provides very good initial values. Then, we compute Mardia's (1970) sample coefficient of multivariate kurtosis κ, on the basis of which we obtain the sequential Method of Moments estimator of η suggested by Fiorentini, Sentana and Calzolari (2003), which exploits the theoretical relationship $\eta = \kappa/(4\kappa+2)$. Next, we could use this estimator as initial value for a univariate optimisation procedure that uses the E04ABF routine to obtain the sequential ML estimator of η discussed by Amengual, Fiorentini and Sentana (2013), which keeps $\boldsymbol{\pi}$, $\mathbf{c}$ and $\boldsymbol{\gamma}$ fixed at their Gaussian PML estimators. The resulting estimates of η, together with the PMLE of $\boldsymbol{\theta}_s$, become the initial values for the t-based ML estimators, which are obtained with the same scoring algorithm as the Gaussian PML estimator, but this time using the analytical expressions for the information matrix $\mathcal{I}(\boldsymbol{\phi}_0)$ in Proposition 1. We rule out numerically problematic solutions by imposing the inequality constraint $0 \leq \eta \leq .499$.

Computational details for the elliptically symmetric semiparametric procedure can be found in Appendix B of Fiorentini and Sentana (2010a). Given that a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise, we have chosen the "optimal" bandwidth in Silverman (1986).

6.2 Finite sample size

The size properties under the null of our proposed LM tests, Hosking's test, the univariate first-order serial correlation test of EWP, and the joint test of univariate first-order autocorre-

[17] The choice of a Cholesky factor is inconsequential for the all estimators of the static factor model parameters that we consider, and for all simulated distributions except the asymmetric t.

lation in all N series introduced in section 3.3 are summarised in Figures 5a-5d using Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. When the distribution is Gaussian, all tests are very accurate. The same conclusion is obtained when the distribution is a Student t, although in this case the elliptically symmetric semiparametric (SSP) tests show some very minor distortions. In contrast, when the true distribution is a DSMN, the tests based on the Student t PMLE's also show some size distortions, but they are very small. Finally, all tests are remarkably reliable when the conditional distribution is an asymmetric Student t, which partly reflects the fact that the elliptically symmetric estimators of the autocorrelation coefficients remain consistent in this case (see Proposition 10 in Fiorentini and Sentana (2010a)).

In turn, Figures 6a-6d show the size of the two-sided versions of our Arch(1) LM tests, Hosking's test applied to $vech[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})']$, a univariate first-order Arch test applied to EWP, the joint test of univariate first-order autocorrelation in all $N(N+1)/2$ squares and cross-products of the (demeaned) observed series introduced in section 4.3, and the analogous test that only focuses on their squares. In the Gaussian case, all tests are fairly accurate, except the SSP tests, which are rather conservative, and Hosking's test, which is rather liberal. This liberality is exacerbated when the true distribution is a Student t, and is shared to some extent by the diagonal version that looks at all $N(N+1)/2$ squares and cross-products, which reflects the imprecision in unrestrictedly estimating higher order moments in this case. As expected, the non-robust version of the normal test rejects far too often, while all the other tests follow a similar pattern: they are liberal for low significance values, and conservative for large ones. Not surprisingly, the sizes of the Student t tests also become highly distorted when the distribution is a DSMN, but the two robust versions of the normal tests are also somewhat unreliable in that context. Finally, those versions of the Gaussian tests that are only robust to kurtosis also suffer substantial size distortions when the conditional distribution is an asymmetric Student t, but the ones that are also robust to asymmetries are not very reliable either.

Figures 7a-7d show the size of all our two-sided LM tests for Garch(1,1) effects calculated with the discount factors $\bar{\beta} = \bar{\beta}^* = .94$ suggested in Riskmetrics (1996). The behaviour of these tests is fairly similar to that of the Arch(1) tests, although in this case the asymptotically valid tests show a stronger tendency to underreject in finite samples.

6.3 Finite sample power

In order to gauge the power of the serial correlation tests we look at a design in which $\rho = .03$ and $\rho_i^* = .045$ but $\alpha = \alpha^* = \beta = \beta^* = 0$. The evidence at the 5% significance level is presented in panels (a) and (b) of Table 1, which include raw rejection rates, as well as size adjusted ones based on the empirical distribution obtained under the null, which in this case provides the closest match because the Gaussian PML estimators of $\boldsymbol{\theta}_s$ that ignore the dynamics in $\mathbf{y}_t$ remain consistent in the presence of serial correlation or conditional heteroskedasticity, as shown by Doz and Lenglart (1999) and Sentana and Fiorentini (2001), respectively.

As expected from our theoretical analysis, the power of the normal tests does not depend much on the actual distribution of the data, while the tests that exploit the leptokurtosis of $\mathbf{y}_t$ offer noticeable power gains in the case of the multivariate t, especially the parametric versions. Another result that we saw in section 3.3 is that in this design the joint test of $H_0 : \boldsymbol{\rho}^{\dagger} = \mathbf{0}$ is only marginally more powerful than the joint test of $H_0 : \boldsymbol{\rho}^{*} = \mathbf{0}$, which in turn is substantially more powerful than the individual test of $H_0 : \rho = 0$. Standard serial correlation tests also behave very much in line with the theoretical analysis in that section.

We also look at a design with $\rho = \rho^* = 0$ but $\alpha = \alpha^* = .05$ and $\beta = \beta^* = 0.75$ to assess the power of the ARCH(1) and GARCH(1,1) tests. A comparison of panels (c)-(e) and (d)-(f) confirms that GARCH(1,1) tests are more powerful than their ARCH(1) counterparts, even though the Riskmetrics values for $\bar{\beta}$ and $\bar{\beta}^*$are much higher than the true values of these parameters. We also find that the power of the fully robust versions of the normal tests is slightly reduced when the distribution of the simulated data is leptokurtic. In contrast, the tests that exploit the leptokurtosis of $\mathbf{y}_t$ clearly become more powerful. Another result that we saw in section 4.3 is that in this design the joint tests of $H_0 : \boldsymbol{\alpha}^{\dagger} = \mathbf{0}$ are more powerful than the joint tests of $H_0 : \boldsymbol{\alpha}^{*} = \mathbf{0}$, which in turn are substantially more powerful than tests of $H_0 : \alpha = 0$. Finally, standard first-order serial correlation tests applied to the squares and cross-products of $\mathbf{y}_t$ do not have much power once we take into account their substantial size distortions under the null, except for the ARCH test applied to the EWP, which is almost as powerful as the analogous test for the common factor.

7 Empirical application

In this section we initially apply the procedures previously developed to the returns on five portfolios of US stocks grouped by industry in excess of the one-month Treasury bill rate (from Ibbotson Associates), which we have obtained from Ken French's Data Library. 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[18] (see <http://mba.tuck.dartmouth.edu/pages/

[18] Industry definitions: Cnsmr: Consumer Durables, NonDurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops). Manuf: Manufacturing, Energy, and Utilities. HiTec: Business Equipment, Telephone and

faculty/ken.french/data_library.html> for further details). We use monthly data from 1952 to 2008, so that our sample starts soon after the March 1951 Treasury - Federal Reserve Accord whereby the Fed stopped its wartime pegging of interest rates. Nevertheless, since we reserve 1952 to compute pre-sample values, we effectively work with 672 observations.

Table 2 contains the sample means, standard deviation and contemporaneous correlations for the excess returns on those portfolios. For our purposes, the two most relevant empirical characteristic are the strong degree of contemporaneous correlation between the series, and their leptokurtosis. Regarding the first aspect, it is customary to look at the ratio of the largest eigenvalue of the sample covariance matrix to its trace in order to judge the representativeness of the first principal component of $\mathbf{y}_t$. However, this measure, which is .79 in our case, fails to take into account the fact that unlike principal components, factor models fully explain the variances of all the $y_{it}'s$ thanks to the inclusion of idiosyncratic components. For that reason, we prefer to look at the fraction of the (square) Frobenius norm of the sample covariance matrix accounted for by a single factor model, which is 99.47%.[19]

As for the Gaussianity of the data, the Kuhn-Tucker test of normality against the alternative of multivariate Student t proposed by Fiorentini, Sentana and Calzolari (2003), which test the restriction on the first two moments of $\varsigma_t(\boldsymbol{\theta}_0)$ implicit in the single condition

$$E\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0)\right] = E[m_{kt}(\boldsymbol{\theta}_0)] = 0,$$

yields a value of 1478.9 despite having one degree of freedom. In contrast, the test of multivariate normal against asymmetric alternatives in Mencía and Sentana (2012), which assesses whether

$$E\left\{\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)\left[\varsigma_t(\boldsymbol{\theta}_0) - (N+2)\right]\right\} = E[m_{st}(\boldsymbol{\theta}_0, 0)] = \mathbf{0}, \tag{35}$$

yields 7.01, whose p-value is 22%. On this basis, we decided to estimate a multivariate t distribution. The ML estimator of the Student tail parameter η is .189, which corresponds to 5.3 degrees of freedom. This confirms our empirical motivation for developing testing procedures that exploit such a prevalent feature of the data.

Nevertheless, both parametric and semiparametric elliptically-based procedures are sensitive to the assumption of elliptical symmetry. For that reason, we follow Mencía and Sentana (2012), and test the null hypothesis of multivariate Student t innovations against the multivariate asym-

Television Transmission. Hlth: Healthcare, Medical Equipment, and Drugs. Other: Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance.

[19] The Frobenius norm of a general matrix $\mathbf{A}$, $||\mathbf{A}||$ say, is the Euclidean norm of $vec(\mathbf{A})$, which can be easily computed as the square root of the sum of its square singular values since $vec'(\mathbf{A})vec(\mathbf{A}) = tr(\mathbf{A}^2)$. Given that $V(\mathbf{y}_t)$ is a real, square symmetric matrix with spectral decomposition $\mathbf{U\Delta U}'$, with $\mathbf{U}$ orthonormal, it is easy to see $||V(\mathbf{y}_t)||^2$ can be additively decomposed as the sum of the square eigenvalues of $V(\mathbf{y}_t)$.

metric t distribution in (34). Their statistic checks the following moment conditions:

$$E\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\left[\varsigma_t(\boldsymbol{\theta})-(N+2)\right]\right]=E[m_{st}(\boldsymbol{\theta}_0,\eta_0)]=\mathbf{0},$$

which reduce to (35) when $\eta=0$. The asymptotic distribution that takes into account the fact that $\boldsymbol{\theta}$ and η have to be replaced by their t-based ML estimators $\hat{\boldsymbol{\theta}}_T$ and $\hat{\eta}_T$ is

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T} m_{st}(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T)\to N\left[0,2(N+2)(N\eta_0+1)\boldsymbol{\Sigma}_0\right].$$

The test statistic is 3.83 with a p-value of 57%, so we cannot reject the null hypothesis that the distribution of $\mathbf{y}_t$ is multivariate Student t at conventional levels.

Table 3 presents the three different estimates of the unconditional covariance parameters, namely Gaussian PMLE, Student t ML, and SSP. As can be seen, the discrepancies are fairly minor, especially in the case of estimators that exploit the leptokurtosis of the data. Consequently, the time series evolution of the corresponding Kalman filter estimates of the common factor are very highly correlated with each other ($>$.999), and also with the excess returns on the Fama and French market portfolio ($\simeq$.978), which corresponds to the value weighted return on all NYSE, AMEX and NASDAQ stocks in CRSP.

Table 4a reports the results of the two multivariate serial correlation tests discussed in section 3.3. As can be seen, there is evidence of first order serial correlation in the industry return series. Nevertheless, it is interesting to understand whether the dependence is due to the common factor or the specific ones. In this sense, note that we have considered not only tests against AR(1) dynamics in common and specific factors, but also tests against restricted AR(3) and AR(12) specifications in which the autoregressive coefficients are all assumed to be the same. The motivation for such tests is twofold. First, there is a substantial body of empirical evidence which suggests that expected returns are smooth processes, while observed returns have a small first order autocorrelation. Second, a rather interesting example of persistent expected returns is an AR(h) model in which $\boldsymbol{\rho}=\rho\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is a vector of h 1's. The results in section 3.2.3 imply that a test of $\rho=0$ in this context essentially involves assessing the significance of the sum of the first h autocorrelations of f_{kt}. In this sense, our procedure is entirely analogous to the one recommended by Jegadeesh (1989) to test for the long run predictability of individual asset returns without introducing overlapping observations (see also Cochrane (1991) and Hodrick (1992)). The intuition is that if returns contain a persistent but mean reverting predictable component, a persistent right hand side variable may pick it up.

The results reported in Table 4a show clear evidence of first order serial correlation in both common and specific factors. There is also some evidence that the idiosyncratic factors may have persistent mean-reverting components. In contrast, there is no evidence that such a component is present in the common factor. This interesting divergence could be due to the market being more closely followed by investors than the hedged components of the industry portfolios.

Table 4b presents our tests for conditional heteroskedasticity. Given the strong evidence for leptokurtosis, we only report the values of the fully robust versions of the different Gaussian tests. Not surprisingly, the multivariate serial dependence tests reject conditional homoskedasticity. We also find very strong evidence of ARCH effects in the idiosyncratic factors. In contrast, the ARCH(1) tests do not provide such a clear evidence in the case of the common factor. Nevertheless, the GARCH(1,1) tests strongly reject the null of conditionally homoskedasticity.

Our conclusions do not seem to be very sensitive to the degree of aggregation of our data. When we repeat exactly the same exercise with the excess returns of the ten portfolios of US stocks grouped by industry in Ken French's Data Library, we obtain rather similar results.

8 Conclusions and extensions

We obtain simple algebraic expressions for the score tests of serial correlation in the levels and squares of common and idiosyncratic factors in static factor models. The implicit orthogonality conditions resemble the orthogonality conditions of models with observed factors but the weighting matrices reflect their unobservability. We robustify our Gaussian tests against non-normality, and derive more powerful versions when the conditional distribution is elliptically symmetric, which can be either parametrically or semipametrically specified. We also explain how to derive tests that focus on either the common factors or the specific factors, or indeed on some of their elements.

Importantly, we show that despite the non-Gaussian nature of the state-space models that we consider, which makes it generally impossible to compute the log-likelihood function and its score without resorting to simulation methods, our tests coincide with the correct score tests.

We conduct Monte Carlo exercises to study the finite sample reliability and power of our proposed tests, and to compare them to existing multivariate serial dependence tests. Our simulation results suggest that the serial correlation tests have fairly accurate finite sample sizes, while the tests for conditional homoskedasticity show some size distortions. Given that $\mathbf{y}_t$ is *i.i.d.* under the null, it would be useful to explore bootstrap procedures, which could also exploit the fact that elliptical distributions are parametric in $N-1$ dimensions, and non-parametric in only one (see Dufour, Khalaf and Beaulieu (2010) for alternative finite-sample refinements of existing multivariate serial dependence tests). We also confirm that there are clear power gains from exploiting the cross-sectional dependence structure implicit in factor models, the leptokurtosis of financial returns, as well as the persistent behaviour of conditional variances.

Finally, we apply our methods to monthly stock returns on US broad industry portfolios. We find clear evidence in favour of first order serial correlation in common and specific factors, weaker evidence for persistent components in the idiosyncratic terms, and no evidence that such a component appears in the common factor. We also find strong evidence for persistent serial correlation in the volatility of common and specific terms.

It should be possible to robustify the serial dependence tests which assume that the return distribution is a Student t when in fact it is not along the lines described by Amengual and Sentana (2010) for mean-variance efficiency tests, and study their relative power in those circumstances. It should also be feasible to develop semiparametric tests that do not impose the assumption of elliptical symmetry. Another interesting extension would be to consider non-parametric alternatives such as the ones studied by Hong and Shehadeh (1999) and Duchesne and Lalancette (2003) among others, in which the lag length is implicitly determined by the choice of bandwidth parameter in a kernel-based estimator of a spectral density matrix. In addition, we could test for the effect of exogenous regressors in either the conditional mean vector or the conditional covariance matrix of returns. Moreover, we could use the test statistics that we have derived to obtain easy to compute indirect estimators of the dynamic models that define our alternative hypothesis along the lines suggested by Calzolari, Sentana and Fiorentini (2004).

The extension of our methods to models in which N/T is non-negligible would also constitute a very valuable addition with many potentially interesting empirical finance applications. In those circumstances, though, we would expect our proposed tests to be more reliable in finite sample and more powerful asymptotically than the Hosking-type multivariate serial correlation tests for the levels, squares and cross products of the variables under consideration, which involve orders of magnitude more degrees of freedom for fixed N.

Another particularly interesting extension would be to allow for serial dependence under the null. Specifically, suppose that we take as our new null hypothesis the factor model with AR(1) dynamics in the latent variables that we considered as the alternative in section 3.1, and as our new alternative a model with a common factor that follows an AR(2) process instead. Although a Lagrange Multiplier test of the new null hypothesis in the time domain is conceptually straightforward (see e.g. Engle and Watson (1981)), the algebra is incredibly tedious and the recursive scores difficult to interpret. In contrast, the frequency domain procedures in Harvey (1989) and Fernández (1990) yield scores which are once again entirely analogous to the univariate frequency domain score obtained if we treated the smoothed estimator of the common factor, $x_{kt|T}$, as if it were observed. We explore this interesting research avenue in Fiorentini and Sentana (2013).

References

Amengual, D., Fiorentini, G. and Sentana, E. (2013): "Sequential estimators of shape parameters in multivariate dynamic models", *Journal of Econometrics* 177, 233-249.

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

Anderson, B. D. O. and Moore, J. B. (1979) *Optimal Filtering*, Prentice Hall.

Anderson, T.W. and Rubin, H. (1956): "Statistical inference in factor analysis", in J. Neymann, ed., *Proceedings of the III Berkeley symposium on mathematical statistics and probability*, University of California, Berkeley.

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

Bai, J. and Ng, S. (2008): "Large dimensional factor analysis", *Foundations and Trends in Econometrics* 3, 89–163.

Bera, A. and Yoon, M.J. (1993): "Specification testing with locally misspecified alternatives", *Econometric Theory* 9, 649-658.

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

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

Box, G. E. P. and D.A. Pierce (1970): "Distribution of the autocorrelations in autoregressive moving average time series models", *Journal of the American Statistical Association* 65, 1509–1526.

Breusch, T. S. and Pagan, A.R. (1980): "The Lagrange Multiplier test and its applications to model specification in econometrics", *Review of Economic Studies* 47, 239-253.

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

Cochrane, J.H. (1991): "Volatility tests and efficient markets: a review essay", *Journal of Monetary Economics* 27, 661-676.

Connor, G. (1984): "A unified beta pricing theory", *Journal of Economic Theory* 34, 13-31.

Connor, G., Goldberg, L.R. Korajczik, R.A. (2010): *Portfolio risk analysis*, Princeton.

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

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

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

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

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

Dempster, A., Laird, N., and Rubin, D. (1977): "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Diebold, F.X. and M. Nerlove (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics* 4, 1-21.

Doz, C. and Lenglart, F. (1999): "Dynamic factor models: test of the number of factors and estimation with an application to the French Industrial Business Survey", *Annales d'Economie et Statistique* 54, 91-128.

Duchesne, P. and Lalancette, S. (2003): "On testing for multivariate ARCH effects in vector time series models", *Canadian Journal of Statistics* 31, 275-292.

Dufour, J.M., Khalaf, L. and Beaulieu, M.C. (2010): "Multivariate residual-based finite-sample tests for serial dependence and ARCH effects with application to asset pricing models", *Journal of Applied Econometrics* 25, 263-285.

Dungey, M., Martin, V.L. and Pagan, A.R. (2000): "A multivariate latent factor decomposition of international bond yield spreads", *Journal of Applied Econometrics* 15, 697-715.

Durbin, J. and Koopman, S.J. (2000): "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives", *Journal of the Royal Statistical Society Series B* 62, 3-56.

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

Engle, R.F. (1984): "Wald, likelihood ratio and Lagrange multiplier tests in econometrics", in *Handbook of Econometrics* vol II, ed. Griliches and Intrilligator (Amsterdam: North Holland), 775-826.

Engle, R.F. and Watson, M.W. (1981): "A one-factor multivariate time series model of metropolitan wage rates", *Journal of the American Statistical Association* 76, 774-781.

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

Fernández, F.J. (1990): "Estimation and testing of a multivariate exponential smoothing model", *Journal of Time Series Analysis* 11, 89–105.

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

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

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

Fiorentini, G. and Sentana, E. (2012): "Tests for serial dependence in static, non-Gaussian factor factor models", CEMFI Working Paper 1211 <http://www.cemfi.es/ftp/wp/1211.pdf>.

Fiorentini, G. and Sentana, E. (2013): "Dynamic specification tests for dynamic factor models", CEMFI Working Paper 1306.

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

Fiorentini, G., Sentana, E. and Shephard, N. (2004): "Likelihood estimation of latent generalised Arch structures", *Econometrica* 72, 1481-1517.

Gallant, A.R. and G. Tauchen (1996): "Which moments to match?", *Econometric Theory* 12, 657-681.

Godfrey, L.G. (1988): *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Econometric Society Monographs.

Guilkey, D.K. (1974): "Alternative tests for a first-order vector autoregressive error specification", *Journal of Econometrics* 4, 95-104.

Hafner, C.M. and Rombouts, J.V.K. (2007): "Semiparametric multivariate volatility models", *Econometric Theory* 23, 251-280.

Harvey, A.C. (1982): "A test of misspecification for systems of equations", LSE Econometrics Programme Discussion Paper A31.

Harvey, A.C. (1989): *Forecasting, structural models and the Kalman filter*, Cambridge University Press, Cambridge.

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

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

Hendry, D.F. (1971): "Maximum likelihood estimation of systems of simultaneous regression equations with errors generated by a vector autoregressive process", *International Economic*

Review 12, 257-272.

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

Hodrick, R. J. (1992): "Dividend yields and stock returns: alternative procedures for inference and measurement", *Review of Financial Studies* 5, 357-386.

Hong, Y. (1996): "Consistent testing for serial correlation of unknown form", *Econometrica* 64, 837-864.

Hong Y. and R.S. Shehadeh (1999): "A new test for Arch effects and its finite sample performance", *Journal of Business and Economic Statistics* 17, 91–108.

Hosking, J. R. M. (1981): "Lagrange-multiplier tests of multivariate time series models", *Journal of the Royal Statistical Society B* 43. 219-230.

Jegadeesh, N. (1989): "On testing for slow decaying components in stock prices", mimeo, Anderson Graduate School of Management, University of California at Los Angeles.

Jegadeesh, N. and G.G. Pennacchi (1996): "The behavior of interest rates implied by the term structure of eurodollar futures", *Journal of Money, Credit and Banking* 28, 426-446.

Kano, Y. (1983): "Consistency of estimators in factor analysis", *Journal of the Japan Statistical Society* 13, 137-144.

King, M.A., Sentana, E. and Wadhwani, S.B. (1994): "Volatility and links between national stock markets", *Econometrica* 62, 901-933.

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

Lawley, D.N. and Maxwell, A.E. (1971): *Factor analysis as a statistical method*, 2nd ed., Butterworths, London.

Lehmann, B., and Modest, D. (1988): "The empirical foundations of the Arbitrage Pricing Theory", *Journal of Financial Economics* 21, 213-254.

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

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, Chichester.

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

Maronna, R., Martin, D. and Yohai, V. (2006): *Robust Statistics - theory and methods*, Wiley.

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

Mencía, J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", *Review of Economics and Statistics* 94, 133-152.

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

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047-70.

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

Nijman, T. and Sentana, E. (1996): "Marginalization and contemporaneous aggregation of multivariate GARCH processes", *Journal of Econometrics* 71, 71-87.

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

Ross, S.A. (1976): "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, 13, 341-360.

Ross, S.A. (1978): "Mutual fund separation in financial theory - the separating distributions", *Journal of Economic Theory* 17, 254-286.

Rubin, D.B. and D.T. Thayer, D.T. (1982): "EM algorithms for ML factor analysis", *Psychometrika* 57, 69-76.

Sentana, E. (1991): *Time-varying volatility and returns on ordinary shares: an empirical investigation*, unpublished Doctoral Dissertation, University of London.

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

Sentana, E. (2004): "Factor representing portfolios in large asset markets", *Journal of Econometrics* 119, 257-289.

Sentana, E., Calzolari, G. and Fiorentini, G. (2008): "Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks", *Journal of Econometrics* 146, 10-25.

Sentana, E. and Fiorentini, G. (2001): "Identification, estimation and testing of conditionally heteroskedastic factor models", *Journal of Econometrics* 102, 143-164.

Sentana, E. and M. Shah (1993): "An index of co-movements in financial time series", CEMFI Working Paper 9415.

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

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

Appendices

A Proofs

Proposition 1

Given that the conditional density of $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$ under ellipticity, the log-likelihood function of a sample of size T 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 $d_t(\boldsymbol{\theta})=-1/2\ln|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ is the Jacobian, $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})=\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})$. Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. Fiorentini and Sentana (2010a) show that if $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank and $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ are differentiable, then

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

where

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

and $\delta(\varsigma_t,\boldsymbol{\eta})$ is defined in (11). Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi})=[\mathbf{e}_{dt}'(\boldsymbol{\phi}),\mathbf{e}_{rt}(\boldsymbol{\phi})]'$ evaluated at the true parameter values follows a vector martingale difference, and therefore, the same is true of the score vector $\mathbf{s}_t(\boldsymbol{\phi})$. His results also imply that, under suitable regularity conditions, the asymptotic distribution of the feasible ML estimator will be $\sqrt{T}(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0)\to N[\mathbf{0},\mathcal{I}^{-1}(\boldsymbol{\phi}_0)]$, where $\mathcal{I}(\boldsymbol{\phi}_0)=E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

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

and $\mathcal{M}(\boldsymbol{\eta})=V\left[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right]$. In particular, Crowder (1976) requires that: (a) $\boldsymbol{\phi}_0$ is locally identified and belongs to the interior of the admissible parameter space, which is a compact subset of $\mathbb{R}^{p+r}$; (b) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\phi}_0$;

(c) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\boldsymbol{\phi})$; and (d) $-E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right] T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi}) \xrightarrow{p} \mathbf{I}_{p+r}$, where $E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$. These regularity conditions are easy to verify in our *i.i.d.* context. In particular, the conditions in the theorem statement guarantee the identification of the factor model parameters and the positive definiteness of the Hessian matrix (see Theorem 12.1 in Anderson and Rubin (1956) and Theorem 2 in Kano (1983)). So the only remaining task is to find out the expression for the unconditional information matrix. In this context, Proposition 2 in Fiorentini and Sentana (2010a) states that:

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

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N,$$

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

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = -E\left\{\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\,\boldsymbol{\phi}\right\} = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}),$$

where $\mathrm{M}_{ll}(\boldsymbol{\eta})$, $\mathrm{M}_{ss}(\boldsymbol{\eta})$, $\mathrm{M}_{sr}(\boldsymbol{\eta})$ and $\mathcal{M}_{rr}(\boldsymbol{\eta})$ are defined in (7), (8), (9) and (10), respectively. Therefore, all we need is the matrix $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s)$, which in turn requires the Jacobian of the conditional mean and covariance functions. Differentiating the Kalman filter prediction equations we obtain $d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi}$ and

$$d\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_s) = d(\mathbf{cc}'+\boldsymbol{\Gamma}) = (d\mathbf{c})\mathbf{c}' + \mathbf{c}(d\mathbf{c}') + d\boldsymbol{\Gamma}$$

(see Magnus and Neudecker (1988)). Hence, the only three non-zero terms of the Jacobian will be:

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

As a result,

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0} \\ \mathbf{0} & [\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)] \\ \mathbf{0} & \frac{1}{2}\mathbf{E}'_N[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)] \end{bmatrix} = \mathbf{Z}_d(\boldsymbol{\phi}).$$

After some straightforward algebraic manipulations, we get that the elliptically symmetric score is

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi}) \\ \mathbf{s}_{\mathbf{c}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \\ \mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) &= \tfrac{1}{2}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \end{aligned} \quad \text{(A1)}$$

Assuming that $\Gamma > \mathbf{0}$ we can use the Woodbury formula

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) = \boldsymbol{\Gamma}^{-1} - (1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{cc}'\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Gamma}^{-1}[\boldsymbol{\Gamma} - (1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\mathbf{cc}']\boldsymbol{\Gamma}^{-1} \quad \text{(A2)}$$

to write

$$\begin{aligned}
&\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\\
=\ &\boldsymbol{\Gamma}^{-1}\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)f_{kt}(\boldsymbol{\theta}_s)-\mathbf{c}\omega_k(\boldsymbol{\theta}_s)\},\\
&\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\\
=\ &\boldsymbol{\Gamma}^{-1}\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}\}\boldsymbol{\Gamma}^{-1},
\end{aligned}$$

and

$$\varsigma_t(\boldsymbol{\theta}_s)=(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})=(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Gamma}^{-1}(\mathbf{y}_t-\boldsymbol{\pi})-f^2_{kt}(\boldsymbol{\theta}_s)/\omega_k(\boldsymbol{\theta}_s),$$

which greatly simplifies the computation of all the elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta})$, as well as $\mathbf{s}_{\boldsymbol{\eta}t}(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\theta})$ (see Sentana (2000)).

If we put all the previous elements together, we can finally obtain the conditional (and unconditional) information matrix, which in view of the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s)$ will be block diagonal between the elements corresponding to $\boldsymbol{\pi}$, and the elements corresponding to $(\mathbf{c},\boldsymbol{\gamma},\boldsymbol{\eta})$, with the diagonal blocks given in the statement of the first part of the proposition. Once again, the Woodbury formula simplifies considerably the computation of the information matrix when $\Gamma>\mathbf{0}$ because $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}=(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{c}$ and $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}=(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}$. Expression (A2) also makes clear that condition (ii) guarantees the full rank of the block of the information matrix corresponding to $\boldsymbol{\gamma}$.

Next, we can use Proposition 7 in Fiorentini and Sentana (2010a) to obtain the elliptically symmetric semiparametric score and corresponding efficiency bound. Specifically, they will be given by:

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0)=\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathbf{W}_d(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]\right\}, \tag{A3}$$

and

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_d(\boldsymbol{\phi}_0)\mathbf{W}'_d(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\}, \tag{A4}$$

respectively, where

$$\begin{aligned}
\mathbf{W}_d(\boldsymbol{\phi})&=\mathbf{Z}_d(\boldsymbol{\phi})[\mathbf{0}',vec'(\mathbf{I}_N)]'=E[\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}][\mathbf{0}',vec'(\mathbf{I}_N)]'\\
=E\left\{\frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})]\right|\left.\boldsymbol{\phi}\right\}&=-E\left\{\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|\,\boldsymbol{\phi}\right\}=E[\mathbf{W}_{dt}(\boldsymbol{\phi})|\boldsymbol{\phi}].
\end{aligned} \tag{A5}$$

But since in the case of a static factor model

$$\mathbf{W}'_{dt}(\boldsymbol{\phi})=\left[\begin{array}{ccc}\mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\end{array}\right]=\mathbf{W}'_d(\boldsymbol{\phi}), \tag{A6}$$

we will have that:

$$
\begin{aligned}
\mathring{\mathbf{s}}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) = &\quad \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}),\\
\mathring{\mathbf{s}}_{\mathbf{c}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) = &\quad \mathbf{s}_{\mathbf{c}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\left[\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta}_s)/N-1\} - \tfrac{2}{(N+2)\kappa+2}\left(\varsigma_t(\boldsymbol{\theta}_s)/N-1\right)\right],\\
\mathring{\mathbf{s}}_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) = &\quad \mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s,\boldsymbol{\eta}) - \tfrac{1}{2}vecd[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})]\left[\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta}_s)/N-1\} - \tfrac{2}{(N+2)\kappa+2}\left(\varsigma_t(\boldsymbol{\theta}_s)/N-1\right)\right].
\end{aligned}
$$

Expression (A6) also implies that the elliptically symmetric semiparametric efficiency bound will be block diagonal between $\boldsymbol{\pi}$ and $(\mathbf{c},\boldsymbol{\gamma})$, where the first block coincides with the first block of the information matrix, and the second one with the expression given in the second part of the proposition.

Finally, the Gaussian PML estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ sets to 0 the average value of $\mathbf{s}_{\boldsymbol{\theta}_s t}(\boldsymbol{\theta},\mathbf{0})$, which is trivially obtained from (A1) by noting that $\delta(\varsigma_t,\mathbf{0})=1$. Given that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied in our context, then we know from Proposition 3 in Fiorentini and Sentana (2010a) that $\sqrt{T}(\bar{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\to N\left[\mathbf{0},\mathcal{C}(\boldsymbol{\phi}_0)\right]$, where

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

$$
\text{and } \mathcal{K}(\kappa) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}] = \begin{bmatrix} \mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & (\kappa+1)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)+\kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{bmatrix}, \quad \text{(A7)}
$$

which only depends on $\boldsymbol{\eta}$ through κ. Hence, we can easily see that $\mathcal{A}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi})$ coincides with $\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\theta}_s,\mathbf{0})$ irrespective of the distribution of $\mathbf{y}_t$ because the model is static and $\mathcal{A}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s t}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s t}(\boldsymbol{\theta},\mathbf{0})|I_{t-1};\boldsymbol{\phi}\right]$ is equal to $\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s t}(\boldsymbol{\theta}_s,\mathbf{0})$. A closely related argument shows that $\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi})$ also mimics the expression for the information matrix if we replace $\text{M}_{ll}(\boldsymbol{\eta})$ by 1 and $\text{M}_{ss}(\boldsymbol{\eta})$ by $(\kappa+1)$.

More generally, if $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0$ 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 $\boldsymbol{\varrho}=\mathbf{0}$ denotes normality, then Proposition 1 in Bollerslev and Wooldridge (1992) coupled with the static nature of the model implies that:

$$
\mathcal{B}_t(\boldsymbol{\phi}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_s)\mathcal{K}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta}_s),
$$

where

$$
\mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}] \quad \text{(A8)}
$$

is the matrix of third and fourth order central moments of $\boldsymbol{\varepsilon}_t^*$, whose first block is the identity matrix of order N. This means that the block diagonality between $\boldsymbol{\pi}$ and $(\mathbf{c},\boldsymbol{\gamma})$ disappears if the true distribution is asymmetric even though $\mathcal{B}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\phi})$ continues to equal $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\theta}_s, \mathbf{0})$. In view of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_s, \mathbf{0})$, an alternative expression will be

$$\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi}) = V\left[\begin{array}{c} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \boldsymbol{\Gamma}^{-1}[\mathbf{v}_{kt}(\boldsymbol{\theta}_s) f_{kt}(\boldsymbol{\theta}_s) - \mathbf{c}\omega_k(\boldsymbol{\theta}_s)] \\ \frac{1}{2} vecd\{\boldsymbol{\Gamma}^{-1}[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}]\boldsymbol{\Gamma}^{-1}\} \end{array}\right],$$

which is more amenable for empirical applications. □

Proposition 2

Initially, the proof follows the same steps as the proof of Proposition 1. Therefore, we need expressions for $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$ to obtain $\mathbf{Z}_{dt}(\boldsymbol{\theta})$. Given our maintained assumption about the coincidence of the first two conditional moments with the Kalman filter prediction equations, we will have that

$$d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi} + d\begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix}\begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} d\begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix}$$

and

$$\begin{aligned} d\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \quad = \quad & d\begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix}\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} d\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} \\ & + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix}\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) d\begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix}, \end{aligned}$$

whence

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'} + [x_{t|t-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'} + \mathbf{c}\frac{\partial x_{t|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial\mathbf{u}_{t|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$$

and

$$\begin{aligned} \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} \quad = \quad & (\mathbf{I}_{N^2} + \mathbf{K}_{NN})[\begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix}\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\begin{pmatrix} \partial\mathbf{c}/\partial\boldsymbol{\theta}' \\ \mathbf{0} \end{pmatrix} \\ & + [\begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \otimes \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix}]\frac{\partial vec[\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}. \end{aligned}$$

Similarly,

$$\frac{\partial x_{t|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = x_{t|t-1}(\boldsymbol{\theta})\frac{\partial\rho}{\partial\boldsymbol{\theta}'} + \rho\frac{\partial x_{t-1|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

and

$$\frac{\partial\mathbf{u}_{t|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = [\mathbf{u}'_{t|t-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{E}_N\frac{\partial\boldsymbol{\rho}^*}{\partial\boldsymbol{\theta}'} + diag(\boldsymbol{\rho}^*)\frac{\partial\mathbf{u}_{t-1|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}.$$

In fact, it is easy to see that this last expression reduces to

$$\frac{\partial u_{it|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = u_{it|t-1}(\boldsymbol{\theta})\frac{\partial\rho_i^*}{\partial\boldsymbol{\theta}'} + \rho_i^*\frac{\partial u_{it-1|t-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}.$$

In addition, if we differentiate the updating equation we obtain

$$\frac{\partial vec[\mathbf{\Omega}_{t|t-1}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} = (\mathbf{I}_{(N+1)^2}+\mathbf{K}_{N+1,N+1})\left\{\begin{bmatrix}\rho & \mathbf{0}\\ \mathbf{0} & diag(\boldsymbol{\rho}^*)\end{bmatrix}\otimes\mathbf{I}_{N+1}\right\}\mathbf{E}_{N+1}\begin{pmatrix}\partial\rho/\partial\boldsymbol{\theta}'\\ \partial\boldsymbol{\rho}^*/\partial\boldsymbol{\theta}'\end{pmatrix}$$
$$+\mathbf{E}_{N+1}\begin{pmatrix}0\\ \partial\boldsymbol{\gamma}/\partial\boldsymbol{\theta}'\end{pmatrix}+\left\{\begin{bmatrix}\rho & \mathbf{0}\\ \mathbf{0} & diag(\boldsymbol{\rho}^*)\end{bmatrix}\otimes\begin{bmatrix}\rho & \mathbf{0}\\ \mathbf{0} & diag(\boldsymbol{\rho}^*)\end{bmatrix}\right\}\frac{\partial vec[\mathbf{\Omega}_{t-1|t-1}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}.$$

In principle, we would need to derive expressions for $\partial x_{t-1|t-1}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, $\partial u_{it-1|t-1}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec[\mathbf{\Omega}_{t-1|t-1}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$. However, since we are only interested in evaluating the score at $\rho=0$ and $\boldsymbol{\rho}^*=\mathbf{0}$, those expressions become unnecessary.

In addition, it is worth noting that under the null $x_{t|t-1}(\boldsymbol{\theta}_s,\mathbf{0})=0$, $\mathbf{u}_{t|t-1}(\boldsymbol{\theta}_s,\mathbf{0})=\mathbf{0}$, $\mathbf{\Omega}_{t|t-1}(\boldsymbol{\theta}_s,\mathbf{0})=diag(1,\boldsymbol{\gamma})$, $\mathbf{\Sigma}_t(\boldsymbol{\theta}_s,\mathbf{0})=\mathbf{cc}'+\mathbf{\Gamma}=\mathbf{\Sigma}(\boldsymbol{\theta}_s)$, $x_{t|t}(\boldsymbol{\theta}_s,\mathbf{0})=f_{kt}(\boldsymbol{\theta}_s)$ and $\mathbf{u}_{t|t}(\boldsymbol{\theta}_s,\mathbf{0})=\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$, so that

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_s,\mathbf{0})}{\partial\boldsymbol{\theta}'}=\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'}+\mathbf{c}f_{kt}(\boldsymbol{\theta}_s)\frac{\partial\rho}{\partial\boldsymbol{\theta}'}+diag[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)]\frac{\partial\boldsymbol{\rho}^*}{\partial\boldsymbol{\theta}'}$$

and

$$\frac{\partial vec[\mathbf{\Sigma}_t(\boldsymbol{\theta}_s,\mathbf{0})]}{\partial\boldsymbol{\theta}'}=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{c}\otimes\mathbf{I}_N)\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}+\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\theta}'}.$$

Hence

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})=\begin{bmatrix}\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\ \mathbf{0} & \frac{1}{2}(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ \mathbf{0} & \frac{1}{2}\mathbf{E}_N'[\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ f_{kt-1}(\boldsymbol{\theta}_s)\mathbf{c}'\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\ diag[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)]\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\end{bmatrix},$$

$$\mathbf{Z}_d(\boldsymbol{\phi})=\begin{bmatrix}\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\ \mathbf{0} & \frac{1}{2}(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ \mathbf{0} & \frac{1}{2}\mathbf{E}_N'[\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$

and

$$\mathbf{W}_d(\boldsymbol{\phi})=\begin{bmatrix}\mathbf{0}' & \mathbf{c}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0}'\end{bmatrix}', \tag{A9}$$

where we have used the fact that

$$\left.\begin{array}{l}E[f_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}]=E[\mathbf{c}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}]=0\\ E[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}]=E[\mathbf{\Gamma}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})|\boldsymbol{\theta}_s,\mathbf{0},\mathbf{0}]=\mathbf{0}\end{array}\right\} \tag{A10}$$

irrespective of the distribution of $\mathbf{y}_t$.

As a result, the elliptically symmetric score under the null will be

$$\begin{bmatrix}s_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\ s_{ct}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\ s_{\gamma t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\ s_{\rho t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\ s_{\rho^* t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\end{bmatrix}=\begin{bmatrix}\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})\\ \delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\\ \frac{1}{2}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)\}\\ f_{kt-1}(\boldsymbol{\theta}_s)\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{c}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})\\ diag[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)]\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})\end{bmatrix}.$$

Therefore, the only difference relative to the static factor model are the scores $s_{\rho t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta})$ and $s_{\rho^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta})$. In this sense, if we assume that $\boldsymbol{\Gamma} > \mathbf{0}$, then we can use the Woodbury formula once again to show that

$$\begin{bmatrix} s_{\rho t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \\ s_{\rho^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}] f_{kt-1}(\boldsymbol{\theta}_s) f_{kt}(\boldsymbol{\theta}_s) \\ \delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}] diag[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)] \boldsymbol{\Gamma}^{-1} \mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{bmatrix}.$$

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, \mathbf{0})$, together with (A10), it is easy to show that the unconditional information matrix $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta})$ will be block diagonal between $\boldsymbol{\pi}$, $(\mathbf{c}, \boldsymbol{\gamma}, \boldsymbol{\eta})$ and $\boldsymbol{\rho}^\dagger$, with the first two blocks as in the static case. Consequently, in computing our ML-based tests we can safely ignore the sampling uncertainty in estimating $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$. In addition, we can write

$$\mathcal{I}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger t}(\boldsymbol{\theta}, \mathbf{0}, \boldsymbol{\eta}) = diag \begin{bmatrix} f_{kt-1}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2} \mathbf{v}_{kt-1}(\boldsymbol{\theta}_s) \end{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta}) diag \begin{bmatrix} f_{kt-1}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2} \mathbf{v}_{kt-1}(\boldsymbol{\theta}_s) \end{bmatrix},$$

where

$$\begin{aligned} \mathcal{V}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta}) &= V \begin{bmatrix} \delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}] f_{kt}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2} \delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}] \mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{bmatrix} = \mathrm{M}_{ll}(\boldsymbol{\eta}) \begin{bmatrix} \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \end{bmatrix} \\ &= \mathrm{M}_{ll}(\boldsymbol{\eta}) \begin{bmatrix} (\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) & \mathbf{c}'\boldsymbol{\Gamma}^{-1/2}/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{c}/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) & \mathbf{I}_N - \boldsymbol{\Gamma}^{-1/2}\mathbf{c}\mathbf{c}'\boldsymbol{\Gamma}^{-1/2}/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) \end{bmatrix}. \end{aligned}$$

Thus, the only remaining item is the calculation of the second moments appearing in $\mathcal{V}_{\boldsymbol{\alpha}^\dagger \boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\eta})$. But since

$$\begin{aligned} & E[f_{kt}^2(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}] = E[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}] \\ = \; & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} = \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}), \\ & E\{\mathbf{v}_{kt}(\boldsymbol{\theta}_s) f_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}\} = E\{[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}\} \\ = \; & \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} = \mathbf{c}/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) \end{aligned}$$

and

$$\begin{aligned} & E\{\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}(\boldsymbol{\theta}_s)']|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}] = E[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}\} \\ = \; & \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} = \boldsymbol{\Gamma} - \mathbf{c}\mathbf{c}'/(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}), \end{aligned}$$

we finally obtain that $\mathcal{V}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\eta})$ mimics $\mathcal{V}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta})$ if we replace $\mathrm{M}_{ll}(\boldsymbol{\eta})$ by 1.

In addition, it follows from (A9) that the elliptically symmetric semiparametric scores for ρ and $\boldsymbol{\rho}^*$ coincide with the parametric ones, and that the elliptically symmetric semiparametric efficiency bound will be block diagonal between $\boldsymbol{\pi}$, $(\rho, \boldsymbol{\rho}^*)$ and $(\mathbf{c}, \boldsymbol{\gamma})$, where the first two blocks coincide with the first two blocks of the information matrix, and the third one with the corresponding bound in the static factor model.

Finally, let us consider the tests based on the Gaussian PML scores $s_{\rho t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ and $s_{\boldsymbol{\rho}^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ when $\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi}$ is *i.i.d.* $D(\boldsymbol{\pi}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_s); \boldsymbol{\varrho})$ but not necessarily normal or elliptical. To do so, let us partition the parameter vector $\boldsymbol{\theta}$ as $(\boldsymbol{\theta}_s, \boldsymbol{\rho}^\dagger)$. It is well known (see e.g. Engle (1984)) that a robust Gaussian pseudo score test of the null hypothesis $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}$ can be computed as

$$\left[\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \mathbf{s}'_{\boldsymbol{\rho}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0}) \right] \mathcal{A}^{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)\, \mathcal{C}^{-1}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)\, \mathcal{A}^{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0) \left[\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \mathbf{s}_{\boldsymbol{\rho}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0}) \right],$$

where $\mathbf{s}_{\boldsymbol{\rho}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0})$ is the Gaussian score evaluated at the restricted PML estimator $\tilde{\boldsymbol{\theta}}_s$, $\mathcal{A}^{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)$ is the relevant block of the inverse of the expected Hessian matrix $\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta} t}(\boldsymbol{\theta}, \mathbf{0}) | \boldsymbol{\phi}\right]$ and $\mathcal{C}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)$ is the corresponding block of the usual sandwich expression $\mathcal{C}(\boldsymbol{\phi}) = \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi})$, with $\mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta} t}(\boldsymbol{\theta}, \mathbf{0}) | \boldsymbol{\phi}\right]$. Once again, the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A10), implies that $\mathcal{A}(\boldsymbol{\phi})$ will be block diagonal between $(\rho, \boldsymbol{\rho}^*)$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$ irrespective of the true distribution of $\mathbf{y}_t$. In addition, $\mathcal{A}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi})$ will coincide with $\mathcal{I}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$. A closely related argument shows that $\mathcal{B}(\boldsymbol{\phi})$ will also be block diagonal between $(\rho, \boldsymbol{\rho}^*)$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$, and that $\mathcal{B}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}) = \mathcal{A}_{\boldsymbol{\rho}^\dagger \boldsymbol{\rho}^\dagger}(\boldsymbol{\phi})$, which validates the expression for $LM_{AR(1)}(\mathbf{0})$. □

Proposition 3

For the sake of brevity, the proof will be developed for the following univariate model:

$$\begin{aligned} y_t &= \pi + x_t + \gamma^{1/2} v_t^*, \\ x_t &= \rho x_{t-1} + f_t, \\ \begin{pmatrix} f_t \\ v_t^* \end{pmatrix} | I_{t-1} &\sim t\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \eta \right]. \end{aligned}$$

Nevertheless, it can be tediously extended to cover the general case. Given that when $\rho = 0$ the log-likelihood function of this model coincides with the log-likelihood function of the model considered in section 2, we only need to look at the score associated to this parameter.

It is easy to see that the joint distribution of y_t and x_t give the past of both variables will be

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} | I_{t-1} \sim t\left[\begin{pmatrix} \rho x_{t-1} \\ \rho x_{t-1} \end{pmatrix}, \begin{pmatrix} 1+\gamma & 1 \\ 1 & 1 \end{pmatrix}, \eta \right].$$

Hence, we can write down the joint log-likelihood as

$$c_2(\eta) - \frac{1}{2} \ln \gamma + g[\varsigma_t(\rho, \gamma); \eta],$$

where

$$c_2(\eta) = \ln\left[\Gamma\left(\frac{2\eta+1}{2\eta}\right)\right] - \ln\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \ln\left(\frac{1-2\eta}{\eta}\right) - \ln \pi$$

is the (log) constant of integration,

$$\gamma = \left| \begin{pmatrix} 1+\gamma & 1 \\ 1 & 1 \end{pmatrix} \right|$$

the Jacobian, and

$$g[\varsigma_t(\rho,\gamma);\eta] = -\left(\frac{2\eta+1}{2\eta}\right) \ln\left[1 + \frac{\eta}{1-2\eta}\varsigma_t(\rho,\gamma)\right],$$

with

$$\begin{aligned} \varsigma_t(\rho,\gamma) &= \begin{pmatrix} y_t - \rho x_{t-1} & x_t - \rho x_{t-1} \end{pmatrix} \begin{pmatrix} 1+\gamma & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_t - \rho x_{t-1} \\ x_t - \rho x_{t-1} \end{pmatrix} \\ &= \gamma^{-1}(y_t - x_t)^2 + (x_t - \rho x_{t-1})^2, \end{aligned}$$

the (log) kernel of the bivariate Student t density.

Given that we can write the standardised residuals as

$$\begin{aligned} \begin{pmatrix} 1+\gamma & 1 \\ 1 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} y_t - \rho x_{t-1} \\ x_t - \rho x_{t-1} \end{pmatrix} &= \begin{pmatrix} \frac{1}{\gamma} & -\frac{1}{\gamma} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t - \rho x_{t-1} \\ x_t - \rho x_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} \gamma^{-1/2}(y_t - x_t) \\ x_t - \rho x_{t-1} \end{pmatrix} \end{aligned}$$

and the gradient of the conditional mean vector with respect to ρ will be x_{t-1} times the vector $(1,1)'$, we will have that the score of the joint log-likelihood function with respect to ρ will be given by

$$\begin{aligned} -\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\rho,\gamma)} x_{t-1} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \gamma^{-1/2} & 0 \\ -\gamma^{-1/2} & 1 \end{pmatrix} \begin{pmatrix} \gamma^{-1/2}(y_t - x_t) \\ x_t - \rho x_{t-1} \end{pmatrix} \\ = -\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\rho,\gamma)}(x_t - \rho x_{t-1})x_{t-1}. \end{aligned}$$

The Kullback inequality then implies that score of the marginal log-likelihood function of y_t with respect to ρ will be given by

$$E\left[-\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\rho,\gamma)}(x_t - \rho x_{t-1})x_{t-1} \middle| \mathbf{Y}_T, \rho \right].$$

This expected value becomes analytically tractable when $\rho = 0$. First of all, the expression inside the expectation simplifies to

$$E\left[\frac{2\eta+1}{1-2\eta+\eta\left[\gamma^{-1}(y_t - x_t)^2 + x_t^2\right]} x_t x_{t-1} \middle| \mathbf{Y}_T, \rho = 0 \right].$$

Second, the joint distribution of y_t and x_t is *i.i.d.* over time, which means that the expected value of this product is equal to

$$E\left[\frac{2\eta+1}{1-2\eta+\eta\left[\gamma^{-1}(y_t - x_t)^2 + x_t^2\right]} x_t \middle| y_t, \rho = 0 \right] E\left[x_{t-1} \middle| y_{t-1}, \rho = 0 \right].$$

But the distribution of $x_t = f_t$ given y_t will also be t with mean

$$f_{kt}(\gamma) = \frac{1}{1+\gamma} y_t, \tag{A11}$$

variance

$$\begin{aligned} \upsilon_{kt}(\gamma,\eta) &= \frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\left(1-\frac{1}{1+\gamma}\right) \\ &= \frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\omega_k(\gamma) \end{aligned} \tag{A12}$$

and shape parameter

$$\frac{\eta}{1+\eta}, \tag{A13}$$

since the degrees of freedom of the conditional distribution of x_t given y_t are 1 plus the degrees of freedom of the joint distribution. Therefore, the second term is simply given by the lagged value of (A11). The first term is trickier, as we need to find the expected value of

$$\frac{2\eta+1}{1-2\eta+\eta\left[\gamma^{-1}(y_t-x_t)^2+x_t^2\right]} x_t. \tag{A14}$$

To do so, it is convenient to follow Fiorentini, Sentana and Calzolari (2003) and write x_t in terms of a conditionally standardised Student t component x_t^* as follows:

$$\begin{aligned} x_t &= \frac{1}{1+\gamma} y_t + \sqrt{\frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}} x_t^*, \\ x_t^* &= \sqrt{\frac{1-\eta}{\eta}} \times \sqrt{\zeta_t/\xi_t} u_t, \end{aligned}$$

where u_t is either 1 or -1 with probability 1/2, ζ_t is a chi-square random variable with 1 degree of freedom and ξ_t is a gamma random variable with mean $1+\eta^{-1}$ and variance $2(1+\eta^{-1})$, with u_t, ζ_t and ξ_t mutually independent and independent of y_t and I_{t-1}.

In turn, this decomposition implies that

$$\begin{aligned} \varsigma_t(0,\gamma) &= \gamma^{-1}(y_t-x_t)^2+x_t^2 = \frac{y_t^2}{1+\gamma} + \left(\frac{1+\gamma}{\gamma}\right)\left(x_t - \frac{1}{1+\gamma} y_t\right)^2 \\ &= \frac{y_t^2}{1+\gamma}\left(1+\frac{\zeta_t}{\xi_t}\right) + \frac{1-2\eta}{\eta}\frac{\zeta_t}{\xi_t}, \end{aligned}$$

so that the denominator of (A14) can be written as

$$1-2\eta+\eta\varsigma_t(0,\gamma) = \left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)\left(\frac{\xi_t+\zeta_t}{\xi_t}\right).$$

As a result, (A14) becomes

$$\begin{aligned} &\left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)^{-1}\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)\frac{2\eta+1}{1+\gamma} y_t \\ &+\left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)^{-1}\sqrt{\frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}}\sqrt{\frac{1-\eta}{\eta}}\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)\sqrt{\frac{\zeta_t}{\xi_t}} u_t. \end{aligned}$$

The expected value of the second summand conditional on y_t is 0 because of the symmetry of u_t. In contrast, we can use the properties of the beta distribution to prove that

$$E\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)=\frac{1+\eta}{1+2\eta}$$

and consequently, that

$$E\left[\left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)^{-1}\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)\frac{2\eta+1}{1+\gamma}y_t\middle|\,y_t\right]=\left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)^{-1}\frac{1+\eta}{1+\gamma}y_t.$$

Therefore, we have proved that

$$\begin{aligned}&E\left[\frac{2\eta+1}{1-2\eta+\eta\left[\gamma^{-1}(y_t-x_t)^2+x_t^2\right]}x_t x_{t-1}\middle|\,Y_T,\rho=0\right]\\&=\left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)^{-1}\frac{1+\eta}{1+\gamma}y_t\frac{1}{1+\gamma}y_{t-1}.\end{aligned}$$

Finally, using the general expressions for the score of the approximating model obtained in the proof of Proposition 2, we will have that the score with respect to ρ of such a univariate log-likelihood function under the null of $\rho=0$ will be given

$$\frac{\eta+1}{1-2\eta+\eta(1+\gamma)^{-1}y_t^2}(1+\gamma)^{-1/2}y_t(1+\gamma)^{-1/2}(1+\gamma)^{-1}y_{t-1},$$

as required. □

Proposition 4

We start again by differentiating the prediction equations, which yield $d\boldsymbol{\mu}_t(\boldsymbol{\theta})=d\boldsymbol{\pi}$ and

$$d\boldsymbol{\Sigma}_t(\boldsymbol{\theta})=(d\mathbf{c})\lambda_t(\boldsymbol{\theta})\mathbf{c}+\mathbf{c}[d\lambda_t(\boldsymbol{\theta})]\mathbf{c}'+\mathbf{c}\lambda_t(\boldsymbol{\theta})d\mathbf{c}'+d\boldsymbol{\Gamma}_t(\boldsymbol{\theta}),$$

whence

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}=\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\theta}'}$$

and

$$\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\mathbf{c}\lambda_{t|t-1}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'}+(\mathbf{c}\otimes\mathbf{c})\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\mathbf{E}_N\frac{\partial\boldsymbol{\gamma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}.$$

But since

$$\begin{aligned}\lambda_t(\boldsymbol{\theta})&=1+\alpha[E(f_{t-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})-1],\\\gamma_{it}(\boldsymbol{\theta})&=\gamma_i+\alpha_i^*[E(v_{it-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})-\gamma_i],\end{aligned}$$

we will have that:

$$\begin{aligned}\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=\alpha\frac{\partial E(f_{t-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\theta}}+\frac{\partial\alpha}{\partial\boldsymbol{\theta}}[E(f_{t-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})-1],\\\frac{\partial\gamma_{it}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=\frac{\partial\gamma_i}{\partial\boldsymbol{\theta}}+\alpha_i^*\frac{\partial E(v_{it-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\theta}}+\frac{\partial\alpha_i^*}{\partial\boldsymbol{\theta}}[E(v_{it-1}^2|Y_{t-1};\boldsymbol{\theta},\mathbf{0})-\gamma_i].\end{aligned}$$

This implies that under the null hypothesis of $\boldsymbol{\alpha}^{\dagger}=\mathbf{0}$,

$$
\begin{aligned}
\frac{\partial \lambda_t(\boldsymbol{\theta}_s,\mathbf{0})}{\partial \boldsymbol{\theta}} &= \frac{\partial \alpha}{\partial \boldsymbol{\theta}}[f^2_{kt-1}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1],\\
\frac{\partial \gamma_{it}(\boldsymbol{\theta}_s,\mathbf{0})}{\partial \boldsymbol{\theta}} &= \frac{\partial \gamma_i}{\partial \boldsymbol{\theta}}+\frac{\partial \alpha_i^*}{\partial \boldsymbol{\theta}}[v^2_{kit-1}(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i],
\end{aligned}
$$

where we have used the fact that $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_s,\mathbf{0})=\mathbf{cc}'+\boldsymbol{\Gamma}=\boldsymbol{\Sigma}(\boldsymbol{\theta}_s)\ \forall t$.

As a result,

$$
\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})=\begin{bmatrix}
\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\
\mathbf{0} & \frac{1}{2}(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]\\
\mathbf{0} & \frac{1}{2}\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]\\
\mathbf{0} & \frac{1}{2}[f^2_{kt-1}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1][\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{c}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]\\
\mathbf{0} & \frac{1}{2}dg[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt-1}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]
\end{bmatrix},
$$

whence it is easy to see that

$$
\mathbf{Z}_d(\boldsymbol{\phi})=\begin{bmatrix}
\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\
\mathbf{0} & \frac{1}{2}(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]\\
\mathbf{0} & \frac{1}{2}\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)]\\
\mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0}
\end{bmatrix}
$$

and

$$
\mathbf{W}_d(\boldsymbol{\phi})=\begin{bmatrix}\mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0}\end{bmatrix}', \tag{A15}
$$

where we have used the fact that

$$
\left.\begin{aligned}
E[f^2_{kt-1}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1|\boldsymbol{\theta}_s,\mathbf{0}]=0\\
E[v^2_{kit-1}(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i|\boldsymbol{\theta}_s,\mathbf{0}]=\mathbf{0}
\end{aligned}\right\} \tag{A16}
$$

irrespective of the true distribution of $\mathbf{y}_t$.

In addition, it follows that the elliptical score under the null will be:

$$
\begin{bmatrix}
s_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\
s_{\mathbf{c}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\
s_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\
s_{\alpha t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})\\
\\
s_{\alpha^* t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})
\end{bmatrix}=\begin{bmatrix}
\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})\\
\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\\
\frac{1}{2}vecd[\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)]\\
\frac{1}{2}[f^2_{kt-1}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]\\
\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\}\\
\frac{1}{2}dg[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt-1}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\\
\times vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\}
\end{bmatrix}.
$$

Therefore, the only difference relative to the static factor model are the scores $s_{\alpha t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})$ and $s_{\alpha^* t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})$. In this sense, if we assume that $\boldsymbol{\Gamma}>\mathbf{0}$ we can use the Woodbury formula to show that

$$
\begin{aligned}
&\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\\
=\ &\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]f^2_{kt}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1,
\end{aligned}
$$

so that

$$\left[\begin{array}{c} s_{\alpha t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \\ s_{\alpha^* t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \end{array}\right] = \left[\begin{array}{c} \frac{1}{2}[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\} \\ \frac{1}{2}dg[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}] \\ \times vecd\{\boldsymbol{\Gamma}^{-1}[\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\boldsymbol{\Gamma}^{-1}\} \end{array}\right].$$

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})$, together with (A16), it is easy to show that the unconditional information matrix $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0},\boldsymbol{\eta})$ will be block diagonal between $\boldsymbol{\pi}$, $(\mathbf{c},\boldsymbol{\gamma},\boldsymbol{\eta})$ and $\boldsymbol{\alpha}^\dagger$, with the first two blocks as in the static case. Consequently, in computing our ML-based tests we can safely ignore the sampling uncertainty in estimating $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$. In addition, we can write

$$\mathcal{I}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger t}(\boldsymbol{\theta},\mathbf{0},\boldsymbol{\eta}) = diag\left[\begin{array}{c} \frac{1}{\sqrt{2}}[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}] \end{array}\right]$$
$$\times\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta})\times diag\left[\begin{array}{c} \frac{1}{\sqrt{2}}[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}] \end{array}\right],$$

where

$$\begin{aligned}\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) &= V\left[\begin{array}{c} \frac{1}{\sqrt{2}}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\} \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}_{kt}'(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}\} \end{array}\right] \\ &= \mathrm{M}_{ss}(\boldsymbol{\eta})\left[\begin{array}{cc} [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \end{array}\right] \\ &+\frac{[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]}{2}\left[\begin{array}{cc} [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] \\ [\mathbf{c}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] & vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] \end{array}\right].\end{aligned} \tag{A17}$$

Thus, the only remaining item is the calculation of fourth order terms appearing in $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$. But if we write

$$f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1 = \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}-[1-\omega_k(\boldsymbol{\theta}_s)],$$

then it is easy to see that

$$\begin{aligned} & E[f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]^2 \\ = \;& E\{vec[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vec'[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]\} \\ & -[1-\omega_k(\boldsymbol{\theta}_s)]^2 \\ = \;& [\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})][\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}] \\ & -[1-\omega_k(\boldsymbol{\theta}_s)]^2 \\ = \;& [\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)](\kappa+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)] \\ & [\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}]-[1-\omega_k(\boldsymbol{\theta}_s)]^2 \\ = \;& (\kappa+1)\{2[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2+[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2\}-[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 = (3\kappa+2)[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2. \end{aligned}$$

Similarly, since

$$
\begin{aligned}
& vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}] \\
= \; & \mathbf{E}'_N\{vec[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]-vec[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]\},
\end{aligned}
$$

we will have that

$$
\begin{aligned}
& E\{vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]vecd'[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\} \\
= \; & \mathbf{E}'_N E\{vec[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]vec'[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]\}\mathbf{E}_N \\
& -vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]vecd'[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)] \\
= \; & \mathbf{E}'_N[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]E[vec(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t)vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t)][\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}] \\
& -vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]vecd'[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)] \\
= \; & \mathbf{E}'_N[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)](\kappa+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)] \\
& \times[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]-vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]vecd'[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)] \\
= \; & (\kappa+1)\{2[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}\odot\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}]+vecd[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]vecd'[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]\} \\
& -vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)]\mathbf{E}_N \\
= \; & 2(\kappa+1)[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}\odot\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]+\kappa vecd[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]vecd'[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]\}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
& E\{vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}][f^2_{kt}(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1]\} \\
= \; & \mathbf{E}'_N E\{vec[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}]vec'[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]\} \\
& -vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)][1-\omega_k(\boldsymbol{\theta}_s)] \\
= \; & \mathbf{E}'_N[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]E[vec(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t)vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t)][\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}] \\
& -vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)][1-\omega_k(\boldsymbol{\theta}_s)] \\
= \; & \mathbf{E}'_N[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)](\kappa+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)] \\
& \times[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{c}]-vecd[\boldsymbol{\Gamma}-\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)][1-\omega_k(\boldsymbol{\theta}_s)] \\
= \; & 2(\kappa+1)[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\odot\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]+\kappa vecd[\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}][\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}].
\end{aligned}
$$

Therefore, $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$ mimics $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta})$ if we replace $\mathrm{M}_{ss}(\boldsymbol{\eta})$ by $\kappa+1$.

In addition, it follows from (A15) that the elliptically symmetric semiparametric scores for $\boldsymbol{\alpha}^\dagger$ coincide with the parametric ones, and that the elliptically symmetric semiparametric efficiency bound will be block diagonal between $\boldsymbol{\pi}$, $(\mathbf{c},\boldsymbol{\gamma})$, and $\boldsymbol{\alpha}^\dagger$, where the first and last blocks coincide with the corresponding blocks of the information matrix, and the second one with the corresponding bound in the static factor model.

Finally, let us consider the tests based on the Gaussian PML scores $s_{\alpha t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ and $s_{\boldsymbol{\alpha}^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ when $\mathbf{y}_t|I_{t-1}; \boldsymbol{\phi}$ is *i.i.d.* $D(\boldsymbol{\pi}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_s); \boldsymbol{\varrho})$ but not necessarily normal or elliptical. Once again, a robust Gaussian pseudo score test of the null hypothesis $H_0 : \boldsymbol{\alpha}_1^\dagger = \mathbf{0}$ can be computed as

$$\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{s}'_{\boldsymbol{\alpha}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0})\right]\mathcal{A}^{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)\,\mathcal{C}^{-1}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)\,\mathcal{A}^{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{s}'_{\boldsymbol{\alpha}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0})\right],$$

where $\mathbf{s}'_{\boldsymbol{\alpha}^\dagger t}(\tilde{\boldsymbol{\theta}}_s, \mathbf{0}, \mathbf{0})$ is the Gaussian score evaluated at the restricted PML estimator $\tilde{\boldsymbol{\theta}}_s$, $\mathcal{A}^{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)$ is the relevant block of the inverse of the expected Hessian matrix $\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta} t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right]$ and $\mathcal{C}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)$ is the corresponding block of the usual sandwich expression $\mathcal{C}(\boldsymbol{\phi}){=}\mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi})$, with $\mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta} t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right]$ (see e.g. Engle (1984)). The structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A16) and the fact that $\mathcal{A}_{\boldsymbol{\theta\theta} t}(\boldsymbol{\phi})$ equals $\mathcal{I}_{\boldsymbol{\theta\theta} t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$, implies that $\mathcal{A}(\boldsymbol{\phi})$ will be block diagonal between $(\alpha, \boldsymbol{\alpha}^*)$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$ irrespective of the true distribution of $\mathbf{y}_t$. In addition, it is easy to see that

$$\mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) = E[\mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger t}(\boldsymbol{\phi})|\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}] = \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \odot \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}),$$

where

$$\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) = V\left[\left.\begin{array}{c}\frac{1}{\sqrt{2}}[f_{kt}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1]\\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s) + \mathbf{cc}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}]\end{array}\right|\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}\right].$$

A closely related argument shows that $\mathcal{B}_t(\boldsymbol{\phi})$ will also be block diagonal between $(\alpha, \boldsymbol{\alpha}^*)$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$. Further, the stationarity of $\mathbf{y}_t$ implies that

$$\mathcal{B}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) = E[\mathcal{B}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger t}(\boldsymbol{\phi})|\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}] = \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \odot \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}),$$

which is generally different from $\mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi})$. As we have seen in (A17) above, $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho})$ will simplify considerably when $\boldsymbol{\varepsilon}_t^*$ is spherical. In any case, the block diagonality of $\mathcal{A}(\boldsymbol{\phi})$ and $\mathcal{B}(\boldsymbol{\phi})$ implies that

$$\mathcal{A}^{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)\,\mathcal{C}^{-1}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)\,\mathcal{A}^{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0) = \mathcal{B}^{-1}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger t}(\boldsymbol{\phi}),$$

which proves the last part of the proposition. □

Proposition 5

For the sake of brevity, the proof will be developed for the following univariate model:

$$y_t = f_t + v_t,$$

$$\begin{pmatrix} f_t \\ v_t \end{pmatrix}|I_{t-1} \sim t\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 + \alpha(f_{t-1}^2 - 1) & 0 \\ 0 & \gamma \end{bmatrix}, \eta\right\},$$

where $\alpha \geq 0$ and $\gamma \geq 0$. Nevertheless, it can be tediously extended to cover the general case. Given that when $\alpha = 0$ the log-likelihood function of this model coincides with the log-likelihood function of the model considered in section 2, we only need to look at the score associated to this parameter.

It is easy to see that the joint distribution of y_t and f_t give the past of both variables will be

$$\begin{pmatrix} y_t \\ f_t \end{pmatrix} | I_{t-1} \sim t\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\alpha(f_{t-1}^2-1)+\gamma & 1+\alpha(f_{t-1}^2-1) \\ 1+\alpha(f_{t-1}^2-1) & 1+\alpha(f_{t-1}^2-1) \end{pmatrix}, \eta\right].$$

Hence, we can write down the joint log-likelihood as

$$c_2(\eta) - \frac{1}{2}\ln\gamma - \frac{1}{2}\ln[1+\alpha(f_{t-1}^2-1)] + g[\varsigma_t(\rho,\gamma);\eta],$$

where

$$c_2(\eta) = \ln\left[\Gamma\left(\frac{2\eta+1}{2\eta}\right)\right] - \ln\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \ln\left(\frac{1-2\eta}{\eta}\right) - \ln\pi$$

is the (log) constant of integration,

$$\gamma[1+\alpha(f_{t-1}^2-1)] = \left|\begin{pmatrix} 1+\alpha(f_{t-1}^2-1)+\gamma & 1+\alpha(f_{t-1}^2-1) \\ 1+\alpha(f_{t-1}^2-1) & 1+\alpha(f_{t-1}^2-1) \end{pmatrix}\right|$$

the Jacobian and

$$g[\varsigma_t(\alpha,\gamma);\eta] = -\left(\frac{2\eta+1}{2\eta}\right)\ln\left[1+\frac{\eta}{1-2\eta}\varsigma_t(\rho,\gamma)\right],$$

with

$$\begin{aligned}
\varsigma_t(\alpha,\gamma) &= \begin{pmatrix} y_t & f_t \end{pmatrix}\begin{pmatrix} 1+\alpha(f_{t-1}^2-1)+\gamma & 1+\alpha(f_{t-1}^2-1) \\ 1+\alpha(f_{t-1}^2-1) & 1+\alpha(f_{t-1}^2-1) \end{pmatrix}^{-1}\begin{pmatrix} y_t \\ f_t \end{pmatrix} \\
&= \gamma^{-1}(y_t-f_t)^2 + [1+\alpha(f_{t-1}^2-1)]^{-1}f_t^2,
\end{aligned}$$

the (log) kernel of the bivariate Student t density.

Given that we can write the standardised residuals as

$$\begin{aligned}
&\begin{pmatrix} 1+\alpha(f_{t-1}^2-1)+\gamma & 1+\alpha(f_{t-1}^2-1) \\ 1+\alpha(f_{t-1}^2-1) & 1+\alpha(f_{t-1}^2-1) \end{pmatrix}^{-1/2}\begin{pmatrix} y_t \\ f_t \end{pmatrix} \\
&= \begin{pmatrix} \gamma^{-1/2} & -\gamma^{-1/2} \\ 0 & [1+\alpha(f_{t-1}^2-1)]^{-1/2} \end{pmatrix}\begin{pmatrix} y_t \\ f_t \end{pmatrix} \\
&= \begin{pmatrix} \gamma^{-1/2}(y_t-f_t) \\ [1+\alpha(f_{t-1}^2-1)]^{-1/2}f_t \end{pmatrix}
\end{aligned}$$

and the gradient of the *vec* of the conditional covariance matrix with respect to α will be f_{t-1}^2-1 times the vector $(1,1,1,1)'$, we will have that the score of the joint log-likelihood function with respect to α will be given by

$$\frac{1}{2}(f_{t-1}^2-1)(1,1,1,1)$$
$$\times\left[\begin{pmatrix}\gamma^{-1/2} & 0\\ -\gamma^{-1/2} & [1+\alpha(f_{t-1}^2-1)]^{-1/2}\end{pmatrix}\otimes\begin{pmatrix}\gamma^{-1/2} & 0\\ -\gamma^{-1/2} & [1+\alpha(f_{t-1}^2-1)]^{-1/2}\end{pmatrix}\right]$$
$$\times vec\begin{pmatrix}\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{(y_t-f_t)^2}{\gamma}-1 & \frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{(y_t-f_t)}{\sqrt{\gamma}}\frac{f_t}{\sqrt{1+\alpha(f_{t-1}^2-1)}}\\ \frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{(y_t-f_t)}{\sqrt{\gamma}}\frac{f_t}{\sqrt{1+\alpha(f_{t-1}^2-1)}} & \frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{f_t^2}{1+\alpha(f_{t-1}^2-1)}-1\end{pmatrix}$$
$$=\frac{1}{2}\left(\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{f_t^2}{1+\alpha(f_{t-1}^2-1)}-1\right)\frac{f_{t-1}^2-1}{1+\alpha(f_{t-1}^2-1)},$$

where we have used the fact that

$$\begin{pmatrix}\gamma^{-1/2} & -\gamma^{-1/2}\\ 0 & [1+\alpha(f_{t-1}^2-1)]^{-1/2}\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}\gamma^{-1/2} & 0\\ -\gamma^{-1/2} & [1+\alpha(f_{t-1}^2-1)]^{-1/2}\end{pmatrix}$$
$$=\begin{pmatrix}0 & 0\\ 0 & [1+\alpha(f_{t-1}^2-1)]^{-1}\end{pmatrix}.$$

The Kullback inequality implies that score of the marginal log-likelihood function of y_t with respect to α will be given by

$$\frac{1}{2}E\left[\left(\frac{2\eta+1}{1-2\eta+\eta\varsigma_t(\alpha,\gamma)}\frac{f_t^2}{1+\alpha(f_{t-1}^2-1)}-1\right)\frac{f_{t-1}^2-1}{1+\alpha(f_{t-1}^2-1)}\middle|\mathbf{Y}_T,\alpha\right].$$

This expected value becomes analytically tractable when $\alpha=0$. First of all, the expression inside the expectation simplifies to

$$E\left[\left(\frac{(2\eta+1)f_t^2}{1-2\eta+\eta\left[\gamma^{-1}(y_t-f_t)^2+f_t^2\right]}-1\right)(f_{t-1}^2-1)\middle|\mathbf{Y}_T,\alpha=0\right].$$

Second, the joint distribution of y_t and f_t is *i.i.d.* over time, which means that the expected value of this product should be equal to

$$E\left[\left(\frac{(2\eta+1)f_t^2}{1-2\eta+\eta\left[\gamma^{-1}(y_t-f_t)^2+f_t^2\right]}-1\right)\middle|y_t,\alpha=0\right]E\left[f_{t-1}^2-1\middle|y_{t-1},\alpha=0\right].$$

But since f_t given y_t has a Student t distribution with (conditional) mean, variance and shape parameter given by (A11), (A12) and (A13), respectively, the second term is simply given by

$$f_{kt-1}^2(\gamma)+\upsilon_{kt-1}(\gamma,\eta)-1.$$

The first term is trickier, as we need to find the expected value of

$$\frac{(2\eta+1)f_t^2}{1-2\eta+\eta\left[\gamma^{-1}(y_t-f_t)^2+f_t^2\right]}-1. \tag{A18}$$

To do so, it is convenient to follow Fiorentini, Sentana and Calzolari (2003) and write f_t in terms of a conditionally standardised Student t component f_t^* as follows:

$$\begin{aligned} f_t &= \frac{1}{1+\gamma}y_t + \sqrt{\frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}}f_t^*, \\ f_t^* &= \sqrt{\frac{1-\eta}{\eta}} \times \sqrt{\zeta_t/\xi_t}u_t, \end{aligned}$$

where u_t is either 1 or -1 with probability 1/2, ζ_t is a chi-square random variable with 1 degree of freedom and ξ_t is a gamma random variable with mean $1+\eta^{-1}$ and variance $2(1+\eta^{-1})$, with u_t, ζ_t and ξ_t mutually independent and independent of y_t and I_{t-1}.

This decomposition allows us to express

$$\begin{aligned} \varsigma_t(0,\gamma) &= \gamma^{-1}(y_t-f_t)^2+f_t^2 = \frac{y_t^2}{1+\gamma}+\left(\frac{1+\gamma}{\gamma}\right)\left(f_t-\frac{1}{1+\gamma}y_t\right)^2 \\ &= \frac{y_t^2}{1+\gamma}+\frac{1-2\eta}{\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\zeta_t}{\xi_t} \\ &= \frac{y_t^2}{1+\gamma}\left(1+\frac{\zeta_t}{\xi_t}\right)+\frac{1-2\eta}{\eta}\frac{\zeta_t}{\xi_t}, \end{aligned}$$

so that the denominator of (A14) can be written as

$$\begin{aligned} 1-2\eta+\eta\varsigma_t(0,\gamma) &= 1-2\eta+\eta\frac{y_t^2}{1+\gamma}\left(1+\frac{\zeta_t}{\xi_t}\right)+(1-2\eta)\frac{\zeta_t}{\xi_t} \\ &= \left(1-2\eta+\frac{\eta y_t^2}{1+\gamma}\right)\left(\frac{\xi_t+\zeta_t}{\xi_t}\right) = (1-2\eta)\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\left(\frac{\xi_t+\zeta_t}{\xi_t}\right). \end{aligned}$$

As for the numerator, we are left with $2\eta+1$ times

$$\begin{aligned} f_t^2 &= \frac{1}{(1+\gamma)^2}y_t^2+\frac{1-2\eta}{\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}\frac{\zeta_t}{\xi_t} \\ &\quad +\frac{2}{1+\gamma}y_t\sqrt{\frac{1-2\eta}{\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}}\sqrt{\frac{\zeta_t}{\xi_t}}u_t. \end{aligned}$$

Therefore, we can re-write (A18) as -1 plus $2\eta+1$ times

$$\begin{aligned} &\frac{\frac{1}{(1+\gamma)^2}y_t^2}{(1-2\eta)\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)}\left(\frac{\xi_t}{\xi_t+\zeta_t}\right) \\ &+\frac{\gamma}{\eta(1+\gamma)}\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right) \\ &+\frac{\frac{2}{1+\gamma}y_t\sqrt{\frac{1-2\eta}{\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\frac{\gamma}{1+\gamma}}\sqrt{\frac{\zeta_t}{\xi_t}}u_t}{(1-2\eta)\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\gamma}\right)\left(\frac{\xi_t+\zeta_t}{\xi_t}\right)}. \end{aligned}$$

The expected value of the last summand is clearly 0 because of the symmetry of u_t. In contrast, we can use the properties of the beta distribution to prove that

$$E\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)=\frac{1+\eta}{1+2\eta}$$

and

$$E\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)=\frac{\eta}{1+2\eta}.$$

If we put all the pieces together we end up with

$$\begin{aligned}
& \frac{1+\eta}{\left(1-2\eta+\eta\frac{y^2}{1+\gamma}\right)}\frac{y_t^2}{(1+\gamma)^2}+\frac{\gamma}{1+\gamma}-1 \\
= & \frac{1+\eta}{\left(1-2\eta+\eta\frac{y^2}{1+\gamma}\right)}f_{kt}^2(\gamma)+\upsilon_{kt}(\gamma,0)-1 \\
= & \frac{1+\eta}{\left(1-2\eta+\eta\frac{y^2}{1+\gamma}\right)}\left[f_{kt}^2(\gamma)+\frac{1-\eta}{1+\eta}\upsilon_{kt}(\gamma,\eta)\right]-1.
\end{aligned}$$

As a result, the score of the true log-likelihood at $\alpha=0$ is

$$\frac{1}{2}\left[\frac{1+\eta}{\left(1-2\eta+\eta\frac{y^2}{1+\gamma}\right)}f_{kt}^2(\gamma)+\upsilon_{kt}(\gamma,0)-1\right]\left[f_{kt-1}^2(\gamma)+\upsilon_{kt-1}(\gamma,\eta)-1\right].$$

Interestingly, note that $\upsilon_{kt}(\gamma,\eta)$ is evaluated in the regressand at its Gaussian value $(=\omega_k(\gamma))$, while in the regressor it is evaluated at the true value of η.

Consider now the following HRS-style auxiliary model

$$y_t=f_t+v_t,$$

$$\begin{aligned}
\begin{pmatrix} f_t \\ v_t \end{pmatrix}|I_{t-1} & \sim t\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix},\begin{bmatrix} 1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1] & 0 \\ 0 & \gamma \end{bmatrix},\eta\right\}, \\
f_{t|t}(\boldsymbol{\theta}) & = \frac{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]}{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]+\gamma}\cdot y_t, \\
\omega_{t|t}(\boldsymbol{\theta},\eta) & = \frac{1-2\eta}{1-\eta}\left(1+\frac{\eta}{1-2\eta}\frac{y_t^2}{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]+\gamma}\right) \\
& \times\frac{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]}{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]+\gamma}\cdot\gamma,
\end{aligned}$$

and $\alpha\geq 0$, $\gamma\geq 0$. In order to compute the score of this model with respect to α, we need the derivative of the conditional variance of y_t with respect to this parameter. This derivative will be

$$[f_{t-1|t-1}^2(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)-1]+\alpha\left[2f_{t-1|t-1}(\boldsymbol{\theta})\frac{\partial f_{t-1|t-1}(\boldsymbol{\theta})}{\partial\alpha}+\frac{\partial\omega_{t-1|t-1}(\boldsymbol{\theta},\eta)}{\partial\alpha}\right].$$

However, since we are only interested in evaluating it at $\alpha=0$ we do not need to compute the second term.

The other component of the derivative will be given by the expression

$$\frac{1+\eta}{1-2\eta+\eta\varepsilon_t^{*2}(\boldsymbol{\theta})}\varepsilon_t^{*2}(\boldsymbol{\theta})-1,$$

where

$$\varepsilon_t^*(\boldsymbol{\theta}) = \frac{y_t}{\sqrt{1+\alpha[f_{t-1|t-1}^2(\boldsymbol{\theta}) + \omega_{t-1|t-1}(\boldsymbol{\theta},\eta) - 1] + \gamma}}.$$

Hence, under the null of $\alpha = 0$ the score with respect to α will be

$$\frac{1}{2}\left(\frac{1+\eta}{1-2\eta+\eta\frac{y_t^2}{1+\gamma}}\frac{y_t^2}{1+\gamma} - 1\right)\frac{1}{1+\gamma}[f_{t-1|t-1}^2(\gamma,0) + \omega_{t-1|t-1}(\gamma,0,\eta) - 1]$$

But since

$$\left(\frac{1+\eta}{1-2\eta+\eta\frac{y_t^2}{1+\gamma}}\frac{y_t^2}{1+\gamma} - 1\right)\frac{1}{1+\gamma} = \frac{1+\eta}{\left(1-2\eta+\eta\frac{y^2}{1+\gamma}\right)}\frac{1}{(1+\gamma)^2}y_t^2 + \frac{\gamma}{1+\gamma} - 1,$$

the pseudo log-likelihood score of the auxiliary model coincides with the score of the true model when we evaluate them at α=0. Hence, the Student t version of HRS auxiliary model smoothly embeds the true model at those parameter values.

Proposition 6

The proof of this proposition combines many elements of the proofs of Propositions 2 and 4. Given that model (32) reduces to model (13) when $\alpha = 0$ and $\boldsymbol{\alpha}^* = \mathbf{0}$ for every possible value of the parameters $\boldsymbol{\pi}, \rho, \boldsymbol{\rho}^*, \mathbf{c}$ and $\boldsymbol{\gamma}$, while it reduces to model (22) when $\rho = 0$ and $\boldsymbol{\rho}^* = \mathbf{0}$ for every possible value of the parameters $\boldsymbol{\pi}, \mathbf{c}$,$\boldsymbol{\gamma}$,$\alpha$ and $\boldsymbol{\alpha}^*$, then it trivially follows that under the joint null of $\boldsymbol{\rho}^\dagger = \mathbf{0}$ and $\boldsymbol{\alpha}^\dagger = \mathbf{0}$ we will have that

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}) = \left[\begin{array}{c} \boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \\ \mathbf{0} \\ \mathbf{0} \\ f_{kt-1}(\boldsymbol{\theta}_s)\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \\ diag[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)]\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \\ 0 \\ \mathbf{0} \end{array}\right.$$

$$\left.\begin{array}{c} \mathbf{0} \\ \frac{1}{2}(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\mathbf{0} \\ \frac{1}{2}\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)] \\ \mathbf{0} \\ \mathbf{0} \\ \frac{1}{2}[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1][\mathbf{c}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{c}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)] \\ \frac{1}{2}dg[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s)+\mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}]\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)] \end{array}\right],$$

whence

$$\mathbf{Z}_d(\boldsymbol{\phi}) = \left[\begin{array}{cc} \boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}) & \mathbf{0} \\ 0 & \frac{1}{2}(\mathbf{c}' \otimes \mathbf{I}_N)(\mathbf{I}_{N^2} + \mathbf{K}_{NN})[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \otimes \boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)] \\ \mathbf{0} & \frac{1}{2}\mathbf{E}_N'[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \otimes \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)] \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right]$$

and

$$\mathbf{W}_d(\boldsymbol{\phi}) = \left[\begin{array}{cccccccc} \mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right]'. \tag{A19}$$

As a result, the score vector under the null will be

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{ct}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\gamma t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ s_{\rho t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\boldsymbol{\rho}^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ s_{\alpha t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\alpha^* t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}) \end{array}\right] = \left[\begin{array}{c} \delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \\ \frac{1}{2}vecd[\delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \\ f_{kt-1}(\boldsymbol{\theta}_s)\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ diag[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)]\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \frac{1}{2}[f_{kt-1}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1]\{\delta[\varsigma(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \\ -\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\} \\ \frac{1}{2}dg[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}] \\ \times vecd[\delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \end{array}\right].$$

But this score is simply made up of the components of the different special cases that we have already studied, so the only thing left to do is to study the blocks of the information matrix and the other efficiency bounds that corresponds to the cross product of

$$[s_{\rho t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\rho}^* t}'(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta})]$$

with

$$[s_{\alpha t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\alpha}^* t}'(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0}, \boldsymbol{\eta})].$$

When the observed variables are elliptically distributed, the vector

$$[f_{kt-1}(\boldsymbol{\theta}_s), \mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s)]$$

is unconditionally orthogonal to the vector

$$\{[f_{kt-1}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1], vecd'[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}_{kt-1}'(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}]\},$$

so all the relevant off-diagonal blocks of $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$, $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$, $\mathcal{A}(\boldsymbol{\phi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_0)$ will be 0, which confirms the additive decomposition of the different joint tests under elliptical symmetry.

For general distributions, though, the expressions for $\mathcal{A}(\boldsymbol{\phi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_0)$ are more involved. Specifically, while it is still true that these matrices will remain block diagonal between $(\boldsymbol{\rho}^\dagger, \boldsymbol{\alpha}^\dagger)$ and $\boldsymbol{\theta}_s$ regardless of the true distribution of $\mathbf{y}_t$ in view of (A10) and (A16), and that $\mathcal{A}(\boldsymbol{\phi}_0)$ will also be block diagonal between $\boldsymbol{\rho}^\dagger$ and $\boldsymbol{\alpha}^\dagger$, with the relevant expressions for $\mathcal{A}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)$ and $\mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}_0)$ as in the proofs of Propositions 2 and 4, respectively, it will no longer be true that $\mathcal{B}(\boldsymbol{\phi}_0)$ will be block diagonal between Ar and Arch parameters, even though $\mathcal{B}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0) = \mathcal{A}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}_0)$. Nevertheless, straightforward calculations show that the blocks of $\mathcal{B}_t(\boldsymbol{\phi}_0)$ corresponding to $(\boldsymbol{\rho}^\dagger, \boldsymbol{\alpha}^\dagger)$ will be given by

$$
diag\begin{bmatrix} f_{kt-1}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s) \\ \frac{1}{\sqrt{2}}[f^2_{kt-1}(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1} vecd[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt-1}(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}] \end{bmatrix}
$$
$$
\times \begin{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \\ \mathcal{V}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \end{bmatrix}
$$
$$
\times diag\begin{bmatrix} f_{kt-1}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s) \\ \frac{1}{\sqrt{2}}[f^2_{kt-1}(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1} vecd[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt-1}(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}] \end{bmatrix},
$$

which confirms (33) in view of the stationarity of $\mathbf{y}_t$. □

B Local power calculations

Let $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the h influence functions used to develop the following moment test of $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$:

$$
M_T = T\bar{\mathbf{m}}'_T(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}), \tag{B20}
$$

where $\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0})$ is the sample average of $\mathbf{m}_t(\boldsymbol{\theta})$ evaluated under the null, $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix and $\boldsymbol{\theta}_{10}$ the true values of the remaining model parameters. In order to obtain the non-centrality parameter of this test under Pitman sequences of local alternatives of the form $H_{la} : \boldsymbol{\theta}_{2T} = \bar{\boldsymbol{\theta}}_2/\sqrt{T}$, it is convenient to linearise $\mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})$ with respect to $\boldsymbol{\theta}_2$ around its true value $\boldsymbol{\theta}_{2T}$. This linearisation yields

$$
\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) = \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{2T}) - \frac{1}{T}\sum\nolimits_{t=1}^{T} \frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}^*_{2T})}{\partial \boldsymbol{\theta}'_2}\bar{\boldsymbol{\theta}}_2,
$$

where $\boldsymbol{\theta}^*_{2T}$ is some "intermediate" value between $\boldsymbol{\theta}_{2T}$ and $\mathbf{0}$. As a result,

$$
\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) \to N[\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2, \boldsymbol{\Psi}],
$$

under standard regularity conditions, where

$$\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0}) = E[-\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})/\partial \boldsymbol{\theta}_2'],$$

so that the non-centrality parameter of the moment test (B20) will be

$$\bar{\boldsymbol{\theta}}_2' \mathbf{M}'(\boldsymbol{\theta}_{10}, \mathbf{0}) \boldsymbol{\Psi}^{-1} \mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0}) \bar{\boldsymbol{\theta}}_2. \tag{B21}$$

On this basis, we can easily obtain the limiting probability of M_T exceeding some pre-specified quantile of a central χ_h^2 distribution from the cdf of a non-central χ^2 distribution with h degrees of freedom and non-centrality parameter (B21). When $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ coincides with a subset of the true scores with respect to $\boldsymbol{\theta}_2$, $\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})$ and $\boldsymbol{\Psi}$ can be readily obtained from the relevant blocks of the information matrix. Similarly, they can be obtained from the $\mathcal{A}(\boldsymbol{\phi})$ and $\mathcal{B}(\boldsymbol{\phi})$ matrices, respectively, when $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ coincides with a subset of the Gaussian scores.

Importantly, (B21) remains valid when we replace $\boldsymbol{\theta}_{10}$ by its ML estimator under the null if $\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0})$ and the scores corresponding to $\boldsymbol{\theta}_1$, $\mathbf{s}_{\boldsymbol{\theta}_1 t}(\boldsymbol{\theta}_1, \mathbf{0})$ say, are asymptotically uncorrelated when H_0 is true, as in all our tests. The same applies to the Gaussian PMLE's because the matrices involved in the asymptotic expansions are block diagonal too. More generally, it would be convenient to work with the alternative influence functions

$$\mathbf{m}_t^{\perp}(\boldsymbol{\theta}_1, \mathbf{0}) = \mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0}) - cov[\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0}), \mathbf{s}_{\boldsymbol{\theta}_1 t}(\boldsymbol{\theta}_1, \mathbf{0})] V^{-1}[\mathbf{s}_{\boldsymbol{\theta}_1 t}(\boldsymbol{\theta}_1, \mathbf{0})] \mathbf{s}_{\boldsymbol{\theta}_1 t}(\boldsymbol{\theta}_1, \mathbf{0}),$$

which can be interpreted as the residual in the regression of $\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0})$ onto $\mathbf{s}_{\boldsymbol{\theta}_1 t}(\boldsymbol{\theta}_1, \mathbf{0})$.

Serial correlation tests

Let us assume without loss of generality that $\boldsymbol{\pi} = \mathbf{0}$. Hosking's test is effectively based on the influence functions

$$\mathbf{m}_{lt}(\boldsymbol{\theta}_s, \boldsymbol{\rho}^{\dagger}) = vec[\mathbf{y}_t \mathbf{y}_{t-1}' - \mathbf{G}_{\mathbf{y}}(1)]$$

evaluated at $\boldsymbol{\rho}^{\dagger} = \mathbf{0}$. But since

$$\mathbf{G}_{\mathbf{y}}(1) = \mathbf{c}\mathbf{c}'\rho + diag(\boldsymbol{\gamma} \odot \boldsymbol{\rho}^*)$$

for the model considered in section 3.3 in view of (15), and

$$vec[\mathbf{G}_{\mathbf{y}}(1)] = (\mathbf{c} \otimes \mathbf{c})\rho + vec[diag(\boldsymbol{\gamma} \odot \boldsymbol{\rho}^*)],$$

it trivially follows that

$$\mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0}) = E[\partial \mathbf{m}_{lt}(\boldsymbol{\theta}_s, \mathbf{0})/\partial \boldsymbol{\rho}^{\dagger\prime}] = -[\ (\mathbf{c} \otimes \mathbf{c}) \quad \mathbf{E}_N \boldsymbol{\Gamma}\ ].$$

Hence, we will have that

$$\mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\rho}}^\dagger = -[(\mathbf{c} \otimes \mathbf{c})\rho + \mathbf{E}_N \boldsymbol{\gamma} \rho^*]$$

when

$$\bar{\boldsymbol{\rho}}^{\dagger\prime} = \begin{pmatrix} \rho & \rho^* \boldsymbol{\iota}_N' \end{pmatrix}.$$

As for the asymptotic covariance matrix, the proof of Proposition 4 in Fiorentini and Sentana (2012) implies that if $\boldsymbol{\rho}^\dagger = \mathbf{0}$, then

$$\sqrt{T}\mathbf{m}_{lt}(\boldsymbol{\theta}_s, \mathbf{0}) = \sqrt{T} vec(\mathbf{y}_t \mathbf{y}_{t-1}') \to N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$$

irrespective of the distribution of $\mathbf{y}_t$.

Since the diagonal serial correlation test uses the influence functions

$$vecd[\mathbf{y}_t \mathbf{y}_{t-1}' - \mathbf{G}_{\mathbf{y}}(1)] = \mathbf{E}_N' vec[\mathbf{y}_t \mathbf{y}_{t-1}' - \mathbf{G}_y(1)],$$

it is easy to obtain the corresponding Jacobian matrix by premultiplying $\mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0})$ by $\mathbf{E}_N'$. Specifically,

$$\mathbf{E}_N' \mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\rho}}^\dagger = -[(\mathbf{c} \odot \mathbf{c})\rho + \boldsymbol{\gamma} \rho^*].$$

We can also exploit the properties of $\mathbf{E}_N$ (see Magnus (1988)) to show that under the null

$$\sqrt{T} vecd(\mathbf{y}_t \mathbf{y}_{t-1}') \to N(\mathbf{0}, \boldsymbol{\Sigma} \odot \boldsymbol{\Sigma}).$$

Finally, to obtain the non-centrality parameter for the serial correlation test of $\mathbf{w}'\mathbf{y}_t$, we simply have to exploit the fact that the relevant influence functions are

$$\mathbf{w}'\mathbf{y}_t \mathbf{y}_{t-1}' \mathbf{w} - \mathbf{w}' \mathbf{G}_y(1) \mathbf{w} = (\mathbf{w}' \otimes \mathbf{w}') vec[\mathbf{y}_t \mathbf{y}_{t-1}' - \mathbf{G}_y(1)],$$

so that the appropriate Jacobian will be $(\mathbf{w}' \otimes \mathbf{w}')\mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0})$, whence

$$(\mathbf{w}' \otimes \mathbf{w}')\mathbf{M}_l(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\rho}}^\dagger = -[(\mathbf{w}'\mathbf{c})^2 \rho + (\mathbf{w}' \boldsymbol{\Gamma} \mathbf{w}) \rho^*].$$

Similarly, it is straightforward to show that

$$\sqrt{T}(\mathbf{w}'\mathbf{y}_t \mathbf{y}_{t-1}' \mathbf{w}) \to N[0, (\mathbf{w}' \boldsymbol{\Sigma} \mathbf{w})^2].$$

In the case of the LM test of $H_0 : \boldsymbol{\rho}^\dagger = \mathbf{0}$, the information matrix equality implies that the Jacobian of the scores of $\boldsymbol{\rho}^\dagger$ with respect to $\boldsymbol{\rho}^\dagger$ will be given by (minus) the information matrix, which also gives us the covariance matrix of the scores under the null. By suitably selecting the relevant elements of $\mathcal{I}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\phi})$, we can also compute the non-centrality parameters for the tests of the null hypotheses $H_0 : \rho = 0$ and $H_0 : \boldsymbol{\rho}^* = \mathbf{0}$. Analogous comments apply to the Gaussian-based LM tests if we replace the elements of the information matrix by the appropriate elements of $\mathcal{A}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\phi})$ or $\mathcal{B}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\phi})$.

ARCH tests

To keep the algebra simple, we assume once again that $\boldsymbol{\pi} = \mathbf{0}$, that the conditional variances of common and specific factors have been generated according to (30) and that the conditional distribution is elliptically symmetric. Hosking's test applied to all the squares and cross-products of $\mathbf{y}_t$ is effectively based on the influence functions that correspond to the first-order autocovariance matrix of $vec(\mathbf{y}_t\mathbf{y}_t')$, $\mathcal{S}_{\mathbf{yy}}(1)$ say, evaluated at $\boldsymbol{\alpha}^\dagger = \mathbf{0}$. More specifically,

$$\mathbf{m}_{st}(\boldsymbol{\theta}_s, \boldsymbol{\alpha}^\dagger) = vec\{[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})] - \mathcal{S}_{\mathbf{y}}(1)\}.$$

But since

$$E(\mathbf{y}_t\mathbf{y}_t'|I_{t-1};\boldsymbol{\theta}) = \mathbf{cc}'\lambda_t + \boldsymbol{\Gamma}_t$$

so that

$$vec[E(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma}|I_{t-1};\boldsymbol{\theta})] = (\mathbf{c}\otimes\mathbf{c})(\lambda_t - 1) + \mathbf{E}_N(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}),$$

and

$$vec(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma}) = (\mathbf{c}\otimes\mathbf{c})(f_{t-1}^2 - 1) + vec(\mathbf{v}_{t-1}\mathbf{v}_{t-1}' - \boldsymbol{\Gamma}) + (\mathbf{I}_{N_2} + \mathbf{K}_{NN})(\mathbf{c}\otimes\mathbf{I}_N)f_{t-1}\mathbf{v}_{t-1},$$

then it follows that

$$\begin{aligned}
\mathcal{S}_{\mathbf{y}}(1) &= E[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})] = E\{E[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})|I_{t-1};\phi]vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})]\} \\
&= E\{[(\mathbf{c}\otimes\mathbf{c})(\lambda_t - 1) + \mathbf{E}_N(\boldsymbol{\gamma}_t - \boldsymbol{\gamma})][(\mathbf{c}'\otimes\mathbf{c}')(f_{t-1}^2 - 1) \\
&\quad + vec'(\mathbf{v}_{t-1}\mathbf{v}_{t-1}' - \boldsymbol{\Gamma}) + f_{t-1}\mathbf{v}_{t-1}'(\mathbf{c}'\otimes\mathbf{I}_N)(\mathbf{I}_{N_2} + \mathbf{K}_{NN})\} \\
&= (\mathbf{cc}'\otimes\mathbf{cc}')E[(\lambda_t - 1)(f_{t-1}^2 - 1)] + (\mathbf{c}\otimes\mathbf{c})E[(\lambda_t - 1)(\mathbf{v}_{t-1}'\odot\mathbf{v}_{t-1}' - \boldsymbol{\gamma}')]\mathbf{E}_N' \\
&\quad \mathbf{E}_N E[(\boldsymbol{\gamma}_t - \boldsymbol{\gamma})(f_{t-1}^2 - 1)](\mathbf{c}'\otimes\mathbf{c}') + \mathbf{E}_N E[(\boldsymbol{\gamma}_t - \boldsymbol{\gamma})(\mathbf{v}_{t-1}'\odot\mathbf{v}_{t-1}' - \boldsymbol{\gamma}')]\mathbf{E}_N'
\end{aligned}$$

because of the assumed elliptical symmetry and lack of cross-sectional correlation between f_t and the $v_{it}'s$, and the fact that we are assuming univariate ARCH(1) processes for them. This last assumption also implies that

$$E[(\lambda_t - 1)(f_{t-1}^2 - 1)] = \alpha V(f_{t-1}^2) = \alpha[E(f_{t-1}^4) - 1] = \alpha\left[\frac{3(\kappa+1)(1-\alpha^2)}{1-3(\kappa+1)\alpha^2} - 1\right] = \alpha\frac{(3\kappa+2)}{1-3(\kappa+1)\alpha^2},$$

where κ is the multivariate excess kurtosis coefficient. Similarly

$$E[(\gamma_{it} - \gamma_i)(v_{it-1}^2 - \gamma_i)] = \alpha_i V(v_{it-1}^2) = \alpha_i\frac{(3\kappa+2)}{1-3(\kappa+1)\alpha_i^2}\gamma_i^2.$$

In addition, we can show that

$$\begin{aligned}
E[(\gamma_{it} - \gamma_i)(v_{jt-1}^2 - \gamma_j)] &= \alpha_i cov(v_{it-1}^2, v_{jt-1}^2) = \alpha_i[E(v_{it-1}^2 v_{jt-1}^2) - \gamma_i\gamma_j] = \alpha_i\gamma_i\gamma_j\frac{\kappa}{1-(\kappa+1)\alpha_i\alpha_j}, \\
E[(\lambda_t - 1)(v_{it-1}^2 - \gamma_i)] &= \alpha cov(f_{t-1}^2, v_{it-1}^2) = \alpha\gamma_i\frac{\kappa}{1-(\kappa+1)\alpha\alpha_i}, \\
E[(\gamma_{it} - \gamma_i)(f_{t-1}^2 - 1)] &= \alpha_i cov(f_{t-1}^2, v_{it-1}^2) = \alpha_i\gamma_i\frac{\kappa}{1-(\kappa+1)\alpha\alpha_i}.
\end{aligned}$$

From here, it is straightforward to see that under the null of conditional homoskedasticity in common and idiosyncratic factors the only non-zero derivatives will be

$$\begin{aligned}
\partial E[(\lambda_t - 1)(f_{t-1}^2 - 1)/\partial\alpha &= (3\kappa + 2) \\
\partial E[(\gamma_{it} - \gamma_i)(v_{it-1}^2 - \gamma_i)]/\partial\alpha_i &= (3\kappa + 2)\gamma_i^2 \\
\partial E[(\gamma_{it} - \gamma_i)(v_{jt-1}^2 - \gamma_j)]/\partial\alpha_i &= \kappa\gamma_i\gamma_j \\
\partial E[(\lambda_t - 1)(v_{it-1}^2 - \gamma_i)]/\partial\alpha &= \kappa\gamma_i \\
\partial E[(\gamma_{it} - \gamma_i)(f_{t-1}^2 - 1)]/\partial\alpha_i &= \kappa\gamma_i
\end{aligned}$$

whence we can obtain the appropriate Jacobian matrix

$$\mathbf{M}_s(\boldsymbol{\theta}_s, \mathbf{0}) = \partial E[\mathbf{m}_t(\boldsymbol{\theta}_s, \mathbf{0})]/\partial\boldsymbol{\alpha}^{\dagger\prime}.$$

Finally, we will have that

$$\begin{aligned}
\mathbf{M}_s(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\alpha}}^{\dagger} &= -vec\{(\mathbf{cc}' \otimes \mathbf{cc}')(3\kappa + 2)\alpha + (\mathbf{c} \otimes \mathbf{c})\boldsymbol{\gamma}'\mathbf{E}_N'\kappa\alpha \\
&\quad + \mathbf{E}_N\boldsymbol{\gamma}(\mathbf{c}' \otimes \mathbf{c}')\kappa\alpha^* + \mathbf{E}_N[2(\kappa + 1)(\boldsymbol{\Gamma} \odot \boldsymbol{\Gamma}) + \kappa\boldsymbol{\gamma}\boldsymbol{\gamma}']\mathbf{E}_N'\alpha^*\} \qquad \text{(B22)}
\end{aligned}$$

when

$$\bar{\boldsymbol{\alpha}}^{\dagger\prime} = \begin{pmatrix} \alpha & \alpha^*\boldsymbol{\iota}_N' \end{pmatrix}.$$

As for the asymptotic covariance matrix, the proof of Proposition 8 in Fiorentini and Sentana (2012) implies that if $\boldsymbol{\rho}^{\dagger} = \mathbf{0}$, then

$$\sqrt{T}\mathbf{m}_{st}(\boldsymbol{\theta}_s, \mathbf{0}) = \sqrt{T}vec[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})] \to N\{0, [\mathbf{H}(\kappa) \otimes \mathbf{H}(\kappa)]\},$$

when the conditional distribution of $\mathbf{y}_t$ is elliptically symmetric, where

$$\mathbf{H}(\kappa) = (\kappa + 1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \kappa vec(\boldsymbol{\Sigma})vec'(\boldsymbol{\Sigma}) = \mathbf{H}(\kappa).$$

But given that the autocovariance matrix of $vech(\mathbf{y}_t\mathbf{y}_t')$ will be

$$\mathbf{D}_N^+ E[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})]\mathbf{D}_N^{+\prime} = \mathbf{D}_N^+ \mathcal{S}_{\mathbf{yy}}(1)\mathbf{D}_N^{+\prime},$$

it is straightforward to obtain the relevant limiting mean vector as

$$(\mathbf{D}_N^+ \otimes \mathbf{D}_N^+)\mathbf{M}_s(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\alpha}}^{\dagger}.$$

Similarly, the proof of Proposition 8 in Fiorentini and Sentana (2012) also implies that

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T} vec[vech(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vech'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})] \to N[\mathbf{0}, (\mathbf{D}_N^+\mathbf{H}(\kappa)\mathbf{D}_N^{+\prime} \otimes \mathbf{D}_N^+\mathbf{H}(\kappa)\mathbf{D}_N^{+\prime})],$$

where

$$\mathbf{D}_N^+\mathbf{H}(\kappa)\mathbf{D}_N^{+\prime} = 2(\kappa+1)\mathbf{D}_N^+(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{D}_N^{+\prime} + \kappa vech(\boldsymbol{\Sigma})vech'(\boldsymbol{\Sigma})].$$

From here, we can obtain the non-centrality parameter for the test that only looks at the marginal autocovariances of $vech(\mathbf{y}_t\mathbf{y}_t')$ by premutiplying by $\mathbf{E}'_{N(N+1)/2}$.

In turn, the diagonalisation matrix $\mathbf{E}_N$ allows us to obtain the autocovariance matrix of $vecd(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})$ as

$$\mathbf{E}'_N E[vec(\mathbf{y}_t\mathbf{y}_t' - \boldsymbol{\Sigma})vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}' - \boldsymbol{\Sigma})]\mathbf{E}_N = \mathbf{E}'_N\mathcal{S}_{\mathbf{y}}(1)\mathbf{E}_N,$$

whence we can obtain the non-centrality parameter for the test that only looks at the marginal autocovariances of $vecd(\mathbf{y}_t\mathbf{y}_t')$ by premutiplying $\mathbf{M}_s(\boldsymbol{\theta}_s,\mathbf{0})\bar{\boldsymbol{\alpha}}^\dagger$ by $(\mathbf{E}'_N\otimes\mathbf{E}'_N)$. An analogous manipulation yields the asymptotic covariance matrix of the relevant influence functions.

Finally, it is straightforward to obtain the autocovariance structure of the squares of any linear combination of $\mathbf{y}_t$, $\mathbf{w}'\mathbf{y}_t$ say, by exploiting the fact that

$$E[(\mathbf{w}'\mathbf{y}_t)^2(\mathbf{w}'\mathbf{y}_{t-1})^2] = vec'(\mathbf{w}\mathbf{w}')E[vec(\mathbf{y}_t\mathbf{y}_t')vec'(\mathbf{y}_{t-1}\mathbf{y}_{t-1}')]vec(\mathbf{w}\mathbf{w}').$$

Similarly, it is easy to prove that

$$\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T}(\mathbf{w}'\mathbf{y}_t)^2(\mathbf{w}'\mathbf{y}_{t-1})^2 \to N[0,(3\kappa+2)(\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w})^2]$$

under the null.

In the case of the LM test of $H_0:\boldsymbol{\alpha}^\dagger=\mathbf{0}$, the information matrix equality implies that the Jacobian of the scores of $\boldsymbol{\alpha}^\dagger$ with respect to $\boldsymbol{\alpha}^\dagger$ will be given by (minus) the information matrix, which also gives us the covariance matrix of the scores under the null. By suitably selecting the relevant elements of $\mathcal{I}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\phi})$, we can also compute the non-centrality parameters for the tests of the null hypotheses $H_0:\alpha=0$ and $H_0:\boldsymbol{\alpha}^*=\mathbf{0}$. Analogous comments apply to the Gaussian-based LM tests if we replace the elements of the information matrix by the appropriate elements of $\mathcal{A}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\phi})$ or $\mathcal{B}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\phi})$.

Table 1

Test power

(a) AR(1) tests. DGP: Gaussian (ρ=.03,ρ_i^*=.045,α=α^*=β=β^*=0)

	Common			Specific			Joint			Hosking		
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP	Gen	Vecd	EWP
Rejection rate	0.121	0.121	0.126	0.395	0.396	0.401	0.402	0.402	0.411	0.203	0.110	0.121
Size adjusted	0.116	0.115	0.117	0.390	0.391	0.376	0.398	0.399	0.381	0.209	0.109	0.117

(b) AR(1) tests. DGP: Student t_6 (ρ=.03,ρ_i^*=.045,α=α^*=β=β^*=0)

	Common			Specific			Joint			Hosking		
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP	Gen	Vecd	EWP
Rejection rate	0.120	0.143	0.155	0.391	0.500	0.524	0.397	0.509	0.539	0.202	0.110	0.120
Size adjusted	0.119	0.143	0.138	0.394	0.502	0.479	0.399	0.511	0.489	0.206	0.110	0.118

(c) ARCH(1) tests. DGP: Gaussian (ρ=ρ^*=0,α=α^*=.05,β=β^*=.75)

	Common			Specific			Joint			Hosking			
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP	Gen	Vech	Vecd	EWP
Rejection rate	0.263	0.261	0.228	0.391	0.391	0.315	0.469	0.473	0.389	0.279	0.197	0.219	0.259
Size adjusted	0.270	0.270	0.264	0.401	0.405	0.391	0.480	0.487	0.475	0.215	0.192	0.222	0.265

(d) ARCH(1) tests. DGP: Student t_6 (ρ=ρ^*=0,α=α^*=.05,β=β^*=.75)

	Common			Specific			Joint			Hosking			
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP	Gen	Vech	Vecd	EWP
Rejection rate	0.229	0.238	0.259	0.377	0.397	0.444	0.438	0.484	0.543	0.510	0.293	0.258	0.226
Size adjusted	0.265	0.267	0.268	0.339	0.384	0.423	0.390	0.467	0.517	0.196	0.189	0.223	0.265

(e) GARCH(1,1) tests ($\bar{\beta}$=$\bar{\beta}^*$=.94). DGP: Gaussian (ρ=ρ^*=0,α=α^*=.05,β=β^*=.75)

	Common			Specific			Joint		
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP
Rejection rate	0.321	0.321	0.292	0.499	0.499	0.437	0.592	0.594	0.525
Size adjusted	0.358	0.355	0.350	0.538	0.540	0.533	0.631	0.635	0.622

(f) GARCH(1,1) tests ($\bar{\beta}$=$\bar{\beta}^*$=.94). DGP: Student t_6 (ρ=ρ^*=0,α=α^*=.05,β=β^*=.75)

	Common			Specific			Joint		
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP
Rejection rate	0.286	0.330	0.352	0.456	0.545	0.600	0.530	0.652	0.714
Size adjusted	0.337	0.372	0.380	0.511	0.554	0.612	0.574	0.662	0.726

Table 2

Descriptive statistics

Industry portfolios

Sector	Means	Std.dev.	Correlations				
			Cnsmr	Manuf	HiTec	Hlth	Other
Cnsmr	.566	4.481	*1*				
Manuf	.543	4.178	.804	*1*			
HiTec	.497	5.320	.734	.718	*1*		
Hlth	.733	4.995	.710	.668	.634	*1*	
Other	.500	4.998	.878	.848	.739	.708	*1*

Notes: Sample: January 1953-December 2008. 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.

Table 3

Estimates of $\mathbf{\Sigma} = \mathbf{cc}' + \Gamma$

Industry portfolios

	Factor Loadings			Specific Variances		
Sector	PML	ML	SSP	PML	ML	SSP
Cnsmr	4.130	4.309	4.292	3.024	3.263	3.215
Manuf	3.708	3.840	3.847	3.710	3.683	3.705
HiTec	4.223	4.337	4.342	10.465	8.453	8.997
Hlth	3.791	4.120	4.075	10.574	10.915	10.870
Other	4.740	4.900	4.909	2.518	3.105	3.062

Notes: Sample: January 1953-December 2008. 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. PML refers to the Gaussian-based ML estimators, ML to the Student t ones, and SSP to the elliptically symmetric semiparametric estimators.

Table 4a

Serial correlation tests (p-values, %)

	AR(1)			AR(3)			AR(12)		
	PML	ML	SSP	PML	ML	SSP	PML	ML	SSP
Common factor	0.35	2.64	1.35	19.75	35.49	24.04	39.59	53.85	59.63
Specific factors	1.46	2.70	1.45	1.40	8.84	4.11	0.06	0.00	0.00
Joint	0.11	0.87	0.30	1.52	11.31	4.71	0.11	0.00	0.00

Table 4b

Conditional heteroskedasticity tests (p-values, %)

	ARCH(1)			GARCH(1,1)		
	PML	ML	SSP	PML	ML	SSP
Common factor	0.36	6.12	1.79	0.00	0.26	0.01
Specific factors	0.00	0.00	0.00	0.00	0.00	0.00
Joint	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Sample: July:1962-June:2007. 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. PML refers to the (fully robust) tests based on the Gaussian ML estimators, ML to the Student t ones, SSP to the elliptically symmetric semiparametric estimators.

Figure 1: Power of mean dependence tests at 5% level against local alternatives

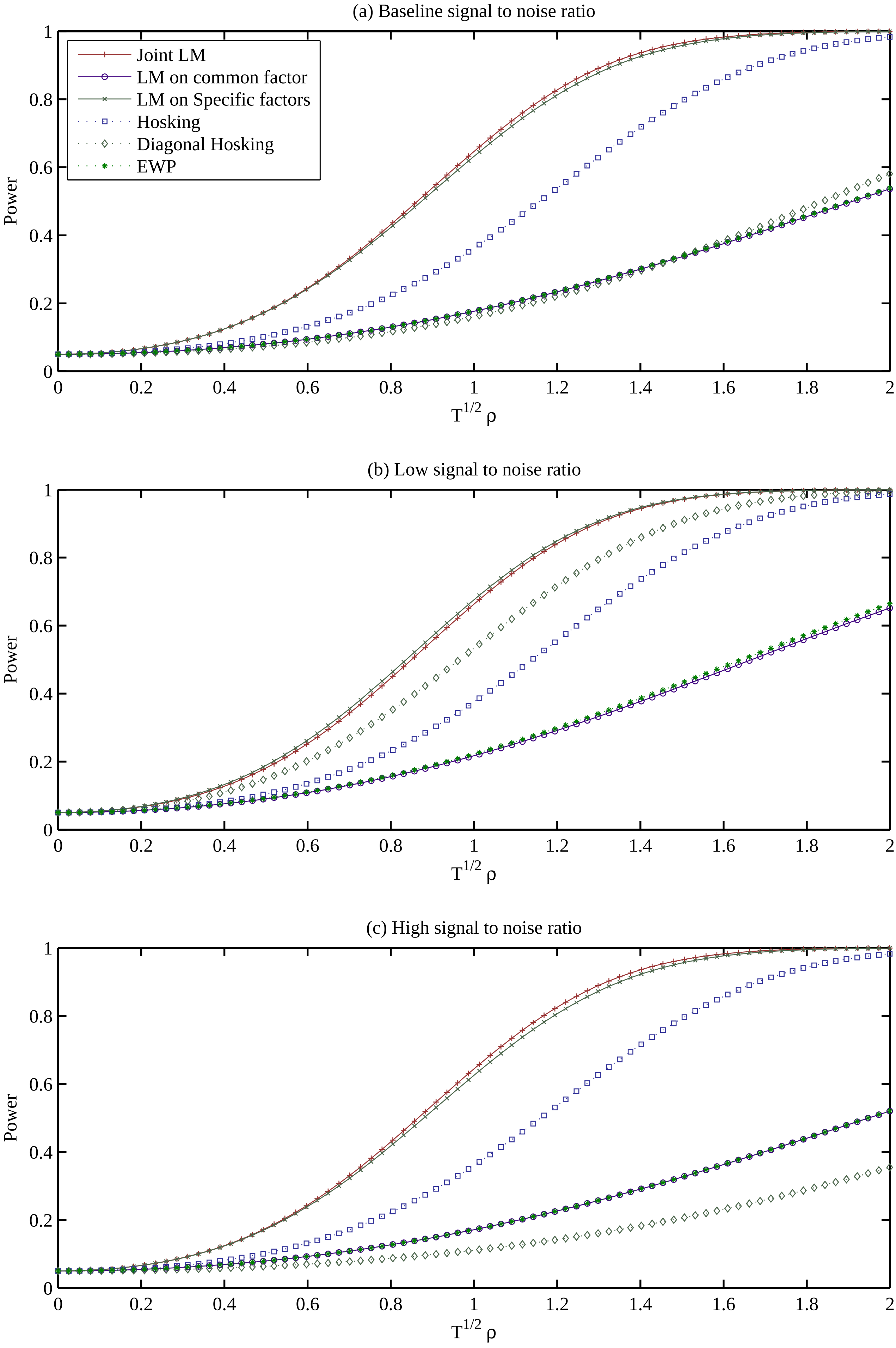


Figure 2: Power of mean dependence tests at 5% level against local alternatives

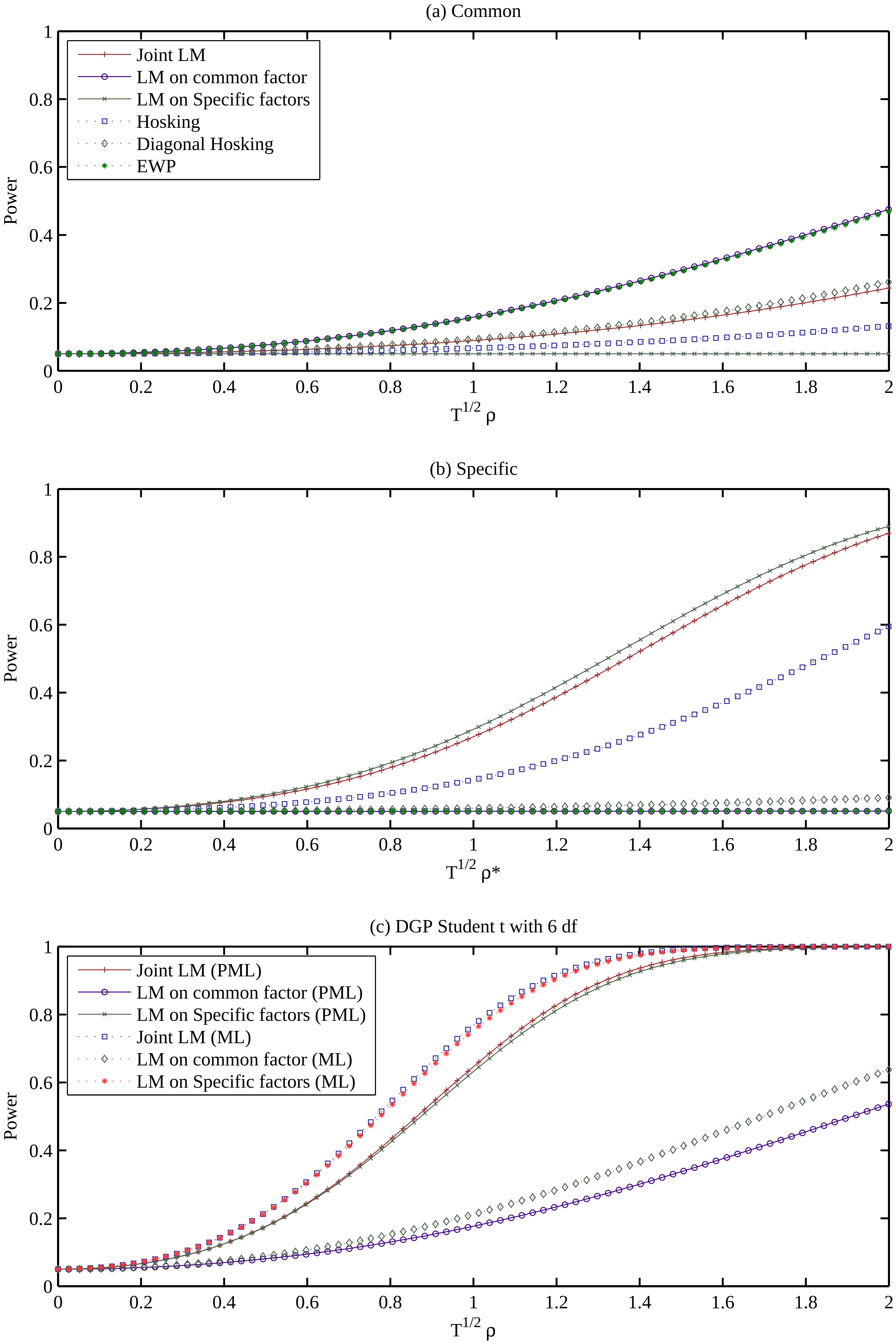


Figure 3: Power of variance dependence tests at 5% level against local alternatives

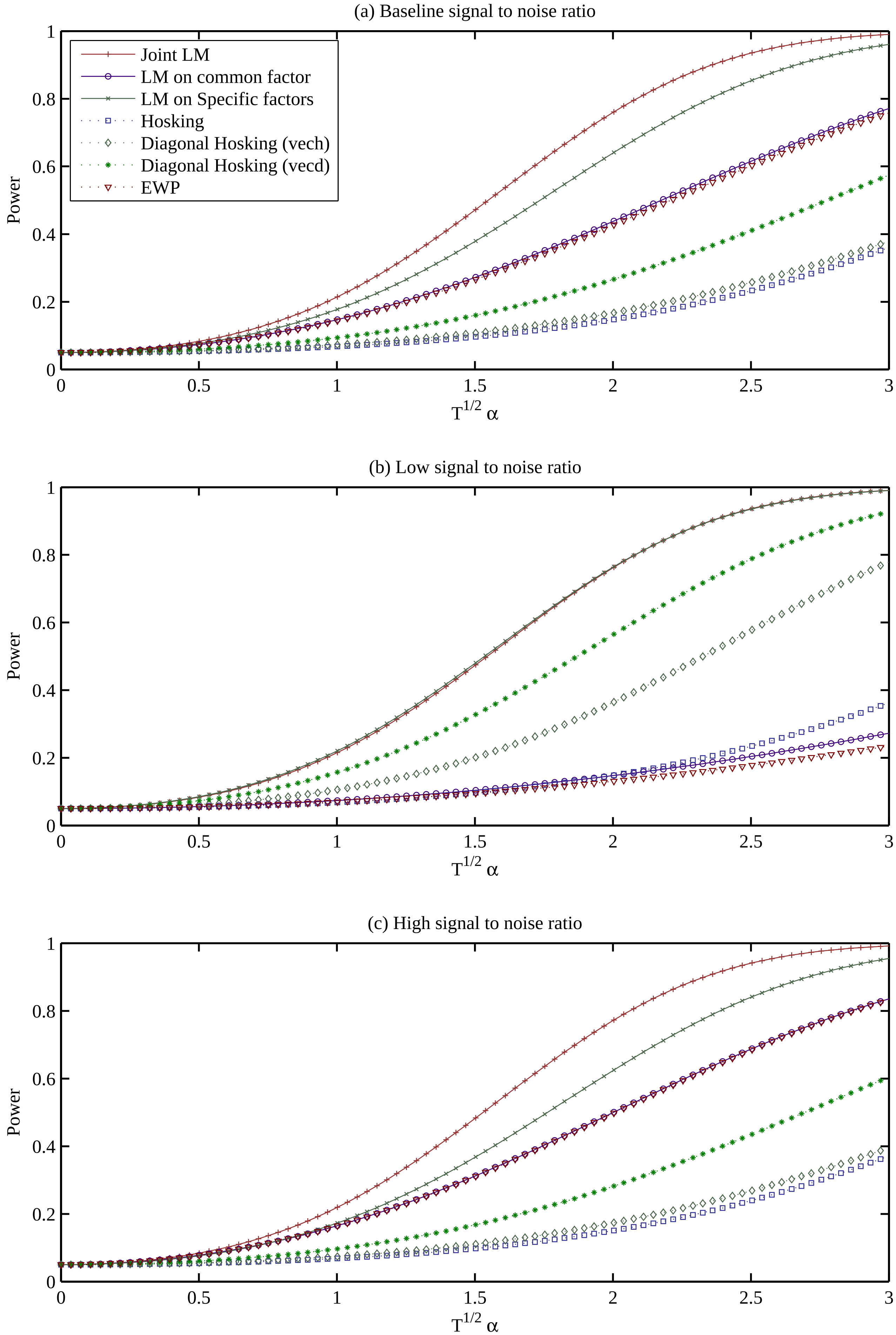


Figure 4: Power of variance dependence tests at 5% level against local alternatives

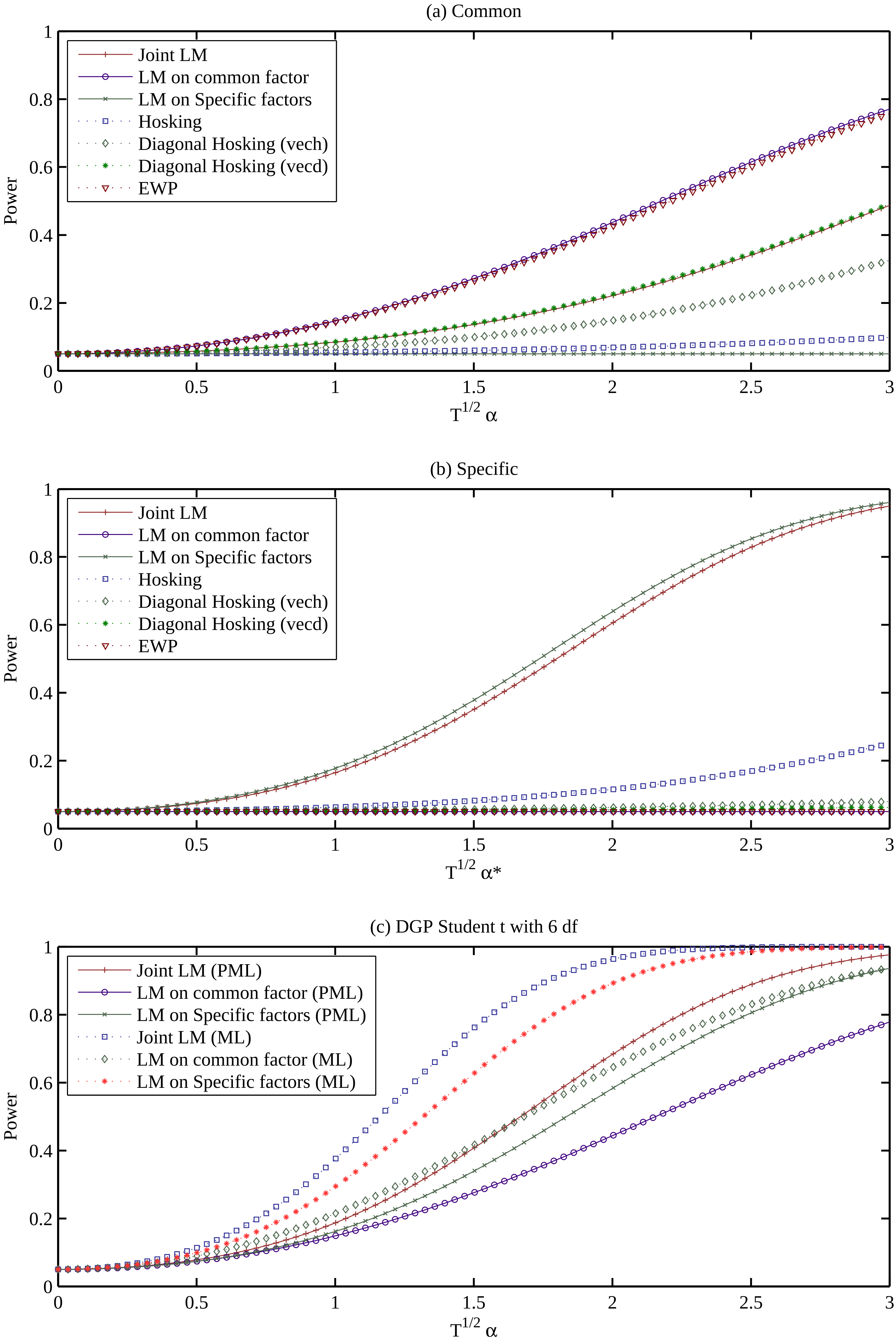
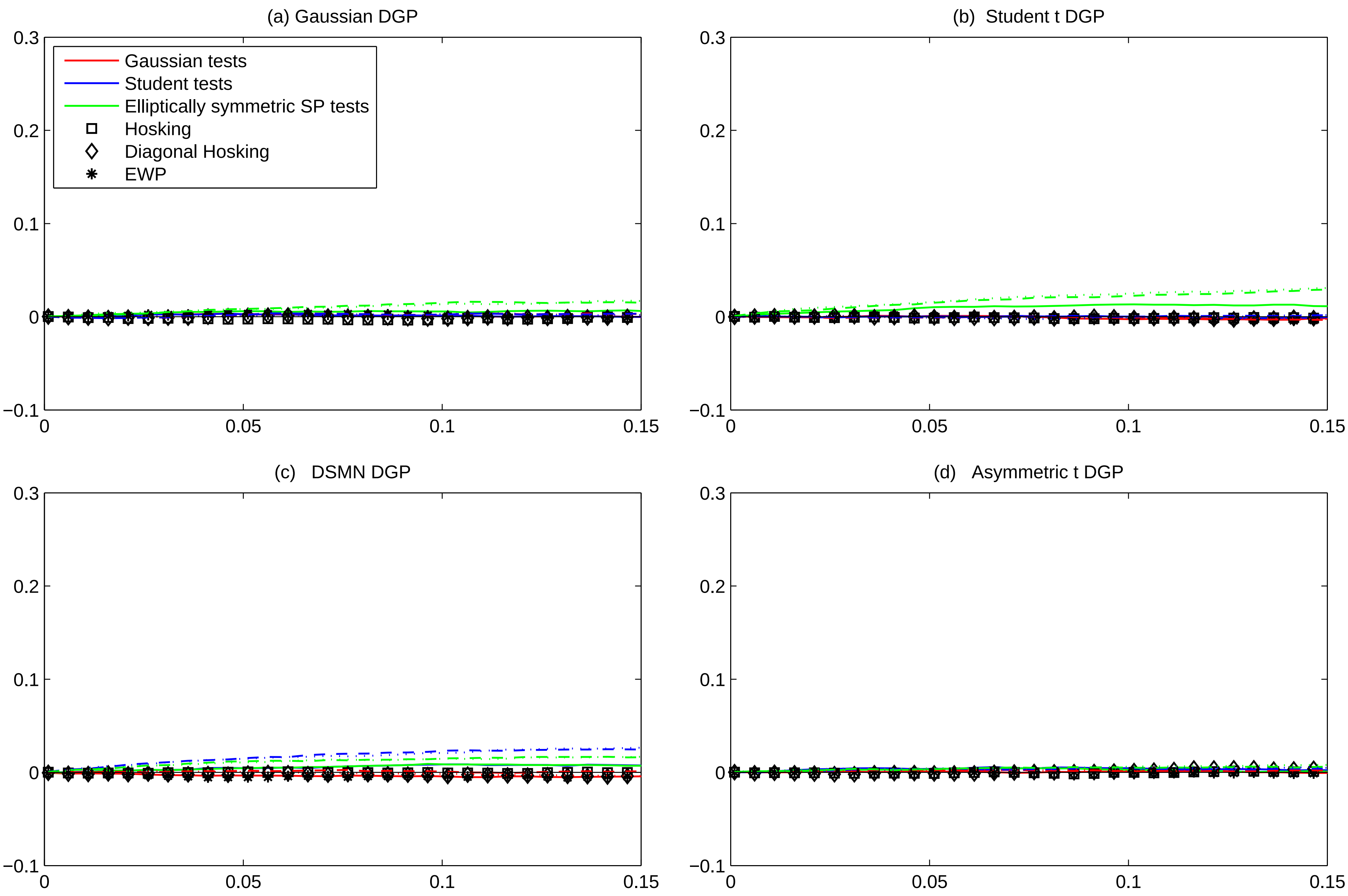


Figure 5: P−value discrepancy plots. Tests against AR(1) alternatives.



___ Common factor dependence tests   − − − Specific factors dependence tests   ... Joint dependence tests

Figure 6: P−value discrepancy plots. Tests against ARCH(1) alternatives.

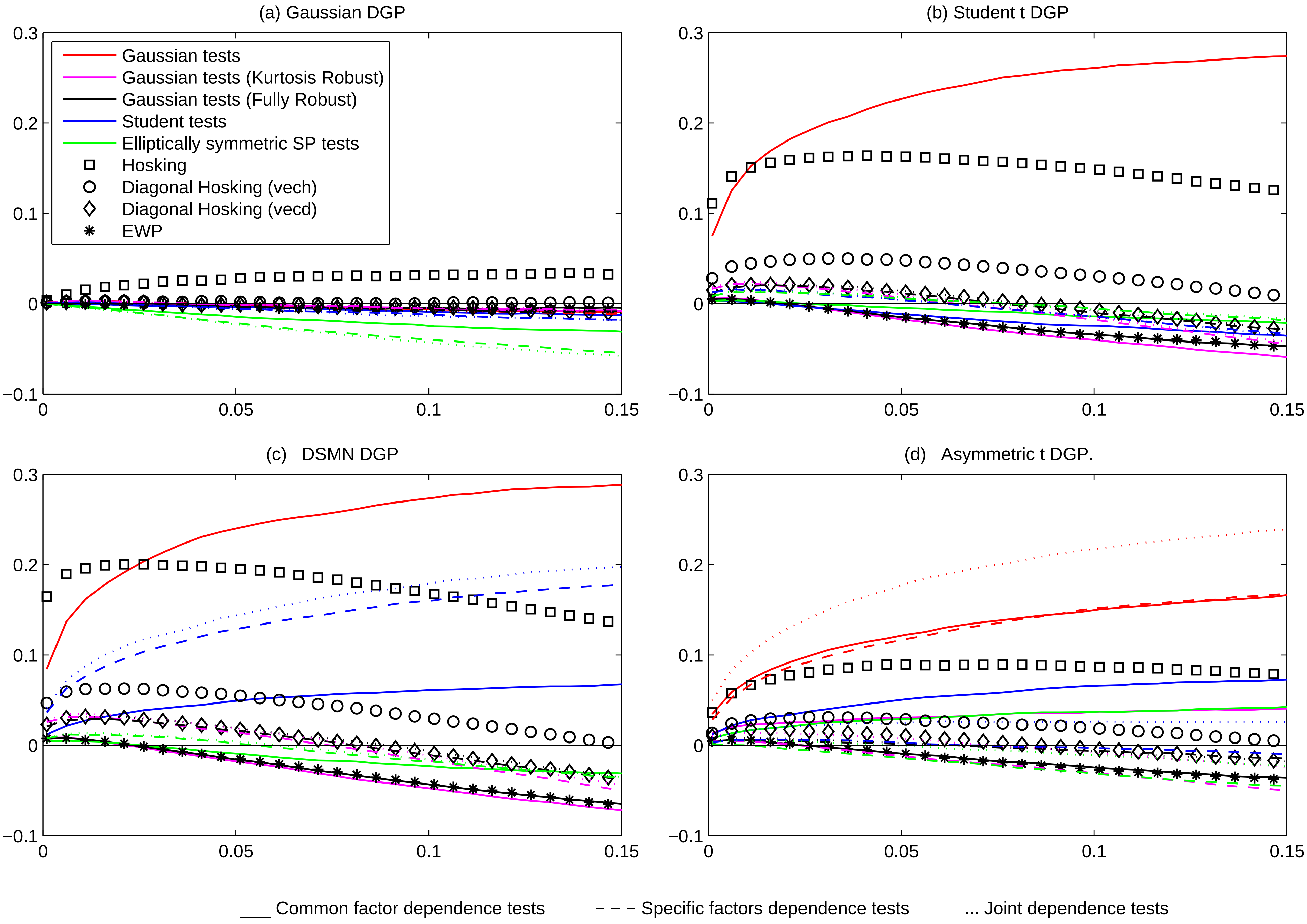


Figure 7: P−value discrepancy plots. Tests against GARCH(1,1) alternatives.