THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 04_10

Gabriele Fiorentini
University of Florence
and The Rimini Centre for Economic Analysis, Italy

Enrique Sentana
CEMFI, Spain

DYNAMIC SPECIFICATION TESTS FOR STATIC FACTOR MODELS

Dynamic specification tests for static factor models*

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

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

December 2009

Abstract

We derive computationally simple 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 derive more powerful tests for elliptically symmetric distributions, which can be either parametrically or semipametrically specified, and robustify the Gaussian tests against general non-normality. 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, Predictability.

JEL: C32, C13, C12, C14, C16

*We are grateful to Dante Amengual, Manuel Arellano, Javier Mencía, Javier Perote and Albert Satorra, as well as to seminar audiences at Cass Business School, CEMFI, Florence, GREQAM, Pablo de Olavide, Toulouse, the 2009 European Meeting of the Econometric Society (Barcelona), the 2008 Finance Forum (Madrid) and the 2009 Symposium on Economic Analysis (Valencia) for helpful comments, discussions and suggestions. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 (Sentana) is gratefully acknowledged.

1 Introduction

There is a long tradition of factor or multi-index models in finance, where they were originally introduced to simplify the computation of the covariance matrix of returns in a mean-variance portfolio allocation framework (see Connor, Goldberg and Korajczyk (2009) for a recent survey). 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 controversial, casts a doubt on the assumption of lack of serial correlation of common and idiosyncratic factors. Second, there is much stronger evidence on time variation in volatilities and correlations at high frequencies such as daily, which is difficult to square with the fairly widespread belief that those effects are irrelevant at monthly and lower frequencies. Finally, many empirical studies with financial time series data indicate that the distribution of asset returns is rather leptokurtic, and possibly somewhat asymmetric. And although it is true that the Gaussian pseudo-maximum likelihood (PML) estimators remain consistent in those circumstances (see e.g. Bollerslev and Wooldridge (1992)), in principle one could obtain more efficient estimators and test procedures by exploiting this third empirical regularity.

The objective of our paper is to provide joint diagnostic tests for serial dependence in the levels and squares of the common and idiosyncratic factors that exploit the non-normality of asset returns, which empirical researchers could easily apply to test the implicit lack of dynamics in the factor analysis models that they estimate. For that reason, 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.

For pedagogical reasons, though, we proceed in steps. We initially assume that the joint distribution of returns conditional on their past is multivariate normal. Under this assumption, we 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. Then, we explain how to modify those tests so that they reflect the more realistic assumption that the joint distribution of asset returns conditional on their past is elliptically symmetric, which can be either parametrically or semipametrically specified. Elliptical distributions are attractive in this context because they generalise the multivariate normal distribution, while retaining its analytical tractability irrespective of the number of assets. In addition, we also explain how to robustify our Gaussian LM tests when the return distribution is neither Gaussian nor elliptical. 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. In section 2, we derive serial dependence tests under normality, which we robustify in section 3, where we obtain more power versions under ellipticity. A Monte Carlo evaluation of all the different tests can be found in section 4, followed by the empirical application to US sectorial stock returns in section 5. Finally, our conclusions can be found in section 6. Proofs and auxiliary results are gathered in appendices.

2 Serial dependence tests under normality

2.1 Static factor models

To keep the notation to a minimum, we initially consider a single factor version of a traditional (i.e. static, conditionally homoskedastic and exact) factor model, which suffices to illustrate our main results. Extensions to multiple factors are considered in sections 2.2.6 and 2.3.7. Specifically:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}f_t + \mathbf{v}_t, \\ \left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\right) | I_{t-1}, \boldsymbol{\theta}_s \sim N\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array}\right)\right] \end{array}\right\} \quad (1)$$

where $\mathbf{y}_t$ is a $N \times 1$ vector of observable variables with constant conditional mean $\boldsymbol{\pi}$, f_t is an unobserved common factor, whose constant variance we have normalised to 1 to avoid the usual scale indeterminacy, $\mathbf{c}$ is the $N \times 1$ vector of factor loadings, $\mathbf{v}_t$ is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to f_t, $\boldsymbol{\Gamma}$ 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 $t-1$ and $\boldsymbol{\theta}_s = (\boldsymbol{\pi}', \mathbf{c}', \boldsymbol{\gamma}')'$, with $\boldsymbol{\gamma} = vecd(\boldsymbol{\Gamma})$. Our assumptions trivially imply that

$$\begin{aligned} \mathbf{y}_t | I_{t-1}; \boldsymbol{\theta}_s &\sim N[\boldsymbol{\pi}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_s)], \\ \boldsymbol{\Sigma}(\boldsymbol{\theta}_s) &= \mathbf{cc}' + \boldsymbol{\Gamma}, \end{aligned} \tag{2}$$

In subsequent sections we shall derive dynamic diagnostic tests for such a static specification.

A non-trivial advantage of these models is that they automatically guarantee a p.s.d. covariance matrix for $\mathbf{y}_t$. But the most distinctive feature of factor models is that they provide a parsimonious specification of the cross-sectional dependence in the observed variables, 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, hedging and portfolio performance evaluation (see Connor, Goldberg and Korajczyk (2009) for details).

The parameters of interest are usually estimated jointly from the log-likelihood function of the observed variables, which can be recursively computed by means of the Kalman filter.[1] In this framework, we prove in appendix B.1 that the Kalman filter updating equations yield:

$$E\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array} \middle| Y_t; \boldsymbol{\theta}_s \right) = \left[\begin{array}{c} \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \end{array} \right] = \left[\begin{array}{c} f_{kt}(\boldsymbol{\theta}_s) \\ \mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{array} \right], \tag{3}$$

and

$$V\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array} \middle| Y_t; \boldsymbol{\theta}_s \right) = \left[\begin{array}{cc} 1\text{-}\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & -\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \\ -\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}\text{-}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \end{array} \right] = \left[\begin{array}{cc} \omega_k(\boldsymbol{\theta}_s) & -\mathbf{c}\omega_k(\boldsymbol{\theta}_s) \\ -\mathbf{c}\omega_k(\boldsymbol{\theta}_s) & \mathbf{cc}'\omega_k(\boldsymbol{\theta}_s) \end{array} \right], \tag{4}$$

a rank 1 matrix 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 (3) coincide with the coefficients in the theoretical regression of each unobserved variable onto the observed series, while (4) coincides with the 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, $f_{kt}(\boldsymbol{\theta}_s)$, $\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$ and $\omega_k(\boldsymbol{\theta}_s)$ constitute the basic ingredients of our tests.

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

In this context, we can use Theorem 12.1 in Anderson and Rubin (1956) and Theorem 2 in Kano (1983) to formally characterise the asymptotic distribution of the maximum likelihood estimators of the static model parameters as follows:

Proposition 1 *Let $\bar{\boldsymbol{\theta}}_s$ denote the Gaussian maximum likelihood estimators of the parameters that characterise model (1). If 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, and we can uniquely decompose $V(\mathbf{y}_t)$ into $\mathbf{cc}'$ and $\boldsymbol{\Gamma}$, then,

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

where

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

$\mathbf{E}_n$ is the unique $n^2 \times n$ "diagonalisation" matrix which transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}'_n vec(\mathbf{A})$ and $\odot$ denotes the Hadamard product of two matrices of equal orders.

2.2 Tests for serial correlation in common and idiosyncratic factors

2.2.1 Baseline case

In this section we shall develop tests of first order serial correlation in the common and idiosyncratic factors under the maintained assumption that their conditional variances are time-invariant. Extensions to higher order serial correlation, multiple factors and conditionally heteroskedastic ones are developed in sections 2.2.5, 2.2.6 and 2.4, respectively. Specifically, the alternative that we consider is the following conditionally homoskedastic dynamic factor model:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t \\ x_t = \rho x_{t-1} + f_t \\ \mathbf{u}_t = diag(\boldsymbol{\rho}^*)\mathbf{u}_{t-1} + \mathbf{v}_t \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1}, \boldsymbol{\theta} \sim N\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\Gamma} \end{pmatrix}\right] \end{array}\right\} \quad (5)$$

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

As is well known, the most precise way to characterise this model is in the frequency domain. Assuming the stationarity conditions $|\rho| < 1$ and $|\rho_i^*| < 1$ $\forall i$ hold, the spectral density matrix of (5) will be given by

$$\mathbf{g}_{\mathbf{yy}}(\lambda) = \mathbf{cc}' g_{xx}(\lambda) + \mathbf{g}_{\mathbf{uu}}(\lambda), \quad (6)$$

which shares the single factor structure of (2) at all frequencies λ. For our purposes, though, it is more interesting to look at the autocovariance matrices of the observed series, which can be trivially obtained from the inverse Fourier transform of the previous expression:

$$\mathbf{G}_{\mathbf{yy}}(j) = \mathbf{cc}'G_{xx}(j) + \mathbf{G}_{\mathbf{uu}}(j). \tag{7}$$

In particular, even though x_t or $\mathbf{u}_t$ are serially correlated, the unconditional covariance matrix of $\mathbf{y}_t$, $\boldsymbol{\Sigma}$ say, can also be written as:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = V(\mathbf{y}_t|\boldsymbol{\theta}) = \mathbf{cc}'G_{xx}(0) + \mathbf{G}_{\mathbf{uu}}(0)$$

(see Doz and Lenglart (1999)). Similarly, it is straightforward to obtain the autocorrelation structure of any linear combination of $\mathbf{y}_t$, $\mathbf{w}'\mathbf{y}_t$ say, by exploiting the fact that its j^{th} autocovariance will be given by $\mathbf{w}'\mathbf{G}_{\mathbf{yy}}(j)\mathbf{w}$ (see also Lütkepohl (1993)). In fact, it is easy to see that the autocovariance structure in (7) corresponds to a special case of a VARMA(2,1) model since

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

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

As the next proposition shows, however, optimally testing the null of multivariate white noise against such a complex VARMA(2,1) specification is extremely easy:

Proposition 2 *Let*

$$\bar{G}_{f_k f_k}(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T} f_{kt}(\boldsymbol{\theta}_s) f_{kt-j}(\boldsymbol{\theta}_s)$$

and

$$\bar{\mathbf{G}}_{\mathbf{v}_k\mathbf{v}_k}(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt-j}(\boldsymbol{\theta}_s)$$

denote the j^{th} sample autocovariances of the Kalman filter estimators of the common and specific factors of model (1), whose expressions are given in (3).

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

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

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

$$\mathcal{I}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}) = \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}) \odot \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}),$$
$$\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}) = V\begin{bmatrix} f_{kt}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\end{bmatrix} = \begin{bmatrix} \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\end{bmatrix}.$$

2. *This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_{s0}$ by its Gaussian maximum likelihood estimator $\bar{\boldsymbol{\theta}}_s$ in Proposition 1.*

Intuitively, we can interpret $LM_{AR(1)}$ as a test based on the $N+1$ orthogonality conditions:

$$E\left[f_{kt}(\boldsymbol{\theta}_s)f_{kt-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0}\right] = 0, \tag{8}$$

$$E\left[\gamma_i^{-1}v_{kit}(\boldsymbol{\theta}_s)v_{kit-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0}\right] = 0 \qquad (i=1,\ldots,N), \tag{9}$$

which are the conditions that we would use to test for first order serial correlation if we treated $f_{kt}(\boldsymbol{\theta}_s)$ or $v_{kit}(\boldsymbol{\theta}_s)$ are the series of interest (see e.g. Breusch and Pagan (1980) or Godfrey (1988)). Given that we have fixed the variance of the innovations in the common factor to 1, these moment conditions closely resemble

$$E(f_t f_{t-1}|\boldsymbol{\theta}_s,\mathbf{0}) = 0,$$

$$E(\gamma_i^{-1}v_{it}v_{it-1}|\boldsymbol{\theta}_s,\mathbf{0}) = 0 \qquad (i=1,\ldots,N),$$

which are the orthogonality conditions that we would 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 $\boldsymbol{\rho}^*=\mathbf{0}$. Proposition 2 implies that the asymptotic variance of $\bar{G}_{f_kf_k}(1)$ is simply $[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2$. But we can use (4) 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 $\bar{G}_{f_kf_k}(1)$ 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 test of $H_0:\rho=0$ will become arbitrarily close to a 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 $\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}=(\mathbf{c}'\Gamma^{-1}\mathbf{c})/[1+(\mathbf{c}'\Gamma^{-1}\mathbf{c})]$ under the assumption that Γ has full rank, the aforementioned R^2 converges to 1 as $N\to\infty$ because $(\mathbf{c}'\Gamma^{-1}\mathbf{c})\to\infty$ in those circumstances due to the pervasive nature of the common factor (see e.g. Sentana (2004)).

Proposition 2 also implies that the asymptotic variance of $\bar{G}_{v_{ki}v_{ki}}(1)$ 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 (4) 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 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 $\sigma^{ii}(\boldsymbol{\theta}_s)=\gamma_i^{-1}-\gamma_i^{-2}c_i^2/[1+(\mathbf{c}'\Gamma^{-1}\mathbf{c})]$ when Γ has full rank, which means that $\gamma_i\sigma^{ii}(\boldsymbol{\theta}_s)$ also converges to 1 as $N\to\infty$ for fixed c_i and γ_i.

Nevertheless, it is important to emphasise that our joint tests also 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, $\mathcal{V}_{\boldsymbol{\rho}^{\dagger}\boldsymbol{\rho}^{\dagger}}(\boldsymbol{\theta}, \mathbf{0}; \mathbf{0})$ 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 application of the multivariate serial correlation test discussed in the next section.

Since the orthogonality conditions (8) and (9) remain valid when $\mathbf{y}_t$ is serially uncorrelated irrespective of $V(\mathbf{y}_t)$ having an exact single factor structure, 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).

2.2.2 Moving average processes

Specification (5) 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

$$\begin{array}{l} x_t = f_t + \varphi f_{t-1}, \\ \mathbf{u}_t = \mathbf{v}_t + diag(\boldsymbol{\varphi}^*)\mathbf{v}_{t-1}. \end{array} \tag{10}$$

Although the single factor structure of the spectral density matrix (6) remains valid, in this case the autocorrelation structure of $\mathbf{y}_t$ corresponds to a restricted VMA(1) process. Therefore, the Kalman filter recursions for this dynamic model are different from the recursions in Appendix B.2. Nevertheless, straightforward algebra shows 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 (5) evaluated at $\boldsymbol{\rho}^{\dagger} = \mathbf{0}$. Hence, we can also interpret $LM_{AR(1)}$ 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)).

2.2.3 Alternative multivariate serial correlation tests

It is illustrative to compare our test of serial correlation in common and specific factors to the multivariate generalisation of the Box and Pierce (1970) test proposed by Hosking (1981). 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). More formally:

Proposition 3 *Consider the following conditionally homoskedastic* VAR*(1) model:*

$$\left.\begin{array}{c} \mathbf{y}_t = (\mathbf{I}_N - \mathbf{P})\boldsymbol{\pi} + \mathbf{P}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t | I_{t-1}, \boldsymbol{\theta}_0 \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \end{array}\right\}, \tag{11}$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}', \mathbf{p}', \boldsymbol{\sigma}')$, *with* $\mathbf{p} = vec(\mathbf{P})$ *and* $\boldsymbol{\sigma} = vech(\boldsymbol{\Sigma})$. *Let*

$$\bar{\boldsymbol{\pi}} = \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{y}_t$$

denote the sample mean of $\mathbf{y}_t$, *and*

$$\bar{\mathbf{G}}_{\mathbf{yy}}(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T}(\mathbf{y}_t - \bar{\boldsymbol{\pi}})(\mathbf{y}_{t-j} - \bar{\boldsymbol{\pi}})'$$

its j^{th} *sample autocovariance matrix. Under the null hypothesis* $H_0 : \mathbf{p} = \mathbf{0}$ *the test statistic*

$$LM_H = Tvec'[\bar{\mathbf{G}}_{\mathbf{yy}}(1)][\bar{\mathbf{G}}_{\mathbf{yy}}^{-1}(0) \otimes \bar{\mathbf{G}}_{\mathbf{yy}}^{-1}(0)]vec[\bar{\mathbf{G}}_{\mathbf{yy}}(1)], \quad (12)$$

will be distributed as a χ^2 *with* N^2 *degrees of freedom for* N *fixed as* T *goes to infinity.*

Apart from the fact that it does not exploit the strong cross-sectional dependence of returns, which results in the number of degrees of freedom being an order of magnitude larger, with the consequent reduction in power, the main problem with this test is that in practice it requires T much larger than N^2 for the asymptotic distribution in Proposition 3 to be reliable in finite samples. In contrast, our joint test only requires that $N/T \to 0$, while our test of $H_0 : \rho = 0$ should remain valid as long as we can consistently estimate the static model parameters.

2.2.4 The relative power of AR tests in multivariate contexts

We compare the power of our proposed LM tests, Hosking's test, 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 ($H_0 : vecd[\mathbf{G}_{\mathbf{yy}}(1)] = \mathbf{0}$), which takes into account that the $y_{it}'s$ are contemporaneously correlated even when they are serially uncorrelated.[2] Note that our joint LM test can also be understood as test of univariate first-order autocorrelation in $[f_{kt}(\boldsymbol{\theta}_s), \mathbf{v}_{kt}'(\boldsymbol{\theta}_s)]$. We consider a non-exchangeable single factor model of the form:

$$\begin{array}{ll} y_{it} = \pi_i + c_i x_t + u_{it} & (i = 1, \ldots, 5) \\ x_t = \rho x_{t-1} + \sqrt{1-\rho^2} f_t & \\ u_{it} = \rho_i^* u_{it-1} + \sqrt{1-\rho_i^{*2}} v_{it} & \end{array}$$

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 5. We evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\boldsymbol{\rho}_{0T}^{\dagger} = \bar{\boldsymbol{\rho}}^{\dagger}/\sqrt{T}$ (see appendix C for details).

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

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

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 $\mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c}$, which coincides with the R^2 in the theoretical least squares projection of f_t on a constant and $\mathbf{y}_t$, is .95.[3] 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.[4]

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 pormanteau test of $H_0 : \mathbf{p} = \mathbf{0}$, the univariate test applied to EWP and finally the test of serial correlation in the common factor, with the "diagonal" serial correlation test of $H_0 : vecd[\mathbf{G_{yy}}(1)] = \mathbf{0}$ somewhere in between. However, this ranking crucially depends on the "signal to noise" ratio $\mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c}$. 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 decreases, while the power of 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. Not surprisingly, the power of the two univariate tests almost coincides because EWP and $f_{kt}(\boldsymbol{\theta}_0)$ become very highly correlated, while the diagonal test is now 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 equal to size, while exactly the opposite happens in the second case.[5]

2.2.5 Higher order serial correlation

Consider the following alternative:

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

[3] 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.

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

[5] The case $\rho = \rho^*$ is rather unsual, in that the reduced form process for the observed series $\mathbf{y}_t$ 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 $\rho = \rho^*$.

so that model (5) corresponds to $h = h_1^* = \ldots = h_N^* = 1$. In view of the discussion in section 2.2.1, it is perhaps not surprising that the score test of $\rho_l = 0$ will be based on the condition

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

while the score test of $\rho_{il}^* = 0$ will be based on

$$E\left[\gamma_i^{-1}v_{kit}(\boldsymbol{\theta}_s)v_{kit-l}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}\right] = 0.$$

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

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

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 instance, imagine that we wanted to test the null against the following novel AR(h) specification for the common factor that we consider in our empirical application:

$$x_t = \sum\nolimits_{l=1}^{h} \rho x_{t-l} + f_t.$$

Then, it is easy to prove that the relevant orthogonality condition 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.

2.2.6 Multiple factor models

So far, we have worked with single factor models to convey the basic intuition while keeping the algebra to a minimum. Nevertheless, it is straightforward to extend our results to models with more than one common factor, which under the null become

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{C}\mathbf{f}_t + \mathbf{v}_t, \\ \left(\begin{array}{c} \mathbf{f}_t \\ \mathbf{v}_t \end{array}\right)|I_{t-1}, \boldsymbol{\theta} \sim N\left[\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array}\right)\right] \end{array}\right\}, \qquad (13)$$

where $\mathbf{f}_t$ is a vector of k unobserved common factors, whose constant covariance matrix we have normalised to the identity matrix, and $\mathbf{C}$ is the corresponding $N \times k$ matrix of factor loadings. In this case, our assumptions trivially imply that

$$\begin{aligned} \mathbf{y}_t | I_{t-1}; \boldsymbol{\theta}_s &\sim N[\boldsymbol{\pi}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_s)], \\ \boldsymbol{\Sigma}(\boldsymbol{\theta}_s) &= \mathbf{C}\mathbf{C}' + \boldsymbol{\Gamma}, \end{aligned}$$

Apart from messier algebraic expressions, the main complication arises from the non-identified nature of the model under the null due to two different issues: (i) the potentially non-unique decomposition of $V(\mathbf{y}_t)$ into a diagonal matrix $\boldsymbol{\Gamma}$ and a reduced rank matrix $\mathbf{C}\mathbf{C}'$, which is related to the so-called Ledermann bound, and (ii) the underidentifiability of $\mathbf{C}$ from $\mathbf{C}\mathbf{C}'$ (see Anderson and Rubin (1956), Dunn (1973), Jennrich (1978), Bekker (1989) or Wegge (1996)). In this sense, it is well known that we can obtain an observationally equivalent model by premultiplying the common factors by an orthogonal matrix of order k, $\mathbf{Q}$ say, and postmultiplying the factor loading matrix by the transpose of this matrix since the unconditional covariance matrix,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_s) = \mathbf{C}^*\mathbf{C}^{*\prime} + \boldsymbol{\Gamma} = \mathbf{C}\mathbf{Q}'\mathbf{Q}\mathbf{C}' + \boldsymbol{\Gamma}$$

remains unchanged. For that reason, empirical researchers often impose a priori restrictions on the matrix $\mathbf{C}$ so that it can be identified (up to permutations and sign changes) from the unconditional covariance matrix of $\mathbf{y}_t$. Although those restrictions are often arbitrary, the factors can be orthogonally rotated to simplify their interpretation once the model has been estimated. In some other cases, identifiability can be achieved by imposing plausible a priori restrictions. For example, if in a two factor model it is believed that the second factor only affects a subset of the variables (say the first N_1, with $N_1 < N$, so that $c_{i2} = 0$ for $i = N_1 + 1, \ldots, N$), then the non-zero elements of $\mathbf{C}$ will always be identifiable. In what follows, we assume that enough restrictions have been imposed to render $\mathbf{C}$ identifiable from knowledge of the unconditional covariance matrix of the observed variables.

Since our main concern in this section is the existence of multiple common factors, to keep the algebra simple the alternative hypothesis that we will consider is as follows:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \\ \mathbf{x}_t = \mathbf{R}\mathbf{x}_{t-1} + \mathbf{f}_t \\ \begin{pmatrix} \mathbf{f}_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1}, \boldsymbol{\theta} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix}\right] \end{array}\right\} \tag{14}$$

which reduces to specification (13) under the null hypothesis that $H_0 : \boldsymbol{\rho} = \mathbf{0}$, where $\boldsymbol{\rho} = vec(\mathbf{R})$. Importantly, it is easy to show that without further restrictions this model will be identified if only if $\mathbf{C}$ can be identified from the static model. We can then prove the following result:

Proposition 4 *Let*

$$\bar{\mathbf{G}}_{\mathbf{f}_k\mathbf{f}_k}(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{f}_{kt}(\boldsymbol{\theta}_s)\mathbf{f}'_{kt-j}(\boldsymbol{\theta}_s)$$

denote the j^{th} *sample autocovariances of the Kalman filter estimators of the common factors of model (13), which are given by*

$$\mathbf{f}_{kt}(\boldsymbol{\theta}_s) = \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}).$$

1. *Under the null hypothesis* $H_0 : \boldsymbol{\rho} = \mathbf{0}$*, the score test statistic*

$$LM_{FVAR(1)} = T \cdot vec'[\bar{\mathbf{G}}_{\mathbf{f}_k\mathbf{f}_k}(1)]\mathcal{I}^{-1}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}_{s0}, \mathbf{0}; \mathbf{0})vec[\bar{\mathbf{G}}_{\mathbf{f}_k\mathbf{f}_k}(1)],$$

will be distributed as a χ^2 *with* k^2 *degrees of freedom as* T *goes to infinity, where*

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}) &= \mathcal{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}) \otimes \mathcal{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}), \\ \mathcal{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}) &= V\left[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\right] = \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}. \end{aligned} \tag{15}$$

2. *This asymptotic null distribution is unaffected if we replace* $\boldsymbol{\theta}_{s0}$ *by its Gaussian maximum likelihood estimator under the null.*

It is easy to see that $LM_{FVAR(1)}$ is numerically invariant to orthogonal rotations of the common factors, so the test result will not depend on the exact identification restriction imposed.

Not surprisingly, this test can be related to a Hosking test applied to the common factors $\mathbf{f}_t$ if they were observed. Unlike the test described in Proposition 3, though, the number of degrees of freedom is k^2 instead of N^2, which still makes a noticeable difference since k is typically much lower than N in practice. Finally, we can easily derive tests for univariate serial correlation in any particular common factor by focusing on the appropriate diagonal element of the autocovariance matrix $\bar{\mathbf{G}}_{\mathbf{f}_k\mathbf{f}_k}(j)$ and the corresponding element of (15).

2.3 Tests for ARCH effects in common and idiosyncratic factors

2.3.1 Baseline case

In this section we shall develop tests of first order ARCH effects in the common and idiosyncratic factors under the maintained assumption that their conditional means are 0. Extensions to higher order effects, multiple factors and serially correlated ones are considered in sections 2.3.5, 2.3.7 and 2.4, respectively. Specifically, the alternative that we consider is the following conditionally heteroskedastic factor model:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}f_t + \mathbf{v}_t, \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1}; \boldsymbol{\theta} \sim N\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}\right], \\ V(f_t|I_{t-1}; \boldsymbol{\theta}) = \lambda_t(\boldsymbol{\theta}) = 1 + \alpha[E(f^2_{t-1}|Y_{t-1}; \boldsymbol{\theta}) - 1], \\ V(v_{it}|I_{t-1}; \boldsymbol{\theta}) = \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha^*_i[E(v^2_{it-1}|Y_{t-1}; \boldsymbol{\theta}) - \gamma_i], \qquad (i = 1, \ldots, N) \end{array}\right\}, \tag{16}$$

where $E(f^2_{t-1}|Y_{t-j}; \boldsymbol{\theta})$ and $E(v^2_{it-j}|Y_{t-1}; \boldsymbol{\theta})$ are the conditionally linear Kalman filter estimators of the squares of the underlying common and idiosyncratic factors obtained from this model (see appendix B.3). Although it is in principle 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 $Y_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$, 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 $\lambda_t(\boldsymbol{\theta}) \notin Y_{t-1}$ until section 2.3.2.

Given (16), the distribution of $\mathbf{y}_t$ conditional on Y_{t-1} is $N(\mathbf{0}, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\Sigma}_t = \mathbf{cc}'\lambda_t + \Gamma_t$ has the usual exact factor structure. For this reason, we shall refer to the data generation process specified by (16) as a multivariate conditionally heteroskedastic exact factor model, which reduces to our baseline specification (1) under the null hypothesis that $H_0 : \boldsymbol{\alpha}^{\dagger} = \mathbf{0}$, where $\boldsymbol{\alpha}^{\dagger} = (\alpha, \boldsymbol{\alpha}^*)$ and $\boldsymbol{\alpha}^* = (\alpha_1, \ldots, \alpha_N)$. But even if even if f_t or $\mathbf{v}_t$ are conditionally heteroskedastic, provided that they are covariance stationary, model (16) also implies an unconditional exact factor structure for $\mathbf{y}_t$. That is, the unconditional covariance matrix, $\boldsymbol{\Sigma}$, can be written as:

$$\boldsymbol{\Sigma} = E(\boldsymbol{\Sigma}_t | \boldsymbol{\theta}) = \mathbf{cc}' + \boldsymbol{\Gamma}, \tag{17}$$

because we have set the unconditional variance of the common factor to 1 to eliminate the usual scale indeterminacy.[6] In this case, the parameters of interest become $\boldsymbol{\theta} = (\boldsymbol{\theta}_s', \boldsymbol{\alpha}^{\dagger\prime})'$.

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 (16) 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 (16) 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 (7) not for the levels but for the squares and cross-products of the observed variables $\mathbf{y}_t$ (see appendix C for further details).

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

In this sense, another empirically appealing feature of (16) 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 hypothesis that $\boldsymbol{\alpha}^* = \mathbf{0}$. The following proposition extends their results to joint tests of ARCH effects in common and specific factors:

Proposition 5 *Let*

$$\bar{S}_{f_k f_k}(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T}[f_{kt}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1][f_{kt-j}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1]$$

and

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

denote the sample autocovariances of the squares of the Kalman filter estimators of the innovations in the common and specific factors of model (1), whose expressions are given in (3).

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

$$\tfrac{T}{4}\left(\bar{S}_{f_k f_k}(1), vecd'[\boldsymbol{\Gamma}^{-1}\bar{\mathbf{S}}_{\mathbf{v}_k\mathbf{v}_k}(1)\boldsymbol{\Gamma}^{-1}]\right)\mathcal{J}^{-1}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0})\left(\bar{S}_{f_k f_k}(1), vecd'[\boldsymbol{\Gamma}^{-1}\bar{\mathbf{S}}_{\mathbf{v}_k\mathbf{v}_k}(1)\boldsymbol{\Gamma}^{-1}]\right)',$$

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

$$\mathcal{J}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}) = \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}) \odot \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}), \qquad (18)$$

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

$$= \begin{bmatrix} [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \odot \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \odot \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \odot \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\end{bmatrix}.$$

2. *This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_{s0}$ by its Gaussian maximum likelihood estimator $\bar{\boldsymbol{\theta}}_s$ in Proposition 1.*

Intuitively, we can interpret $LM_{ARCH(1)}$ as a test based on the $N+1$ orthogonality conditions:

$$E\left\{[f_{kt}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1][f_{kt-1}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1]|\boldsymbol{\theta}_s, \mathbf{0}\right\} = 0, \qquad (19)$$

$$E\left\{\gamma_i^{-2}[v_{kit}^2(\boldsymbol{\theta}_s) + c_i^2\omega_k(\boldsymbol{\theta}_s) - \gamma_i][v_{kit-1}^2(\boldsymbol{\theta}_s) + c_i^2\omega_k(\boldsymbol{\theta}_s) - \gamma_i]|\boldsymbol{\theta}_s, \mathbf{0}\right\} = 0 \ (i = 1, \ldots, N) \qquad (20)$$

which are the orthogonality conditions that we would use to test for first order ARCH effects if we treated $f_{kt}(\boldsymbol{\theta}_s)$ or $v_{kit}(\boldsymbol{\theta}_s)$ as the series of interest (see e.g. Engle (1982)). Once again, given that we normalise $V(f_t)$ to 1, these moment conditions closely resemble

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

which are the orthogonality conditions that we would use to test for first order ARCH effects if we could observe the latent variables.

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 $\boldsymbol{\alpha}^* = \mathbf{0}$. Proposition 5 implies that the asymptotic variance of $\bar{S}_{f_k f_k}(1)$ is simply $2[\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^4$. But as we saw in section 2.2.1, 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 $\bar{S}_{f_k f_k}(1)$ will be to 2, which is the asymptotic variance of the first sample autocovariance of f_t^2. 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.

Proposition 5 also implies that the asymptotic variance of $\bar{S}_{v_{ki} v_{ki}}(1)$ is $2[\gamma_i \sigma^{ii}(\boldsymbol{\theta}_s)]^4$, 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 $\rho = 0$ and the remaining elements of $\boldsymbol{\rho}^*$ 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.

Again, it would be straightforward to adapt Proposition 5 to handle large N panel data restrictions such as $\alpha_i^* = \alpha^* \; \forall i$, as in Sentana, Calzolari and Fiorentini (2008). Given that the orthogonality conditions (19) and (20) 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.

2.3.2 Unobservable conditional variances

Specification (16) 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, we could assume that

$$V(f_t|I_{t-1}; \boldsymbol{\theta}_0) = 1 + \alpha(f_{t-j}^2 - 1), \tag{21}$$

$$V(v_{it}|I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \gamma_i + \alpha_i^*(v_{it-1}^2 - \gamma_i), \qquad (i = 1, \ldots, N). \tag{22}$$

The problem with this formulation is that the log-likelihood function can no longer be written in closed form except under the null hypothesis $\boldsymbol{\alpha}^\dagger = \mathbf{0}$. For our purposes, though, the nonmeasurability of λ_t and $\mathbf{\Gamma}_t$ is inconsequential because Proposition 1 in Sentana, Calzolari and Fiorentini (2008) shows that not only the log-likelihood function but also the score of this modified model coincides the score of model (16) under the null of conditionally homoskedasticity. Therefore, $LM_{ARCH(1)}$ can be interpreted as LM tests of $H_0 : \boldsymbol{\alpha}^{\dagger} = \mathbf{0}$ in this context too.

2.3.3 Alternative multivariate ARCH tests

It is again illustrative to compare our tests of ARCH effects in the latent factors to Hosking-style general multivariate ARCH test of the type discussed by Duchesne and Lalancette (2003):

Proposition 6 *Consider the following vech specification of the multivariate* ARCH*(1) model:*

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t | I_{t-1}, \boldsymbol{\theta}_0 \sim N(\mathbf{0}, \mathbf{\Sigma}_t) \\ vech(\mathbf{\Sigma}_t) = vech(\mathbf{\Sigma}) + \mathbf{A} vech(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} - \mathbf{\Sigma}). \end{array}\right\} \quad (23)$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}', \boldsymbol{\sigma}', \mathbf{a}')$, *with* $\boldsymbol{\sigma} = vech(\mathbf{\Sigma})$ *and* $\mathbf{a} = vec(\mathbf{A})$*. Let*

$$\bar{\boldsymbol{\pi}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t$$

denote the sample mean of $\mathbf{y}_t$*,*

$$\bar{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{y}_t - \bar{\boldsymbol{\pi}})(\mathbf{y}_{t-j} - \bar{\boldsymbol{\pi}})',$$

its sample covariance matrix, and

$$\bar{\mathbf{S}}_{\mathbf{yy}}(j) = \frac{1}{T}\sum_{t=1}^{T} vech[(\mathbf{y}_t - \bar{\boldsymbol{\pi}})(\mathbf{y}_t - \bar{\boldsymbol{\pi}})' - \bar{\mathbf{\Sigma}}] vech'[(\mathbf{y}_{t-j} - \bar{\boldsymbol{\pi}})(\mathbf{y}_{t-j} - \bar{\boldsymbol{\pi}})' - \bar{\mathbf{\Sigma}}]$$

the j^{th} *sample autocovariance matrix of* $vech[(\mathbf{y}_t - \bar{\boldsymbol{\pi}})(\mathbf{y}_t - \bar{\boldsymbol{\pi}})']$*. Under the null hypothesis that* $H_0 : \mathbf{a} = \mathbf{0}$ *the asymptotic distribution of the test statistic*

$$LM_{VECH(1)} = \frac{T}{4} vec'[\bar{\mathbf{S}}_{\mathbf{yy}}(1)]\{[\mathbf{D}'_N(\bar{\mathbf{\Sigma}}^{-1} \otimes \bar{\mathbf{\Sigma}}^{-1})\mathbf{D}_N] \otimes [\mathbf{D}'_N(\bar{\mathbf{\Sigma}}^{-1} \otimes \bar{\mathbf{\Sigma}}^{-1})\mathbf{D}_N]\} vec[\bar{\mathbf{S}}_{\mathbf{yy}}(1)], \quad (24)$$

will be a χ^2 *with* $N^2(N+1)^2/4$ *degrees of freedom for* N *fixed as* T *goes to infinity, where* $\mathbf{D}_N$ *is the duplication matrix of order* N*.*

Apart from the fact that it does not exploit the strong cross-sectional dependence of returns, which results in the number of degrees of freedom being three orders of magnitude larger, with the consequent reduction in power, the main problem with (24) is that in practice it requires T much larger than N^4 for the asymptotic distribution in Proposition 6 to be reliable in finite samples. In contrast, our joint test only requires that $N/T \to 0$, while our test of $H_0 : \alpha = 0$ should remain valid as long as we can consistently estimate the model parameters.[7]

[7] Another implication of the single factor structure of $\mathbf{\Sigma}$ is that $LM_{VECH(1)}$ differs from the multivariate ARCH test considered by Dufour, Khalaf and Beaulieu (2008), 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 $\bar{\mathbf{\Sigma}}$.

2.3.4 The relative power of ARCH tests in multivariate contexts

We compare the power of our LM tests, Hosking's test applied to $vech[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})']$, a standard univariate ARCH(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 $[f_{kt}(\boldsymbol{\theta}_s), \mathbf{v}'_{kt}(\boldsymbol{\theta}_s)]$. 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 2.2.4. We evaluate power against *compatible* sequences of local alternatives of the form $\boldsymbol{\alpha}_{0T}^{\dagger} = \bar{\boldsymbol{\alpha}}^{\dagger}/\sqrt{T}$ (see appendix C for details).

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 $H_0 : \mathbf{a} = \mathbf{0}$, which suffers from having a 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 serial correlation in the common factor becomes the most powerful, with the test of serial correlation in the specific factors having power equal to size, while exactly the opposite happens in the second case.

2.3.5 Higher order ARCH effects

Consider the following alternative specification:

$$\begin{aligned}V(f_t|I_{t-1};\boldsymbol{\theta}_0) &= \lambda_t(\boldsymbol{\theta}) = 1+\sum\nolimits_{j=1}^{q}\alpha_j[E(f_{t-j}^2|Y_{t-j};\boldsymbol{\theta})-1],\\ V(v_{it}|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) &= \gamma_{it}(\boldsymbol{\theta})=\gamma_i+\sum\nolimits_{j=1}^{q_i^*}\alpha_{ij}^*[E(v_{it-j}^2|Y_{t-j};\boldsymbol{\theta})-\gamma_i], \qquad (i=1,\ldots,N),\end{aligned}$$

so that model (16) corresponds to $q=q_1^*=\ldots=q_N^*=1$. In view of the discussion in section 2.3.1, it is perhaps not surprising to find out that the score test of $\alpha_j=0$ will be based on the orthogonality condition

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

while the score test of $\alpha_{ij}^*=0$ will be based on

$$E\left\{\gamma_i^{-2}[v_{kit}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i][v_{kit-j}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i]|\boldsymbol{\theta}_s,\mathbf{0}\right\}=0$$

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

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

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.

2.3.6 GARCH tests

The univariate empirical evidence suggests that GARCH(1,1) specifications of the form

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

should be more realistic than unrestricted ARCH(q) ones. As Bollerslev (1986) noted in a univariate context, however, one cannot derive a score test for conditional homoskedasticity versus these GARCH(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

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

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

whose asymptotic covariance matrix would be

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

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}.

2.3.7 Multiple factor models

As in section 2.2.6, we assume that enough restrictions have been imposed to render $\mathbf{C}$ identifiable from knowledge of the unconditional covariance matrix of the observed variables. Since our main concern in this section is the existence of multiple common factors, to keep the algebra simple the alternative hypothesis that we will consider is as follows:

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{C}\mathbf{f}_t + \mathbf{v}_t, \\ \left(\begin{array}{c} \mathbf{f}_t \\ \mathbf{v}_t \end{array}\right)|I_{t-1}; \boldsymbol{\theta} \sim N\left[\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \Lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array}\right)\right], \\ vech[V(\mathbf{f}_t|I_{t-1}; \boldsymbol{\theta})] = vech[\Lambda_t(\boldsymbol{\theta})] = \mathbf{I}_k + \mathbf{A}vech[E(\mathbf{f}_{t-1}\mathbf{f}_{t-1}'|Y_{t-1}; \boldsymbol{\theta}) - \mathbf{I}_k], \end{array}\right\}, \quad (25)$$

which reduces to our baseline specification (13) under the null hypothesis that $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, where $\boldsymbol{\alpha} = vec(\mathbf{A})$. Importantly, it is easy to show that without further restrictions on the matrix $\mathbf{A}$ this model will be identified if and only if $\mathbf{C}$ can be identified from the static model (cf. Sentana and Fiorentini (2001)). We can then prove the following result:

Proposition 7 *Let*

$$\begin{aligned} \bar{\mathbf{S}}_{\mathbf{f}_k\mathbf{f}_k}(j) &= \frac{1}{T}\sum\nolimits_{t=1}^{T} vech[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\mathbf{f}_{kt}'(\boldsymbol{\theta}_s) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}_s) - \mathbf{I}_k] \\ &\quad \times vech'[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\mathbf{f}_{kt}'(\boldsymbol{\theta}_s) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}_s) - \mathbf{I}_k] \end{aligned}$$

denote the sample autocovariances of the squares and cross-products of the Kalman filter estimators of the innovations in the common factors of model (13), where

$$\boldsymbol{\Omega}_k(\boldsymbol{\theta}_s) = \mathbf{I}_k - \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}.$$

1. *Under the null hypothesis* $H_0 : \boldsymbol{\alpha} = \mathbf{0}$*, the score test statistic*

$$\begin{aligned} LM_{FVECH(1)} &= \frac{T}{4} \cdot vec'[\bar{\mathbf{S}}_{\mathbf{f}_k\mathbf{f}_k}(1)]\{\mathbf{D}_k'[(\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C})^{-1} \otimes (\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C})^{-1}]\mathbf{D}_k] \\ & \otimes[\mathbf{D}_k'[(\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C})^{-1} \otimes (\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C})^{-1}]\mathbf{D}_k]\}vec[\bar{\mathbf{S}}_{\mathbf{f}_k\mathbf{f}_k}(1)], \end{aligned}$$

will be distributed as a χ^2 *with* $k^2(k+1)^2/4$ *degrees of freedom as* T *goes to infinity.*

2. *This asymptotic null distribution is unaffected if we replace* $\boldsymbol{\theta}_{s0}$ *by its Gaussian maximum likelihood estimator under the null.*

It is easy to see that $LM_{FVECH(1)}$ is numerically invariant to orthogonal rotations of the common factors, so the test result will not depend on the exact identification restriction imposed.

Not surprisingly, this test can be related to the test discussed in Proposition 6 applied to $vech(\mathbf{f}_t\mathbf{f}_t')$ if the common factors were observed. Unlike the test described in that proposition, though, the number of degrees of freedom is $O(k^4)$ instead of $O(N^4)$, which makes a tremendous difference in practice since k is typically much smaller than N.

Finally, note that Proposition (7) also allows us to derive tests for univariate ARCH effects in any particular common factor by focusing on the corresponding autocovariance. In this sense, our multiple factor test nests the tests proposed in Sentana and Fiorentini (2001), who assumed that the common factors followed conditionally orthogonal univariate ARCH processes instead.

2.4 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,

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t \\ x_t = \rho x_{t-l} + f_t \\ \mathbf{u}_t = diag(\boldsymbol{\rho}^*)\mathbf{u}_{t-1} + \mathbf{v}_t \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1}; \boldsymbol{\theta} \sim N\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}\right], \\ V(f_t|I_{t-1};\boldsymbol{\theta}_0) = \lambda_t(\boldsymbol{\theta}) = 1 + \alpha[E(f_{t-1}^2|Y_{t-1};\boldsymbol{\theta}) - 1], \\ V(v_{it}|I_{t-1};\boldsymbol{\theta}_0) = \gamma_{it}(\boldsymbol{\theta}) = \gamma_i + \alpha_i^*[E(v_{it-1}^2|Y_{t-1};\boldsymbol{\theta}) - \gamma_i], \qquad (i = 1, \ldots, N) \end{array}\right\}. \tag{26}$$

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 (5). 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 (16). Finally, under the null hypothesis of lack of predictability in mean ($\boldsymbol{\rho}^\dagger = \mathbf{0}$) and variance ($\boldsymbol{\alpha}^\dagger = \mathbf{0}$), model (26) 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) is simply the sum of the separate tests:

Proposition 8 *1. Under the joint null hypothesis $H_0 : \boldsymbol{\rho}^{\dagger} = \mathbf{0}, \boldsymbol{\alpha}^{\dagger} = \mathbf{0}$ the score test statistic*

$$LM_{AR(1)-ARCH(1)} = LM_{AR(1)} + LM_{ARCH(1)},$$

will be distributed as a χ^2 with $2(N+1)$ degrees of freedom for N fixed as T goes to infinity.

2. This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_s$ by its Gaussian maximum likelihood estimator $\bar{\boldsymbol{\theta}}_s$ in Proposition 1.

Intuitively, the reason is that the serial correlation orthogonality conditions (8)-(9) are asymptotically orthogonal to the ARCH orthogonality conditions (19)-(20) because all odd order moments of the multivariate normal distribution are 0.

3 Non-normal distributions for returns

As mentioned in the introduction, many empirical studies with financial time series data indicate that the distribution of asset returns is rather leptokurtic. For our purposes, it is important to distinguish between the LM tests that we have obtained under the normality assumption, which we may have to robustify, and the more powerful LM tests that could be obtained by exploiting the non-normality of the conditional distribution.

3.1 Serial dependence tests that exploit ellipticity

We first extend our previous results to the case in which the conditional mean vector and covariance matrix of $\mathbf{y}_t$ is the same as in section 2 (see appendix B), but the conditional distribution is elliptically symmetric. More formally, if we define the standardised innovations

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)[\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)] \quad (27)$$

as a vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$, we shall assume that its conditional distribution is spherical, but not necessarily multivariate normal. If the corresponding density is well defined, then it will be characterised by some additional r parameters $\boldsymbol{\eta}$ that determine the shape of the conditional density of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. The most prominent elliptical example is, of course, the spherical normal distribution, which we assume corresponds to $\boldsymbol{\eta} = \mathbf{0}$. For illustrative purposes, though, we shall also look in some detail at the special case in which $\boldsymbol{\varepsilon}_t^*$ follows a standardised multivariate t with ν_0 degrees of freedom, or *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ for short. As is well known, the multivariate student t approaches the multivariate normal as $\nu_0 \to \infty$, but has generally fatter tails. For that reason, we shall define η as $1/\nu$, which will always remain in the finite range $0 \leq \eta_0 < 1/2$ under our assumptions.

In this case, the scores that we should use to test for serial dependence should correspond to the correct elliptical log-likelihood function. The derivations in the proofs of Proposition 2 and

5 imply that the optimal score tests will be based on the modified moment conditions

$$
\begin{aligned}
E\left[\delta_t(\boldsymbol{\theta},\boldsymbol{\eta}) f_{kt}(\boldsymbol{\theta}_s) f_{kt-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0}\right] &= 0, \\
E\left[\gamma_i^{-1}\delta_t(\boldsymbol{\theta},\boldsymbol{\eta}) v_{kit}(\boldsymbol{\theta}_s) v_{kit-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0}\right] &= 0 \qquad (i=1,\ldots,N),
\end{aligned}
$$

and

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

where

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

and $g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ is the kernel of the elliptical density (see appendix D for details). The factor $\delta_t(\boldsymbol{\theta},\boldsymbol{\eta})$ is equal to 1 under Gaussianity and to $(N\eta+1)/[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})]$ for the Student t, so it can be regarded as a damping factor for big observations because it is a decreasing function of $\varsigma_t(\boldsymbol{\theta})$ for fixed $\eta>0$, the more so the higher η is.

In this context, we show in the proof of Proposition 2 that the asymptotic covariance matrix of the moment conditions corresponding to the serial correlation tests above is

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

where

$$
\begin{aligned}
\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) &= V\left[\begin{array}{c} \delta_t(\boldsymbol{\theta},\boldsymbol{\eta}) f_{kt}(\boldsymbol{\theta}_s) \\ \delta_t(\boldsymbol{\theta},\boldsymbol{\eta})\boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \end{array}\right] = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}), \\
\mathrm{M}_{ll}(\boldsymbol{\eta}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta}_0)/N\right|\boldsymbol{\eta}\},
\end{aligned}
$$

and $\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$ equals $\mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0})$ in Proposition 2.

Similarly, the asymptotic covariance matrix of the orthogonality conditions of the ARCH tests above becomes

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

where

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

is given in (A12) as a function of both $\boldsymbol{\theta}_s$ and

$$
\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2}\left\{1+V\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\theta}_0)/N\right|\boldsymbol{\eta}]\right\},
$$

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 $\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1$ by

$$\kappa = \frac{E(\varsigma_t^2)}{N(N+1)} - 1,$$

which is Mardia's coefficient of multivariate excess kurtosis. In addition, the conditional mean and conditional variance orthogonality conditions are asymptotically independent, which means that the joint test is simply the sum of its two components.

Importantly, we show in the proofs of Propositions 2, 5 and 8 that all these tests remain valid if we replace $\delta[\varsigma_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]$ by either a feasible parametric estimator of $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$ obtained by fitting a specific elliptical distribution to $\mathbf{y}_t$ under the null, or an elliptically symmetric semiparametric estimator of $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ obtained from a nonparametric estimate of the density of $\varsigma_t(\boldsymbol{\theta}_s)$.

3.2 Robustifying the score tests based on normality

Let us now study the effects of general forms of non-normality on the tests derived in section 2. To do so, let us partition the parameter vector $\boldsymbol{\theta}$ as $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. It is well known that a robust Gaussian pseudo score test of the null hypothesis $H_0 : \boldsymbol{\theta}_1 = \mathbf{0}$ can be computed as

$$\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T} \mathbf{s}'_{\boldsymbol{\theta}_1 t}(\mathbf{0}, \tilde{\boldsymbol{\theta}}_2, \mathbf{0})\right] \mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\, \mathcal{C}^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\, \mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) \left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T} \mathbf{s}_{\boldsymbol{\theta}_1 t}(\mathbf{0}, \tilde{\boldsymbol{\theta}}_2, \mathbf{0})\right],$$

where $\mathbf{s}_{\boldsymbol{\theta}_1 t}(\mathbf{0}, \tilde{\boldsymbol{\theta}}_2, \mathbf{0})$ is the Gaussian score evaluated at the restricted PML estimator $\tilde{\boldsymbol{\theta}}_2$, $\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)$ is the relevant block of the inverse of the expected Hessian matrix $\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right]$ and $\mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)$ 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)). But if $\mathcal{A}(\boldsymbol{\phi})$ and $\mathcal{B}(\boldsymbol{\phi})$ are block diagonal between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, then the matrix in the middle simplifies to $\mathcal{B}^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi})$.

Taking $\boldsymbol{\theta}_1$ as $\boldsymbol{\rho}^\dagger = (\rho, \boldsymbol{\rho}^{*\prime})'$, we show in the proof of Proposition 2 that both $\mathcal{A}(\boldsymbol{\phi})$ and $\mathcal{B}(\boldsymbol{\phi})$ are block diagonal with respect to $\boldsymbol{\pi}$, $\boldsymbol{\rho}^\dagger$ and $(\mathbf{c}, \boldsymbol{\gamma})$, with identical blocks for $\boldsymbol{\rho}^\dagger$, when the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ in (27) is *i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ regardless of its sphericity. Further, $\mathcal{B}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi})$ coincides with the expression for $\mathcal{I}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0})$ in Proposition 2. Therefore, it is not necessary to robustify the Gaussian tests for serial correlation that we derived in section 2.2.1.

Effectively, this result mimics the fact that under conditional homoskedasticity, standard score tests for serial correlation in observed series are robust to non-normality in the conditional distribution. 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. For analogous reasons, the multivariate serial correlation test in (12) remains valid regardless of the true distribution of the data.

Similarly, if we take $\boldsymbol{\theta}_1$ as $\boldsymbol{\alpha}^\dagger = (\alpha, \boldsymbol{\alpha}^{*\prime})'$, then we show in the proof of Proposition 5 that $\mathcal{A}(\boldsymbol{\phi})$ and $\mathcal{B}(\boldsymbol{\phi})$ are also block diagonal with respect to $\boldsymbol{\pi}$, $(\mathbf{c}, \boldsymbol{\gamma})$ and $\boldsymbol{\alpha}^\dagger$ irrespective of the distribution of $\boldsymbol{\varepsilon}_t^*$, but the blocks for $\boldsymbol{\alpha}^\dagger$ no longer coincide with $\mathcal{J}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0})$ because

$$\begin{aligned} \mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) &= \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \odot \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0}), \\ \mathcal{B}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi}) &= \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \odot \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}). \end{aligned} \tag{28}$$

Consequently, it is necessary to modify the ARCH tests derived in section 2.3.1 by using (28). As we mentioned in the previous section, $\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho})$ simplifies considerably when $\boldsymbol{\varepsilon}_t^*$ is spherical, which we can exploit to 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 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 $H_0: \alpha = 0$ 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 $V[f_{kt}^2(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}]$ as the residual variance of the regression. 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, then our F test will be incorrectly sized when the conditional distribution is not Gaussian.

It is also worth mentioning that we show in the proof of Proposition 8 that the Gaussian orthogonality conditions corresponding to the conditional mean and conditional variance parameters continue to be asymptotically orthogonal under the sphericity assumption, so that the joint tests can still be obtained as the sum of the two components. This additivity, though, no longer holds for non-spherical distributions, in which case:

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

$$\begin{bmatrix} \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \\ \mathcal{V}'_{\boldsymbol{\rho}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) & \mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \end{bmatrix} = V \begin{bmatrix} f_{kt}(\boldsymbol{\theta}_s) \\ \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_{kt}(\boldsymbol{\theta}_s) \\ \frac{1}{\sqrt{2}}[f_{kt}^2(\boldsymbol{\theta}_s) + \omega_k(\boldsymbol{\theta}_s) - 1] \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1} vecd[\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s) + \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) - \boldsymbol{\Gamma}] \end{bmatrix} \tag{29}$$

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.

Finally, the simplest way to make the test proposed in Proposition 6 robust to any departures from normality is by applying expression ((12) in Proposition 3 to $vech[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})]$ (see the proof of Proposition 6 for a simplified expression in the elliptically symmetric case).

3.3 The relative power of the normality tests

To keep the algebra simple, we shall initially compare the individual tests of $H_0 : \rho = 0$ and $H_0 : \alpha = 0$ under the maintained assumption that all the remaining dynamic parameters are 0. It is not difficult to show that the ratios of non-centrality parameters of the normality tests and elliptical likelihood tests are $\text{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$ for the AR(h) tests, and $4/\{[3\text{M}_{ss}(\boldsymbol{\eta}_0) - 1](3\kappa_0 + 2)\}$ for the individual ARCH(q) tests. In the multivariate student t case with $\nu_0 > 4$, in particular, these asymptotic efficiency ratios become

$$\begin{array}{c} \dfrac{(\nu_0 - 2)(\nu_0 + N + 2)}{\nu_0(\nu_0 + N)} \\ \dfrac{(\nu_0 + N + 2)(\nu_0 - 4)}{(\nu_0 - 1)(\nu_0 + N - 1)}, \end{array} \tag{30}$$

respectively. For any given N, these ratios are monotonically increasing in ν_0, and approach 1 from below as $\nu_0 \to \infty$, and 0 from above as $\nu_0 \to 2^+$ or $\nu_0 \to 4^+$. For instance, for $N = 1$ and $\nu_0 = 9$, they take the value of $.9\mathring{3}$ and $.8\mathring{3}$, respectively, while for $\nu_0 = 5$, their values are only .8 and .4. At the same time, these ratios are decreasing in N for a given ν_0, which reflects the fact that Fisher's information is "increasing" in N. For $\nu_0 = 9$ and $N = 3$, for instance, they take the value of .907 and .795, respectively, while for $\nu_0 = 5$, their values are only .75 and .357. Exactly the same results apply to tests of $H_0 : \rho_i^* = 0$ and $H_0 : \alpha_i^* = 0$.

More generally, we can use the asymptotic distribution of the different estimators of $\boldsymbol{\rho}^\dagger$ and $\boldsymbol{\alpha}^\dagger$ under the null derived in the proofs of Proposition 2 and 5 to obtain the non-centrality parameters of joint tests of $\boldsymbol{\rho}^* = \mathbf{0}$, $\boldsymbol{\alpha}^* = \mathbf{0}$, $\boldsymbol{\rho}^\dagger = \mathbf{0}$ or $\boldsymbol{\alpha}^\dagger = \mathbf{0}$. In the case of the mean parameters, the asymptotic efficiency ratio (30) applies to the joint tests too. In addition, the non-centrality parameters of the Gaussian tests are invariant to the true conditional distribution of the data. In the case of the variance 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 designs in sections 2.2.4 and 2.3.4, respectively, under the assumption that the true conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is a multivariate t_6. Figure 5a shows that the power gains that accrue to our proposed serial correlation tests by exploiting the leptokurtosis of the t distribution are far from trivial. Furthermore, Figure 5b 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.

4 Monte Carlo analysis

4.1 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 2.2.4 and 2.3.4, and is thereby adapted to the empirical application in section 5. 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 considered in Proposition 6 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.[8] All the examples of the DGP in (26) considered can be written as nonexchangeable single factor models of the form:

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

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 Mencia and Sentana (2009a,b) 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

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

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

where ε_t° 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 (2009b), we generate a standardised asymmetric multivariate t by choosing

$$\varepsilon_t^* = \boldsymbol{\beta}\left[\xi_t^{-1} - c(\boldsymbol{\beta},\eta)\right] + \sqrt{\frac{\zeta_t}{\xi_t}}\boldsymbol{\Xi}^{1/2}\varepsilon_t^\circ, \tag{31}$$

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.[9] 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 in Appendix B.1 and the $\mathcal{A}(\boldsymbol{\phi}_0)$ matrix 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

[9] 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.

of which we obtain the sequential Method of Moments estimator of η suggested by Fiorentini, Sentana and Calzolari (2004), 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 a sequential ML estimator of η, keeping $\boldsymbol{\pi}$, $\mathbf{c}$ 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 (2007). 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).

4.2 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 autocorrelation in all N series introduced in section 2.2.4 are summarised in Figures 6a-6d 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 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 17 in Fiorentini and Sentana (2007)).

In turn, Figures 7a-7d 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 2.3.4, 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 8a-8d 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.

4.3 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 2.2.4 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 2.3.4

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.

5 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[10] (see <http://mba.tuck.dartmouth.edu/pages/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 in order to its trace 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%.[11]

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

[10] 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.

[11] 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 $V(\mathbf{y}_t)$.

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 Mencia and Sentana (2009b), which assesses whether

$$E\left\{\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0)\left[\varsigma_t(\boldsymbol{\theta}_0)-(N+2)\right]\right\}=E[m_{st}(\boldsymbol{\theta}_0,0)]=\mathbf{0}, \tag{32}$$

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 (2009b), and test the null hypothesis of multivariate Student t innovations against the multivariate asymmetric t distribution in (31). 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 (32) 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 $\tilde{\boldsymbol{\theta}}_T$ and $\tilde{\eta}_T$ is

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

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 (≃.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 2.2.4. 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 2.2.5 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 3a 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.

6 Conclusions

We derive computationally simple 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 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 (2008) 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 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. Finally, 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). We are currently exploring these interesting research avenues.

Appendix

A Proofs

Proposition 1

The asymptotic normality of the Gaussian ML estimators follows directly from 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. Given the discussion in appendix D, to find the score function and conditional information matrix 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. In view of (B21) and (B22), it is clear that $d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi}$ 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,

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

and

$$\mathbf{W}'_{dt}(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] \end{bmatrix} = \mathbf{W}'_d(\boldsymbol{\phi}). \tag{A1}$$

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

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \mathbf{s}_{\mathbf{c}t}(\boldsymbol{\theta}_s, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} \\ \mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s, \boldsymbol{\eta}) &= \tfrac{1}{2}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \end{aligned} \tag{A2}$$

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

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

and

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

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

Then, we can use Proposition 1 in Fiorentini and Sentana (2007) to 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 first block being given by $\mathrm{M}_{ll}(\boldsymbol{\eta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)$, and the second block by

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

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

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

If we then set $\mathrm{M}_{ll}(\mathbf{0})=1$, $\mathrm{M}_{ss}(\mathbf{0})=1$ and $\mathrm{M}_{sr}(\mathbf{0})=\mathbf{0}$, then we finally obtain the expressions for the information matrix under normality reported in the statement of Proposition 1. For other elliptical distributions we can proceed analogously.

In order to obtain the elliptically symmetric semiparametric score we must use expression (D37), which in view of (A1) leads to

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

(A1) 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 corresponding block of the information matrix minus

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

It is also worth mentioning that if we reparametrised the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_s)$ as $\vartheta_2\boldsymbol{\Sigma}^{\circ}(\boldsymbol{\vartheta}_1)$, where

$$\vartheta_2=\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_s)|=\ln|\boldsymbol{\Gamma}|+\ln(1+\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})=\sum\nolimits_{i=1}^{N}\ln\gamma_i+\ln\left(1+\sum\nolimits_{i=1}^{N}\frac{c_i^2}{\gamma_i}\right),$$

then Proposition 12 in Fiorentini and Sentana (2007) would imply that the information matrix would be block diagonal between $\boldsymbol{\vartheta}_1$ and $(\vartheta_2,\boldsymbol{\eta})$, with $\boldsymbol{\vartheta}_1$ being as efficiently estimated as if we knew $\boldsymbol{\eta}$, while we could only achieve the asymptotic efficiency of the Gaussian pseudo Maximum likelihood estimator of ϑ_2, which would be given by the expression:

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

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

These Gaussian PML estimators set to 0 the average value of $\mathbf{s}_{\boldsymbol{\theta}_s t}(\boldsymbol{\theta},\mathbf{0})$, which is trivially obtained from (A2) by noting that $\delta[\varsigma_t(\boldsymbol{\theta}_s);\mathbf{0}]=1$. Similarly, 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 $\mathcal{I}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s t}(\boldsymbol{\theta}_s,\mathbf{0})$ from Proposition 1 in Bollerslev and Wooldridge (1992). However, in order to derive an expression for $\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi})=V\left[\mathbf{s}_{\boldsymbol{\theta}_s t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]$ we must take into account the true distribution of $\mathbf{y}_t$. When this distribution is elliptically symmetric, Proposition 2 in Fiorentini and Sentana (2007) implies that $\mathcal{B}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}(\boldsymbol{\phi})$ also mimics the expression for the information matrix if we replace $\mathrm{M}_{ll}(\boldsymbol{\eta})$ by 1 and $\mathrm{M}_{ss}(\boldsymbol{\eta})$ by $(\kappa+1)$. In more general cases, Proposition 1 in Bollerslev and Wooldridge (1992) coupled with the static nature of the model implies that:

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

where $\mathcal{K}\left(\boldsymbol{\varrho}\right)$ is the matrix of unconditional third and fourth central moments of $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$ defined in (D36). 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

Once again, in order to obtain $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ we need expressions for $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$. But given (B23) and (B25) we will have that

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

and

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

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

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

Now, equation (B24) implies that

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

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

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

Similarly, equation (B26) implies that

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

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 $\partial vec[\boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta})]/\partial \boldsymbol{\theta}'$. 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 $x_{t|t-1}(\boldsymbol{\theta}_s, \mathbf{0}) = 0$, $\mathbf{u}_{t|t-1}(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{0}$, $\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}_s, \mathbf{0}) = diag(1, \boldsymbol{\gamma})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{cc}' + \boldsymbol{\Gamma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_s)$, $x_{t|t}(\boldsymbol{\theta}_s, \mathbf{0}) = f_{kt}(\boldsymbol{\theta}_s)$ 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

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

Hence

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

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

and

$$\mathbf{W}_d(\phi) = \begin{bmatrix} \mathbf{0}' & \mathbf{c}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0}' \end{bmatrix}', \tag{A3}$$

where we have used the fact that

$$\left.\begin{array}{r} E[f_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}] = E[\mathbf{c}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})|\boldsymbol{\theta}_s, \mathbf{0}] = 0 \\ E[\mathbf{v}_{kt-1}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, \mathbf{0}] = E[\mathbf{\Gamma}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})|\boldsymbol{\theta}_s, \mathbf{0}] = \mathbf{0} \end{array}\right\} \tag{A4}$$

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 $\mathbf{\Gamma} > \mathbf{0}$, then we can use the Woodbury formula once again to show that

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

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, \mathbf{0})$, together with (A4), it is easy to show that the unconditional information matrix $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta})$ 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

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

where

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

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

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

and

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

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, which confirms the expressions for the information matrix under normality reported in the statement of Proposition 2. For other elliptical distributions we can proceed analogously.

In addition, it follows from (A3) 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. Once again, the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A4), 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 $\mathcal{B}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi}) = \mathcal{A}_{\boldsymbol{\rho}^\dagger\boldsymbol{\rho}^\dagger}(\boldsymbol{\phi})$. As a result, the Gaussian-based LM test for $H_0:\boldsymbol{\rho}^\dagger=\mathbf{0}$ remains valid irrespective of the true distribution of $\mathbf{y}_t$. □

Proposition 3

Given that in model (11)

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = (\mathbf{I}-\mathbf{P})\boldsymbol{\pi} + \mathbf{P}\mathbf{y}_{t-1} = \boldsymbol{\pi} + \mathbf{P}(\mathbf{y}_{t-1}-\boldsymbol{\pi})$$

and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}$, we will have that

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I}_N-\mathbf{P})\boldsymbol{\Sigma}^{-1/2\prime} & \mathbf{0} \\ [(\mathbf{y}_{t-1}-\boldsymbol{\pi})\otimes\mathbf{I}_N]\boldsymbol{\Sigma}^{-1/2\prime} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1/2\prime}\otimes\boldsymbol{\Sigma}^{-1/2\prime}) \end{bmatrix}.$$

Hence, the Gaussian score vector will be given by

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\pi},\mathbf{p},\boldsymbol{\sigma}) \\ \mathbf{s}_{\mathbf{p}t}(\boldsymbol{\pi},\mathbf{p},\boldsymbol{\sigma}) \\ \mathbf{s}_{\boldsymbol{\sigma}t}(\boldsymbol{\pi},\mathbf{p},\boldsymbol{\sigma}) \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_N-\mathbf{P})\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t-\boldsymbol{\pi}) \\ (\mathbf{I}_N\otimes\boldsymbol{\Sigma}^{-1})vec[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'] \\ \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N vech\{[(\mathbf{y}_t-\boldsymbol{\pi})-\mathbf{P}(\mathbf{y}_{t-1}-\boldsymbol{\pi})] \\ \times[(\mathbf{y}_t-\boldsymbol{\pi})-\mathbf{P}(\mathbf{y}_{t-1}-\boldsymbol{\pi})]'-\boldsymbol{\Sigma}\} \end{bmatrix}$$

and the conditional information matrix by

$$\mathcal{I}_t(\boldsymbol{\theta}) = \left[\begin{array}{cc} (\mathbf{I}_N-\mathbf{P})\boldsymbol{\Sigma}^{-1}(\mathbf{I}_N-\mathbf{P})' & (\mathbf{I}_N-\mathbf{P})[(\mathbf{y}_{t-1}-\boldsymbol{\pi})'\otimes\boldsymbol{\Sigma}^{-1}] \\ [(\mathbf{y}_{t-1}-\boldsymbol{\pi})\otimes\boldsymbol{\Sigma}^{-1}](\mathbf{I}_N-\mathbf{P})' & [(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'\otimes\boldsymbol{\Sigma}^{-1}] \\ \mathbf{0} & \mathbf{0} \end{array}\right.$$
$$\left.\begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N \end{array}\right].$$

If we define $\boldsymbol{\Upsilon} = V(\mathbf{y}_t)$, which can be obtained from the relationship $\boldsymbol{\Upsilon} = \mathbf{P}\boldsymbol{\Upsilon}\mathbf{P}' + \boldsymbol{\Sigma}$, we can finally obtain the following expression for the unconditional information matrix:

$$\mathcal{I}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} (\mathbf{I}_N-\mathbf{P})\boldsymbol{\Sigma}^{-1}(\mathbf{I}_N-\mathbf{P})' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Upsilon}\otimes\boldsymbol{\Sigma}^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N \end{array}\right],$$

where we have used the fact that

$$E(\mathbf{y}_t) = \boldsymbol{\pi}. \tag{A5}$$

From here, it is trivial to show that the score under the null will be

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\pi},\mathbf{0},\boldsymbol{\sigma}) \\ \mathbf{s}_{\mathbf{p}t}(\boldsymbol{\pi},\mathbf{0},\boldsymbol{\sigma}) \\ \mathbf{s}_{\boldsymbol{\sigma}t}(\boldsymbol{\pi},\mathbf{0},\boldsymbol{\sigma}) \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t-\boldsymbol{\pi}) \\ (\mathbf{I}_N\otimes\boldsymbol{\Sigma}^{-1})vec[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'] \\ \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'-\boldsymbol{\Sigma}] \end{array}\right],$$

and

$$\mathcal{I}(\boldsymbol{\pi},\mathbf{0},\boldsymbol{\sigma}) = \left[\begin{array}{ccc} \boldsymbol{\Sigma}^{-1} & 0 & 0 \\ 0 & (\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}^{-1}) & 0 \\ 0 & 0 & \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N \end{array}\right].$$

Given that we are basing our test in the sample average of $vec[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})']$, the above expression confirms that the LM test for $H_0 : \mathbf{p} = \mathbf{0}$ will be given by (12). Finally, the asymptotic distribution follows from standard arguments (see e.g. Newey and McFadden (1994)).

Finally, let us consider the test in Proposition 3, which is based on a full rank linear transformation of the Gaussian scores $s_{\mathbf{p}t}(\boldsymbol{\pi},\mathbf{0},\boldsymbol{\sigma})$, when the conditional distribution of $\mathbf{y}_t$ is not multivariate normal. Once again, the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A5) and the fact that $\mathcal{A}_t(\boldsymbol{\theta})$ and $\mathcal{I}_t(\boldsymbol{\theta})$ coincide, implies that $\mathcal{A}(\boldsymbol{\theta})$ will be block diagonal between $\boldsymbol{\pi}$, $\mathbf{p}$ and $\boldsymbol{\sigma}$ irrespective of the true distribution of $\mathbf{y}_t$. In addition, $\mathcal{A}_{\mathbf{pp}}(\boldsymbol{\theta})$ will coincide with $\mathcal{I}_{\mathbf{pp}}(\boldsymbol{\theta})$. A closely related argument shows that $\mathcal{B}(\boldsymbol{\phi})$ will also be block diagonal between $\mathbf{p}$ and $(\boldsymbol{\pi},\boldsymbol{\sigma})$, and that $\mathcal{B}_{\mathbf{pp}}(\boldsymbol{\theta}) = \mathcal{A}_{\mathbf{pp}}(\boldsymbol{\theta})$. As a result, the Gaussian-based LM test for $H_0 : \mathbf{p} = \mathbf{0}$ remains valid regardless of the true distribution of $\mathbf{y}_t$. □

Proposition 4

Once again, in order to obtain $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ we need expressions for $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$. Since we are assuming that only the common factors can be serially correlated, we can write

(14) in state space representation with $\mathbf{f}_t$ as the only state variable. Then, a straightforward application of the Kalman filter implies that

$$\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\pi} + \mathbf{C}\mathbf{x}_{t|t-1}(\boldsymbol{\theta}), \\
\mathbf{x}_{t|t-1}(\boldsymbol{\theta}) &= \mathbf{R}\mathbf{x}_{t-1|t-1}(\boldsymbol{\theta}), \qquad \text{(A6)} \\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{C}\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}, \\
\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) &= \mathbf{R}\boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta})\mathbf{R}' + \mathbf{I}_k, \qquad \text{(A7)}
\end{aligned}$$

whence

$$d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi} + (d\mathbf{C})\mathbf{x}_{t|t-1}(\boldsymbol{\theta}) + \mathbf{C}d\mathbf{x}_{t|t-1}(\boldsymbol{\theta})$$

and

$$d\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = (d\mathbf{C})\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\mathbf{C}' + \mathbf{C}d\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\mathbf{C}' + \mathbf{C}\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})(d\mathbf{C}') + d\boldsymbol{\Gamma}.$$

As a result,

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

and

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

Now, equation (A6) implies that

$$\frac{\partial \mathbf{x}_{t|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = [\mathbf{x}'_{t-1|t-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_k]\frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\theta}'} + \mathbf{R}\frac{\partial \mathbf{x}_{t-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'},$$

while equation (A7) implies that

$$\frac{\partial vec[\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} = (\mathbf{I}_{k^2} + \mathbf{K}_{kk})(\mathbf{R}\boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_k)\frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\theta}'} + (\mathbf{R} \otimes \mathbf{R})\frac{\partial vec[\boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}.$$

Under the null $\mathbf{x}_{t|t-1}(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{0}$, $\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{I}_k$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{C}\mathbf{C}' + \boldsymbol{\Gamma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_s)$ and $\mathbf{x}_{t|t}(\boldsymbol{\theta}_s, \mathbf{0}) = \mathbf{f}_{kt}(\boldsymbol{\theta}_s)$, so that

$$\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta}_s, \mathbf{0})}{\partial \boldsymbol{\theta}'} = \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}'} + \mathbf{C}[\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) \otimes \mathbf{I}_k]\frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\theta}'}$$

and

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

Hence, if we define $\mathbf{J}$ as the matrix that implicitly imposes the identifiability conditions on $\mathbf{C}$ through the relationship $vec(\mathbf{C}) = \mathbf{J}\mathbf{c}$, then we will have that

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

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

and

$$\mathbf{W}_d(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{0}' & \mathbf{J}'\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0}' \end{bmatrix}', \tag{A8}$$

where we have used the fact that

$$E[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}] = E[\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})|\boldsymbol{\theta}_s,\mathbf{0}] = \mathbf{0} \tag{A9}$$

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

As a result, the score under the null will be

$$\begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \\ s_{ct}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \\ s_{\gamma t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \\ s_{\rho t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi}) \\ \mathbf{J}'vec\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \\ \frac{1}{2}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \\ [\mathbf{f}_{kt-1}(\boldsymbol{\theta})\otimes\mathbf{I}_k]\delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi}) \end{bmatrix}.$$

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

$$s_{\boldsymbol{\rho}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) = \delta[\varsigma_t(\boldsymbol{\theta}_s);\boldsymbol{\eta}]vec[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\mathbf{f}_{kt-1}'(\boldsymbol{\theta}_s)].$$

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})$, together with (A4), it is easy to show that the unconditional information matrix $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta})$ will be block diagonal between $\boldsymbol{\pi}$, $(\mathbf{c},\boldsymbol{\gamma},\boldsymbol{\eta})$ and $\boldsymbol{\rho}$, with the first two blocks being exactly the same as in the static factor model after excluding the restricted elements of $\mathbf{C}$. 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

$$\mathcal{I}_{\boldsymbol{\rho\rho}t}(\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}) = [\mathbf{f}_{kt-1}(\boldsymbol{\theta})\otimes\mathbf{I}_k]\mathcal{V}_{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta})[\mathbf{f}_{kt-1}'(\boldsymbol{\theta})\otimes\mathbf{I}_k],$$

where

$$\mathcal{V}_{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) = V\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\} = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}.$$

But since

$$E[\mathbf{f}_{kt}(\boldsymbol{\theta}_s)\mathbf{f}_{kt}'(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s,\mathbf{0},\boldsymbol{\eta}] = E[\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}|\boldsymbol{\theta}_s,\mathbf{0}] = \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C},$$

we finally obtain that $\mathcal{V}_{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta})$ 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, which confirms the expressions for the information matrix under normality reported in the statement of Proposition 4. For other elliptical distributions we can proceed analogously.

In addition, it follows from (A8) that the elliptically symmetric semiparametric scores for $\boldsymbol{\rho}$ coincide with the parametric ones, and that the elliptically symmetric semiparametric efficiency bound will be block diagonal between $\boldsymbol{\pi}$, $\boldsymbol{\rho}$ 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 $s_{\boldsymbol{\rho}t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ 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. Once again, the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A9), implies that $\mathcal{A}(\boldsymbol{\phi})$ will be block diagonal between $\boldsymbol{\rho}$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$ irrespective of the true distribution of $\mathbf{y}_t$. In addition, $\mathcal{A}_{\boldsymbol{\rho\rho}}(\boldsymbol{\phi})$ will coincide with $\mathcal{I}_{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$. A closely related argument shows that $\mathcal{B}(\boldsymbol{\phi})$ will also be block diagonal between $\boldsymbol{\rho}$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$, and that $\mathcal{B}_{\boldsymbol{\rho\rho}}(\boldsymbol{\phi}) = \mathcal{A}_{\boldsymbol{\rho\rho}}(\boldsymbol{\phi})$. As a result, the Gaussian-based LM test for $H_0 : \boldsymbol{\rho} = \mathbf{0}$ remains valid irrespective of the true distribution of $\mathbf{y}_t$. □

Proposition 5

Given (B27) and (B28) it is clear that $d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi}$ 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

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

we will have that:

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

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

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

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,

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

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

$$\mathbf{W}_d(\boldsymbol{\phi})=\begin{bmatrix}\mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0}\end{bmatrix}', \tag{A10}$$

where we have used the fact that

$$\left.\begin{array}{c}E[f_{kt-1}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1|\boldsymbol{\theta}_s,\mathbf{0}]=0\\ E[v_{kit-1}^2(\boldsymbol{\theta}_s)+c_i^2\omega_k(\boldsymbol{\theta}_s)-\gamma_i|\boldsymbol{\theta}_s,\mathbf{0}]=\mathbf{0}\end{array}\right\} \tag{A11}$$

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

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

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

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

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

so that

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

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})$, together with (A11), it is easy to show that the unconditional information matrix $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0},\boldsymbol{\eta})$ 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

$$\mathcal{V}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\theta}_s,\boldsymbol{\eta};\boldsymbol{\eta}) = V\left[\begin{array}{c} \frac{1}{\sqrt{2}}\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]f_{kt}^2(\boldsymbol{\theta}_s)+\omega_k(\boldsymbol{\theta}_s)-1\} \\ \frac{1}{\sqrt{2}}\boldsymbol{\Gamma}^{-1}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}]\mathbf{v}_{kt}(\boldsymbol{\theta}_s)\mathbf{v}'_{kt}(\boldsymbol{\theta}_s)+\mathbf{cc}'\omega_k(\boldsymbol{\theta}_s)-\boldsymbol{\Gamma}\} \end{array}\right]$$
$$= \mathrm{M}_{ss}(\boldsymbol{\eta})\left[\begin{array}{cc} [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \\ \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}\odot\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2} \end{array}\right]$$
$$+\frac{[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]}{2}\left[\begin{array}{cc} [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]^2 & [\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] \\ [\mathbf{c}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c}]vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] & vecd[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}]vecd'[\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{1/2}] \end{array}\right]. \tag{A12}$$

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 $\text{M}_{ss}(\boldsymbol{\eta})$ by $\kappa+1$. If we set $\text{M}_{ss}(\boldsymbol{\eta})=1$ and $\kappa=0$, then we obtain the expressions for the information matrix under normality reported in the statement of Proposition 5. For other elliptical distributions we can proceed analogously.

In addition, it follows from (A10) 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. The structure of

$\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A11) 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

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

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 $\mathcal{A}_{\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}^\dagger}(\boldsymbol{\phi})$. □

Proposition 6

Given that in model (23) $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\pi}$ and

$$vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})] = \mathbf{D}_N vech(\boldsymbol{\Sigma}) + \mathbf{D}_N \mathbf{A} vech[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})' - \boldsymbol{\Sigma}],$$

we will have that $d\boldsymbol{\mu}_t(\boldsymbol{\theta}) = d\boldsymbol{\pi}$ and

$$\begin{aligned} dvec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})] &= \mathbf{D}_N dvech(\boldsymbol{\Sigma}) + \{vech'[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})' - \boldsymbol{\Sigma}] \otimes \mathbf{D}_N\} dvec(\mathbf{A}) \\ &\quad - \mathbf{D}_N \mathbf{A} \{\mathbf{D}_N^+ (\mathbf{I}_{N^2} + \mathbf{K}_{NN})[(\mathbf{y}_{t-1} - \boldsymbol{\pi}) \otimes \mathbf{I}_N] d\boldsymbol{\pi} + dvech(\boldsymbol{\Sigma})\} \end{aligned}$$

so that the only non-zero elements of the Jacobian will be $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\pi}' = \mathbf{I}_N$,

$$\begin{aligned} \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\pi}'} &= -\mathbf{D}_N \mathbf{A} \mathbf{D}_N^+ (\mathbf{I}_{N^2} + \mathbf{K}_{NN})[(\mathbf{y}_{t-1} - \boldsymbol{\pi}) \otimes \mathbf{I}_N] \\ \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\sigma}'} &= \mathbf{D}_N (\mathbf{I}_{N(N+1)/2} - \mathbf{A}), \\ \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \mathbf{a}'} &= \{vech'[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})' - \boldsymbol{\Sigma}] \otimes \mathbf{D}_N\}, \end{aligned}$$

where $\mathbf{D}_N^+ = (\mathbf{D}_N' \mathbf{D}_N)^{-1} \mathbf{D}_N$ is the Moore-Penrose inverse of $\mathbf{D}_N$. But since we are only interested in evaluating these derivatives under the null hypothesis of $\boldsymbol{\alpha} = \mathbf{0}$, we will have that

$$\mathbf{Z}_{dt}(\boldsymbol{\pi}, \boldsymbol{\sigma}, \mathbf{0}) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1/2\prime} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1/2\prime} \otimes \boldsymbol{\Sigma}^{-1/2\prime}) \\ \mathbf{0} & \frac{1}{2}\{vech[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})' - \boldsymbol{\Sigma}] \otimes \mathbf{D}_N'\}(\boldsymbol{\Sigma}^{-1/2\prime} \otimes \boldsymbol{\Sigma}^{-1/2\prime}) \end{bmatrix}.$$

Hence, the Gaussian score vector under the null will be given by

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\pi}, \boldsymbol{\sigma}, \mathbf{0}) \\ \mathbf{s}_{\boldsymbol{\sigma}t}(\boldsymbol{\pi}, \boldsymbol{\sigma}, \mathbf{0}) \\ \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\pi}, \boldsymbol{\sigma}, \mathbf{0}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\pi}) \\ \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{D}_N vech[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})' - \boldsymbol{\Sigma}] \\ \frac{1}{2}\{\mathbf{I}_{N(N+1)/2} \otimes [\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{D}_N]\} \\ \times vec\{vech[(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})' - \boldsymbol{\Sigma}] vech'[(\mathbf{y}_{t-1} - \boldsymbol{\pi})(\mathbf{y}_{t-1} - \boldsymbol{\pi})' - \boldsymbol{\Sigma}]\} \end{bmatrix},$$

and the conditional information matrix by

$$\mathcal{I}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \boldsymbol{\Sigma}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N & \frac{1}{2}\{vech'[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\otimes\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N\} \\ \mathbf{0} & \frac{1}{2}\{vech[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\otimes\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N\} & \frac{1}{2}\{vech[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]vech'[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\otimes\mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N\} \end{array}\right].$$

But since

$$E[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'] = \boldsymbol{\Sigma} \tag{A13}$$

regardless of the distribution of $\mathbf{y}_t$, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ will be block diagonal between $\boldsymbol{\pi}$, $\boldsymbol{\sigma}$ and $\mathbf{a}$. Consequently, in computing our ML-based tests we can safely ignore the sampling uncertainty in the sample means, variances and covariances of $\mathbf{y}_t$.

Given that we are basing our test on

$$\begin{aligned} & vec\{vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]vech'[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\} \\ = & \{vech[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\otimes\mathbf{I}_{N(N+1)/2}\}vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'-\boldsymbol{\Sigma}] \\ = & 2\{\mathbf{I}_{N(N+1)/2}\otimes[\mathbf{D}_N^+(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{D}_N^{+\prime}]\}\mathbf{s}_{\mathbf{a}t}(\boldsymbol{\pi},\boldsymbol{\sigma},\mathbf{0}) \end{aligned}$$

the asymptotic covariance matrix of $vec[\bar{\mathbf{S}}_{\mathbf{yy}}(1)]$ will be

$$4\{\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0})\otimes\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0})\},$$

where

$$\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) = V\left\{\frac{1}{\sqrt{2}}vech[(\mathbf{y}_t-\boldsymbol{\pi})(\mathbf{y}_t-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\right\}.$$

But since

$$\begin{aligned} & V\{vec[(\mathbf{y}_{t-1}-\boldsymbol{\pi})(\mathbf{y}_{t-1}-\boldsymbol{\pi})'-\boldsymbol{\Sigma}]\} \\ = & (\boldsymbol{\Sigma}^{1/2}\otimes\boldsymbol{\Sigma}^{1/2})V[vec(\varepsilon^*_{t-1}\varepsilon^{*\prime}_{t-1}-\mathbf{I}_N)](\boldsymbol{\Sigma}^{1/2\prime}\otimes\boldsymbol{\Sigma}^{1/2\prime}) \\ = & (\boldsymbol{\Sigma}^{1/2}\otimes\boldsymbol{\Sigma}^{1/2})[(\kappa+1)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+\kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)](\boldsymbol{\Sigma}^{1/2\prime}\otimes\boldsymbol{\Sigma}^{1/2\prime}) \\ = & (\kappa+1)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\kappa vec(\boldsymbol{\Sigma})vec'(\boldsymbol{\Sigma}) = \mathbf{H}(\kappa), \end{aligned} \tag{A14}$$

when the conditional distribution of $\mathbf{y}_t$ is elliptically symmetric, we will have that

$$\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\eta}) = \tfrac{1}{2}\mathbf{D}_N^+\mathbf{H}(\kappa)\mathbf{D}_N^{+\prime} = (\kappa+1)\mathbf{D}_N^+(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{D}_N^{+\prime} + \frac{\kappa}{2}vech(\boldsymbol{\Sigma})vech'(\boldsymbol{\Sigma})]. \tag{A15}$$

Finally, the result in Proposition 6 follows from the fact that

$$\mathcal{V}^{-1}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\mathbf{0}) = \mathbf{D}_N'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_N.$$

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, the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A13) and the fact that $\mathcal{A}_t(\boldsymbol{\theta})$ and $\mathcal{I}_t(\boldsymbol{\theta})$ coincide, implies that $\mathcal{A}(\boldsymbol{\theta})$ will also be block diagonal between $\boldsymbol{\pi}$, $\boldsymbol{\sigma}$ and $\mathbf{a}$ irrespective of the true distribution of $\mathbf{y}_t$. Likewise, it is easy to see that $\mathcal{B}(\boldsymbol{\theta})$ will also be block diagonal between $(\boldsymbol{\pi},\boldsymbol{\sigma})$ and $\mathbf{a}$. As a result, the asymptotic covariance matrix of $vec[\bar{\mathbf{S}}_{\mathbf{yy}}(1)]$ will be

$$4\{\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho})\otimes\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho})\}.$$

In non-elliptical cases, we can find $\mathcal{V}_{\mathbf{aa}}(\boldsymbol{\theta}_s,\mathbf{0};\boldsymbol{\varrho})$ by replacing $V[vec(\boldsymbol{\varepsilon}^*_{t-1}\boldsymbol{\varepsilon}^{*\prime}_{t-1}-\mathbf{I}_N)]$ in (A14) by the 2,2 block of $\mathcal{K}(\boldsymbol{\varrho})$. □

Proposition 7

Given that in model (25) $\boldsymbol{\mu}_t(\boldsymbol{\theta})=\boldsymbol{\pi}$ and

$$\boldsymbol{\Sigma}_t(\boldsymbol{\theta})=\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'+\boldsymbol{\Gamma},$$

we will have that $d\boldsymbol{\mu}_t(\boldsymbol{\theta})=d\boldsymbol{\pi}$ and

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

so that

$$\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial vec(\mathbf{C})}{\partial\boldsymbol{\theta}'}+(\mathbf{C}\otimes\mathbf{C})\frac{\partial vec[\boldsymbol{\Lambda}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}+\mathbf{E}_N\frac{\partial vecd(\boldsymbol{\Gamma})}{\partial\boldsymbol{\theta}'},$$

where

$$\begin{aligned}\frac{\partial vech[\boldsymbol{\Lambda}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} &= \{vech'[E(\mathbf{f}_{t-1}\mathbf{f}_{t-1}-\mathbf{I}_k|Y_{t-1},\boldsymbol{\theta})]\otimes\mathbf{I}_{k(k+1)/2}\}\frac{\partial vec(\mathbf{A})}{\partial\boldsymbol{\theta}'}\\ &\quad+\mathbf{A}\left\{\frac{\partial vech[E(\mathbf{f}_{t-1}\mathbf{f}_{t-1}|Y_{t-1},\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}\right\}.\end{aligned}$$

But since we are only interested in evaluating these derivatives under the null hypothesis of $\boldsymbol{\alpha}=\mathbf{0}$, in which case $\boldsymbol{\Lambda}_t(\boldsymbol{\theta})=\mathbf{I}_k$ and $E(\mathbf{f}_{t-1}\mathbf{f}_{t-1}|Y_{t-1},\boldsymbol{\theta})]=\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta})+\boldsymbol{\Omega}_k(\boldsymbol{\theta})$, we will have that

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s,\mathbf{0})=\begin{bmatrix}\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) & \mathbf{0}\\ \mathbf{0} & \frac{1}{2}\mathbf{J}'(\mathbf{C}'\otimes\mathbf{I}_N)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ \mathbf{0} & \frac{1}{2}\mathbf{E}'_N[\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\\ \mathbf{0} & \begin{array}{c}\frac{1}{2}\{vech[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta})+\boldsymbol{\Omega}_k(\boldsymbol{\theta})-\mathbf{I}_k]\otimes\mathbf{D}'_k\}\\ \times[\mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)\otimes\mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]\end{array}\end{bmatrix},$$

where $\mathbf{J}$ as the matrix that implicitly imposes the identifiability conditions on $\mathbf{C}$ through the relationship $vec(\mathbf{C})=\mathbf{J}\mathbf{c}$. Consequently

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

and

$$\mathbf{W}_d(\boldsymbol{\phi}) = \left[\begin{array}{cccc} \mathbf{0}' & \mathbf{J}'\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0}' \end{array} \right]', \tag{A16}$$

where we have used the fact that

$$E[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k|\boldsymbol{\theta}_s, \mathbf{0}] = \mathbf{0}. \tag{A17}$$

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

Hence, the score vector under the null will be given by

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\mathbf{c}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \\ \mathbf{s}_{\boldsymbol{\alpha}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) \end{array} \right] = \left[\begin{array}{l} \delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \mathbf{J}'vec\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \\ \frac{1}{2}vecd\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\} \\ \frac{1}{2}[\mathbf{I}_{N(N+1)/2} \otimes (\mathbf{D}'_k\mathbf{D}_k)]vec[vech\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) \\ \times(\mathbf{y}_t - \boldsymbol{\pi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}\}vech'[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k]] \end{array} \right],$$

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

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\alpha}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) &= \frac{1}{2}[\mathbf{I}_{N(N+1)/2} \otimes (\mathbf{D}'_k\mathbf{D}_k)]vec[vech\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{f}_{kt}(\boldsymbol{\theta})\mathbf{f}'_{kt}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k\} \\ &\quad \times vech'\{\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k\}] \end{aligned}$$

Using the expression for $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta})$ it is also easy to show that the conditional information matrix $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}_s, \mathbf{0})$ will be block diagonal between $\boldsymbol{\pi}$, $(\mathbf{c}, \boldsymbol{\gamma}, \boldsymbol{\eta})$ and $\boldsymbol{\alpha}$, with the first two blocks being exactly the same as in the static factor model after excluding the restricted elements of $\mathbf{C}$. Thus, in computing our ML-based tests we can safely ignore the sampling uncertainty in estimating $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$.

In addition, given that we can also express

$$\mathbf{s}_{\boldsymbol{\alpha}t}(\boldsymbol{\theta}_s, \mathbf{0}, \boldsymbol{\eta}) = \tfrac{1}{2}\{vech[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k] \otimes (\mathbf{D}'_k\mathbf{D}_k)\}vech\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{f}_{kt}(\boldsymbol{\theta})\mathbf{f}'_{kt}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k\}$$

we can write

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\alpha}\boldsymbol{\alpha}t}(\boldsymbol{\theta}, \mathbf{0}, \boldsymbol{\eta}) &= [\tfrac{1}{\sqrt{2}}vech[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k] \otimes (\mathbf{D}'_k\mathbf{D}_k)] \\ &\times \mathcal{V}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta}) \times [\tfrac{1}{\sqrt{2}}vech'[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k] \otimes (\mathbf{D}'_k\mathbf{D}_k)] \end{aligned}$$

where

$$\begin{aligned} \mathcal{V}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta}) &= V\left[\frac{1}{\sqrt{2}}vech\{\delta[\varsigma_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]\mathbf{f}_{kt}(\boldsymbol{\theta})\mathbf{f}'_{kt}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k\}\right] \\ &= \mathrm{M}_{ss}(\boldsymbol{\eta})\mathbf{D}_k^+[\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C} \otimes \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}]\mathbf{D}_k^{+\prime} \\ &+ \frac{[\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]}{2}vech[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}]vech'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}]. \end{aligned}$$

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

$$
\begin{aligned}
& V\{vech[\mathbf{f}_{kt-1}(\boldsymbol{\theta})\mathbf{f}'_{kt-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_k(\boldsymbol{\theta}) - \mathbf{I}_k]\} \qquad \text{(A18)} \\
&= \mathbf{D}_k^+[\mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \otimes \mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)]V[vec(\boldsymbol{\varepsilon}^*_{t-1}\boldsymbol{\varepsilon}^{*\prime}_{t-1} - \mathbf{I}_k)][\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{C} \otimes \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{C}]\mathbf{D}_k^{+\prime} \\
&= \mathbf{D}_k^+[\mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s) \otimes \mathbf{C}'\boldsymbol{\Sigma}^{-1/2\prime}(\boldsymbol{\theta}_s)][(\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)] \\
&\quad \times[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{C} \otimes \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_s)\mathbf{C}]\mathbf{D}_k^{+\prime} \\
&= 2(\kappa+1)\mathbf{D}_k^+[\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C} \otimes \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}]\mathbf{D}_k^{+\prime} + \kappa vech[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}]vech'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{C}],
\end{aligned}
$$

so $\mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\eta})$ mimics $\mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta})$ if we replace $\mathrm{M}_{ss}(\boldsymbol{\eta})$ by $\kappa + 1$. Hence, the information matrix will be

$$
\mathcal{J}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}) = \mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\eta}) \otimes (\mathbf{D}'_k\mathbf{D}_k)\mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \boldsymbol{\eta}; \boldsymbol{\eta})(\mathbf{D}'_k\mathbf{D}_k). \qquad \text{(A19)}
$$

If we then set $\mathrm{M}_{ss}(\boldsymbol{\eta}) = 1$ and $\kappa = 0$, we can use the same argument as in the proof of Proposition 6 to derive the test statistic in the statement of Proposition 7. For other elliptical distributions we can proceed analogously.

In addition, it follows from (A16) that the elliptically symmetric semiparametric scores for $\boldsymbol{\alpha}$ 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 $\boldsymbol{\alpha}$ 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 $s_{\boldsymbol{\alpha}t}(\boldsymbol{\theta}_s, \mathbf{0}, \mathbf{0})$ 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. The structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$, together with (A17) and the fact that $\mathcal{A}_t(\boldsymbol{\theta})$ and $\mathcal{I}_t(\boldsymbol{\theta})$ coincide, implies that $\mathcal{A}(\boldsymbol{\phi})$ will be block diagonal between $\boldsymbol{\alpha}$ and $(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\gamma})$ irrespective of the true distribution of $\mathbf{y}_t$. In addition, it is easy to prove that

$$
\mathcal{A}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}) = E[\mathcal{A}_{\boldsymbol{\alpha\alpha}t}(\boldsymbol{\phi})|\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}] = \mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \otimes (\mathbf{D}'_k\mathbf{D}_k)\mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \mathbf{0})(\mathbf{D}'_k\mathbf{D}_k).
$$

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

$$
\mathcal{B}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}) = E[\mathcal{B}_{\boldsymbol{\alpha\alpha}t}(\boldsymbol{\phi})|\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}] = \mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho}) \odot (\mathbf{D}'_k\mathbf{D}_k)\mathcal{V}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\theta}_s, \mathbf{0}; \boldsymbol{\varrho})(\mathbf{D}'_k\mathbf{D}_k),
$$

which is generally different from $\mathcal{A}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi})$. □

Proposition 8

The proof of this proposition combines many elements of the proofs of Propositions 2 and 5. Given that model (26) reduces to model (5) 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 (16) 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

$$\mathbf{W}_d(\boldsymbol{\phi}) = \left[\begin{array}{ccccccc} \mathbf{0} & \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) & \frac{1}{2}vecd'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)] & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right]'. \quad \text{(A20)}$$

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}_{\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

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

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 (A4) and (A11), 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 5, 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

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

which confirms (29) in view of the stationarity of $\mathbf{y}_t$. □

References

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

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

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.

Balestra, P., and Holly, A. (1990), "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.

Bekker, P.A. (1989): "Identification in restricted factor models and the evaluation of rank conditions", *Journal of Econometrics* 41, 5-16.

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. (2009): *Portfolio risk analysis*, mimeo, http://economics.nuim.ie/staff/connor/documents/cgk_january_version_final_version.pdf

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. (2008): "Multivariate residual-based finite-sample tests for serial dependence and ARCH effects with application to asset pricing models", *Journal of Applied Econometrics*, forthcoming.

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.

Dunn, J.E. (1973): "A note on a sufficiency condition for uniqueness of a restricted factor matrix", *Psychometrika* 38, 141-143.

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.

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

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

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.

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., 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.

Jennrich, R.I. (1978): "Rotational equivalence of factor loading matrices with specified values", *Psychometrika* 43, 421-426.

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.

Lütkepohl, H. (1993): *Introduction to multiple time series analysis*, 2nd. ed., Springer, Berlin.

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.

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

Mencía, J. and E. Sentana (2009b): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", mimeo, CEMFI.

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. (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.

Wegge, L.E. (1996): "Local identifiability of the factor analysis and measurement error model parameter", *Journal of Econometrics* 70, 351-382.

B Kalman filter recursions

B.1 Static factor models

Model (1) can be regarded as a time-series state-space representation, with f_t as the state, $\mathbf{y}_t = \mathbf{c}f_t + \mathbf{v}_t$ as the measurement equation, and $f_t = 0 \cdot f_{t-1} + f_t$ as transition equation. In this framework, it is straightforward to prove that the Kalman filter prediction equations are

$$
\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}_s) &= \boldsymbol{\pi}, \\
f_{t|t-1}(\boldsymbol{\theta}_s) &= 0,
\end{aligned}
\tag{B21}
$$

and

$$
\begin{aligned}
\Sigma_t(\boldsymbol{\theta}_s) &= \mathbf{c}\mathbf{c}' + \boldsymbol{\Gamma}, \\
\omega_{t|t-1}(\boldsymbol{\theta}_s) &= 1,
\end{aligned}
\tag{B22}
$$

while the updating equations are:

$$
\begin{aligned}
f_{t|t}(\boldsymbol{\theta}_s) = \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) = f_{kt}(\boldsymbol{\theta}_s), \\
\omega_{t|t}(\boldsymbol{\theta}_s) = 1 - \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} = \omega_k(\boldsymbol{\theta}_s).
\end{aligned}
$$

If we define $\mathbf{v}_{t|t}(\boldsymbol{\theta}_s) = E(\mathbf{v}_t|Y_t;\boldsymbol{\theta}_s) = \Gamma\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)(\mathbf{y}_t - \boldsymbol{\pi}) = \mathbf{v}_{kt}(\boldsymbol{\theta}_s)$, then the matrix

$$
V\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array} \middle| Y_t;\boldsymbol{\theta}_s \right) = \begin{pmatrix} 1 - \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & -\mathbf{c}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \\ -\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\mathbf{c} & \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s)\boldsymbol{\Gamma} \end{pmatrix} = \begin{pmatrix} \omega_k(\boldsymbol{\theta}_s) & -\mathbf{c}\omega_k(\boldsymbol{\theta}_s) \\ -\mathbf{c}\omega_k(\boldsymbol{\theta}_s) & \mathbf{c}\mathbf{c}'\omega_k(\boldsymbol{\theta}_s) \end{pmatrix}
$$

will be of rank 1 because $\mathbf{v}_{kt}(\boldsymbol{\theta}_s) = \mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}f_{kt}(\boldsymbol{\theta}_s)$. Similarly, $V[(\begin{array}{cc} f_{kt} & \mathbf{v}'_{kt} \end{array})'|\boldsymbol{\theta}_s]$ will be of rank N because $f_{kt}(\boldsymbol{\theta}_s) = \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{v}_{kt}(\boldsymbol{\theta}_s)$.

Importantly, given the degenerate nature of the transition equation, smoothing is unnecessary in this case, so that $f_{kt}(\boldsymbol{\theta}_s) = E(f_t|Y_T;\boldsymbol{\theta}_s)$ and $\omega_{kt}(\boldsymbol{\theta}_s) = V(f_t|Y_T;\boldsymbol{\theta}_s)$ (see e.g. Diebold and Nerlove (1989) or Harvey (1989)).

Finally, if $\boldsymbol{\Gamma} > \mathbf{0}$, then we can use the Woodbury formula to prove that

$$
\begin{aligned}
f_{kt}(\boldsymbol{\theta}_s) &= \omega_{kt}(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Gamma}^{-1}(\mathbf{y}_t - \boldsymbol{\pi}), \\
\omega_k(\boldsymbol{\theta}_s) &= (1 + \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})^{-1}, \\
\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_s) &= \boldsymbol{\Gamma}^{-1} - \omega_k(\boldsymbol{\theta}_s)\boldsymbol{\Gamma}^{-1}\mathbf{c}\mathbf{c}'\boldsymbol{\Gamma}^{-1},
\end{aligned}
$$

which greatly simplifies the computations (see Sentana (2000)).

B.2 Conditionally homoskedastic dynamic factor models

Although from a computational point of view this is not the most efficient formulation, for our purposes it is convenient to write model (5) in state-space form as

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\pi} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} x_t \\ \mathbf{u}_t \end{pmatrix}, \\ \begin{pmatrix} x_t \\ \mathbf{u}_t \end{pmatrix} &= \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{bmatrix} \begin{pmatrix} x_{t-1} \\ \mathbf{u}_{t-1} \end{pmatrix} + \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix}.\end{aligned}$$

Subject to an assumption about initialisation, such as that $(x_0, \mathbf{u}_0')$ is drawn from its stationary distribution, the Kalman filter prediction equations will be

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\pi} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix}, \tag{B23}$$

$$\begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} = \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{bmatrix} \begin{pmatrix} x_{t-1|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t-1|t-1}(\boldsymbol{\theta}) \end{pmatrix}, \tag{B24}$$

and

$$\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix}, \tag{B25}$$

$$\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) = \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{bmatrix} \boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta}) \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{bmatrix} + \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix}, \tag{B26}$$

while the updating equations will be

$$\begin{pmatrix} x_{t|t}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} + \boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \left[\mathbf{y}_t - \boldsymbol{\pi} - \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} \right]$$

and

$$\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) = \boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) - \boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}).$$

In this sense, note that

$$\begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} = \mathbf{0},$$

which simply reflects the fact that $\mathbf{u}_{t|t}(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}x_{t|t}(\boldsymbol{\theta})$.

B.3 Conditionally heteroskedastic factor models with constant conditional means

If we define $(f_t, \mathbf{v}_t')$ as the state variables, the state-space representation of model (16) is

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\pi} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix}, \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} &= \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} f_{t-1} \\ \mathbf{v}_{t-1} \end{pmatrix} + \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix}, \\ V\left[\begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} \middle| I_{t-1}; \boldsymbol{\theta} \right] &= \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}.\end{aligned}$$

Subject to an assumption about initialisation, such as that $(f_0, \mathbf{v}_0')$ are drawn from their stationary distribution, the Kalman filter prediction equations will be

$$
\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\pi}, \\
\begin{pmatrix} f_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{v}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} &= \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix},
\end{aligned} \tag{B27}
$$

and

$$
\begin{aligned}
\Sigma_t(\boldsymbol{\theta}) &= \mathbf{cc}'\lambda_t(\boldsymbol{\theta}) + \boldsymbol{\Gamma}_t(\boldsymbol{\theta}), \\
\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) &= \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix},
\end{aligned} \tag{B28}
$$

while the updating equations will be

$$
\begin{pmatrix} f_{t|t}(\boldsymbol{\theta}) \\ \mathbf{v}_{t|t}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \lambda_t(\boldsymbol{\theta})\mathbf{c}' \\ \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})(\mathbf{y}_t - \boldsymbol{\pi})
$$

and

$$
\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) = \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) - \lambda_t^2(\boldsymbol{\theta})\mathbf{c}'\Sigma_t^{-1}(\boldsymbol{\theta})\mathbf{c} & -\lambda_t(\boldsymbol{\theta})\mathbf{c}'\Sigma_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \\ -\boldsymbol{\Gamma}_t(\boldsymbol{\theta})\Sigma_t^{-1}(\boldsymbol{\theta})\mathbf{c}\lambda_t(\boldsymbol{\theta}) & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) - \boldsymbol{\Gamma}_t(\boldsymbol{\theta})\Sigma_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}.
$$

If $\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) > \mathbf{0}$, then we can use the Woodbury formula to prove that

$$
\begin{aligned}
f_{t|t}(\boldsymbol{\theta}) &= \omega_{t|t}(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Gamma}_t^{-1}(\boldsymbol{\theta})(\mathbf{y}_t - \boldsymbol{\pi}), \\
\omega_{t|t}(\boldsymbol{\theta}) &= [\mathbf{c}'\boldsymbol{\Gamma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} + \lambda_t^{-1}(\boldsymbol{\theta})]^{-1}, \\
\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) &= \boldsymbol{\Gamma}_t^{-1}(\boldsymbol{\theta}) - \omega_{t|t}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t^{-1}(\boldsymbol{\theta})\mathbf{cc}'\boldsymbol{\Gamma}_t^{-1}(\boldsymbol{\theta}),
\end{aligned}
$$

which greatly simplifies the computations (see Sentana (2000)).

The degenerate nature of the transition equation implies that smoothing is also unnecessary in this case, so that $f_{t|t}(\boldsymbol{\theta}) = E(f_t|Y_T;\boldsymbol{\theta})$ and $\omega_{t|t}(\boldsymbol{\theta}) = V(f_t|Y_T;\boldsymbol{\theta})$ (see Diebold and Nerlove (1989)).

B.4 Conditionally heteroskedastic dynamic factor models

Following Harvey, Ruiz and Sentana (1992), we can write model (26) using the following state representation:

$$
\mathbf{y}_t = \boldsymbol{\pi} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_t \\ \mathbf{u}_t \\ f_t \\ \mathbf{v}_t \end{pmatrix},
$$

$$
\begin{pmatrix} x_t \\ \mathbf{u}_t \\ f_t \\ \mathbf{v}_t \end{pmatrix} = \begin{pmatrix} \rho & \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \mathbf{u}_{t-1} \\ f_{t-1} \\ \mathbf{v}_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \\ 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix},
$$

$$
V\left[\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\right)\middle| I_{t-1};\boldsymbol{\theta}\right] = \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}.
$$

Subject to some initial conditions, the prediction equations will be

$$
\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\pi} + \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix}, \\
\begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} &= \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{bmatrix} \begin{pmatrix} x_{t-1|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t-1|t-1}(\boldsymbol{\theta}) \end{pmatrix}, \\
\begin{pmatrix} f_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{v}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} &= \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix},
\end{aligned} \tag{B29}
$$

and

$$
\begin{aligned}
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix}, \\
\boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) &= \begin{pmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{pmatrix} \boldsymbol{\Omega}_{11t-1|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \rho & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\rho}^*) \end{pmatrix} + \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}, \\
\boldsymbol{\Omega}_{12t|t-1}(\boldsymbol{\theta}) &= \boldsymbol{\Omega}_{22t|t-1}(\boldsymbol{\theta}) = \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix},
\end{aligned} \tag{B30}
$$

while the updating equations will be

$$
\begin{aligned}
\begin{pmatrix} x_{t|t}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t}(\boldsymbol{\theta}) \end{pmatrix} &= \begin{pmatrix} x_{t|t-1}(\boldsymbol{\theta}) \\ \mathbf{u}_{t|t-1}(\boldsymbol{\theta}) \end{pmatrix} + \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \left[\mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}x_{t|t-1}(\boldsymbol{\theta}) - \mathbf{u}_{t|t-1}(\boldsymbol{\theta})\right], \\
\begin{pmatrix} f_{t|t}(\boldsymbol{\theta}) \\ \mathbf{v}_{t|t}(\boldsymbol{\theta}) \end{pmatrix} &= \begin{pmatrix} \lambda_t(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \left[\mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}x_{t|t-1}(\boldsymbol{\theta}) - \mathbf{u}_{t|t-1}(\boldsymbol{\theta})\right] \\ \boldsymbol{\Gamma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \left[\mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}x_{t|t-1}(\boldsymbol{\theta}) - \mathbf{u}_{t|t-1}(\boldsymbol{\theta})\right] \end{pmatrix},
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\Omega}_{11t|t}(\boldsymbol{\theta}) &= \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) - \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c}' \\ \mathbf{I}_N \end{pmatrix} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{c} & \mathbf{I}_N \end{pmatrix} \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}), \\
\boldsymbol{\Omega}_{12t|t}(\boldsymbol{\theta}) &= \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix} - \boldsymbol{\Omega}_{11t|t-1}(\boldsymbol{\theta}) \begin{pmatrix} \lambda_t(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} & \mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \\ \lambda_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} & \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}, \\
\boldsymbol{\Omega}_{22t|t}(\boldsymbol{\theta}) &= \begin{pmatrix} \lambda_t(\boldsymbol{\theta}) - \lambda_t^2(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} & -\lambda_t(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \\ -\lambda_t(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{c} & \boldsymbol{\Gamma}_t(\boldsymbol{\theta}) - \boldsymbol{\Gamma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) \end{pmatrix}.
\end{aligned}
$$

Once again, if $\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) > \mathbf{0}$ then we can use the Woodbury formula to simplify the computations. Interestingly, the expression for $\boldsymbol{\Omega}_{22t|t}(\boldsymbol{\theta})$ coincides with the analogous expression when there are no dynamics in the mean, although the expression for $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is obviously different.

C 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 $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$:

$$
M_T = T\bar{\mathbf{m}}_T'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}), \tag{C31}
$$

where $\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0})$ is the sample average of $\mathbf{m}_t(\boldsymbol{\theta})$ evaluated under the null, and $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix. In order to obtain the non-centrality parameter of this test under Pitman sequences of local alternatives of the form $H_0 : \boldsymbol{\theta}_{2T} = \bar{\boldsymbol{\theta}}_2/\sqrt{T}$, 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 (C31) will be

$$\bar{\boldsymbol{\theta}}'_2\mathbf{M}'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2. \tag{C32}$$

On this basis, we can easily obtain the limiting probability of M_T exceeding some pre-specified quantile of a central χ^2_h distribution from the cdf of a non-central χ^2 distribution with h degrees of freedom and non-centrality parameter (C32).

Finally, note that (C32) 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 $\boldsymbol{\theta}_1$ are asymptotically uncorrelated when H_0 is true, as in all our tests. In addition, both $\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})$ and $\boldsymbol{\Psi}$ coincide with the (2,2) block of the information matrix when $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are the scores with respect to $\boldsymbol{\theta}_2$. This result confirms that the non-centrality parameters of LM and Wald tests will be the same under sequences of local alternatives, which simplifies their computation.

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

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

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

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

for the model considered in section 2.2.4 in view of (7), and

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

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

$$\bar{\boldsymbol{\rho}}^{\dagger\prime} = (\begin{array}{cc} \rho & \rho^* \boldsymbol{\iota}_N' \end{array}).$$

As for the asymptotic covariance matrix, the proof of Proposition 3 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

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

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

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

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

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

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 (21) and (22), respectively, 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,

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

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

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

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

$$\begin{aligned}
\mathbf{M}_s(\boldsymbol{\theta}_s, \mathbf{0})\bar{\boldsymbol{\alpha}}^{\dagger} &= -vec\{(\mathbf{cc}' \otimes \mathbf{cc}')(3\kappa + 2)\alpha + (\mathbf{c} \otimes \mathbf{c})\boldsymbol{\gamma}'\mathbf{E}_N'\kappa\alpha \\
&\quad + \mathbf{E}_N\boldsymbol{\gamma}(\mathbf{c}' \otimes \mathbf{c}')\kappa\alpha^* + \mathbf{E}_N[2(\kappa+1)(\boldsymbol{\Gamma} \odot \boldsymbol{\Gamma}) + \kappa\boldsymbol{\gamma}\boldsymbol{\gamma}']\mathbf{E}_N'\alpha^*\} \qquad \text{(C33)}
\end{aligned}$$

when

$$\bar{\boldsymbol{\alpha}}^{\dagger\prime} = (\ \alpha \quad \alpha^*\boldsymbol{\iota}_N'\ ).$$

As for the asymptotic covariance matrix, the proof of Proposition 6 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 $\mathbf{H}(\kappa)$ is defined in (A14). 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 6 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 $\frac{1}{2}\mathbf{D}_N^+\mathbf{H}(\boldsymbol{\kappa})\mathbf{D}_N^{+\prime}$ is defined in (A15).

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 $vecd(\mathbf{y}_t\mathbf{y}_t' - \mathbf{\Sigma})$ as

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

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

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

Similarly, it is easy to prove that

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

under the null.

D Inference with elliptical innovations

Some useful distribution results

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

Then, it is easy to combine the representation of elliptical distributions above with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of a spherically symmetric distribution with $V(\varepsilon_t^\circ) = \mathbf{I}_N$ are given by

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

and

$$E(\varepsilon_t^\circ \varepsilon_t^{\circ\prime} \otimes \varepsilon_t^\circ \varepsilon_t^{\circ\prime}) = E[vec(\varepsilon_t^\circ \varepsilon_t^{\circ\prime}) vec'(\varepsilon_t^\circ \varepsilon_t^\circ)] = (\kappa_0 + 1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N)], \tag{A2}$$

respectively.

D.1 Log-likelihood function, score vector and information matrix

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $p + r$ parameters of interest, which we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T based on a particular parametric spherical assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln |\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed elliptical density, and $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta}) = \varepsilon_t^{*\prime}(\boldsymbol{\theta})\varepsilon_t^*(\boldsymbol{\theta})$, $\varepsilon_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\varepsilon_t(\boldsymbol{\theta})$ and $\varepsilon_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

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. Then, it is straightforward to 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,

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

where

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

and $\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial \boldsymbol{\theta}'$ depend on the particular specification adopted. For example, $\delta\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ is equal to $(N\eta + 1)/[1 - 2\eta + \eta\varsigma_t(\boldsymbol{\theta})]$ in the student t case, and to 1 under Gaussianity

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

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

and $\mathcal{M}(\boldsymbol{\eta}) = V\left[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right]$. In this context, Proposition 1 in Fiorentini and Sentana (2007) states that:

Proposition 9 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}$ *is i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ *with density* $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$, *then*

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

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

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

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

$$\mathcal{M}_{rr}(\boldsymbol{\eta}) = V[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = -E[\partial \mathbf{e}_{rt}(\boldsymbol{\phi})/\partial \boldsymbol{\eta}'|\,\boldsymbol{\phi}],$$

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

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

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

In the multivariate standardised student t case, in particular:

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

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

where $\psi(.)$ is the di-gamma function (see Abramowitz and Stegun (1964)), which under normality reduce to 1, 1, 0 and $N(N+2)/2$, respectively.

D.2 Gaussian pseudo maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote the Gaussian pseudo-ML (PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ in which $\boldsymbol{\eta}$ is set to zero. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. Proposition 2 in Fiorentini and Sentana (2007) derives the asymptotic distribution of the pseudo-ML estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is elliptical:

Proposition 10 *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with $\kappa_0<\infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\to N[\mathbf{0},\mathcal{C}(\boldsymbol{\phi}_0)]$, where*

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

and $$\mathcal{K}(\kappa)=V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}]=\begin{bmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & (\kappa+1)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)+\kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{bmatrix}, \tag{D34}$$

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

$$\kappa=E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)]-1. \tag{D35}$$

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

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

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 2 in Fiorentini and Sentana (2007) remains valid except for the fact that:

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

where

$$\mathcal{K}\left(\boldsymbol{\varrho}\right)=V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}] \tag{D36}$$

is the matrix of third and fourth order central moments of $\boldsymbol{\varepsilon}_t^*$, whose first block is the identity matrix of order N.

D.3 Elliptically symmetric semiparametric estimators of $\boldsymbol{\theta}$

Hodgson and Vorkink (2001), Hafner and Rombouts (2007) and other authors have suggested semi-parametric estimators of $\boldsymbol{\theta}$ which limit the admissible distributions of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}_0$ to the class of spherically symmetric ones. Proposition 7 in Fiorentini and Sentana (2007) provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 11 *When $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t,I_{t-1},\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with $-2/(N+2)<\kappa_0<\infty$, the elliptically symmetric semiparametric efficient score is given by:*

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0)=\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]\right\}, \tag{D37}$$

where

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\phi}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]'$$
$$= E\left\{ \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]/\partial\boldsymbol{\theta}\cdot vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)] \right| \boldsymbol{\phi}_0 \right\}, \tag{D38}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{ \left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1 \right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]} \right\}. \tag{D39}$$

In practice, $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the joint density of $\boldsymbol{\varepsilon}_t^*$. However, the elliptical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h\left(\varsigma_t; \boldsymbol{\eta}\right)$, avoiding in this way the curse of dimensionality (see appendix B1 in Fiorentini and Sentana (2007) for details).

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

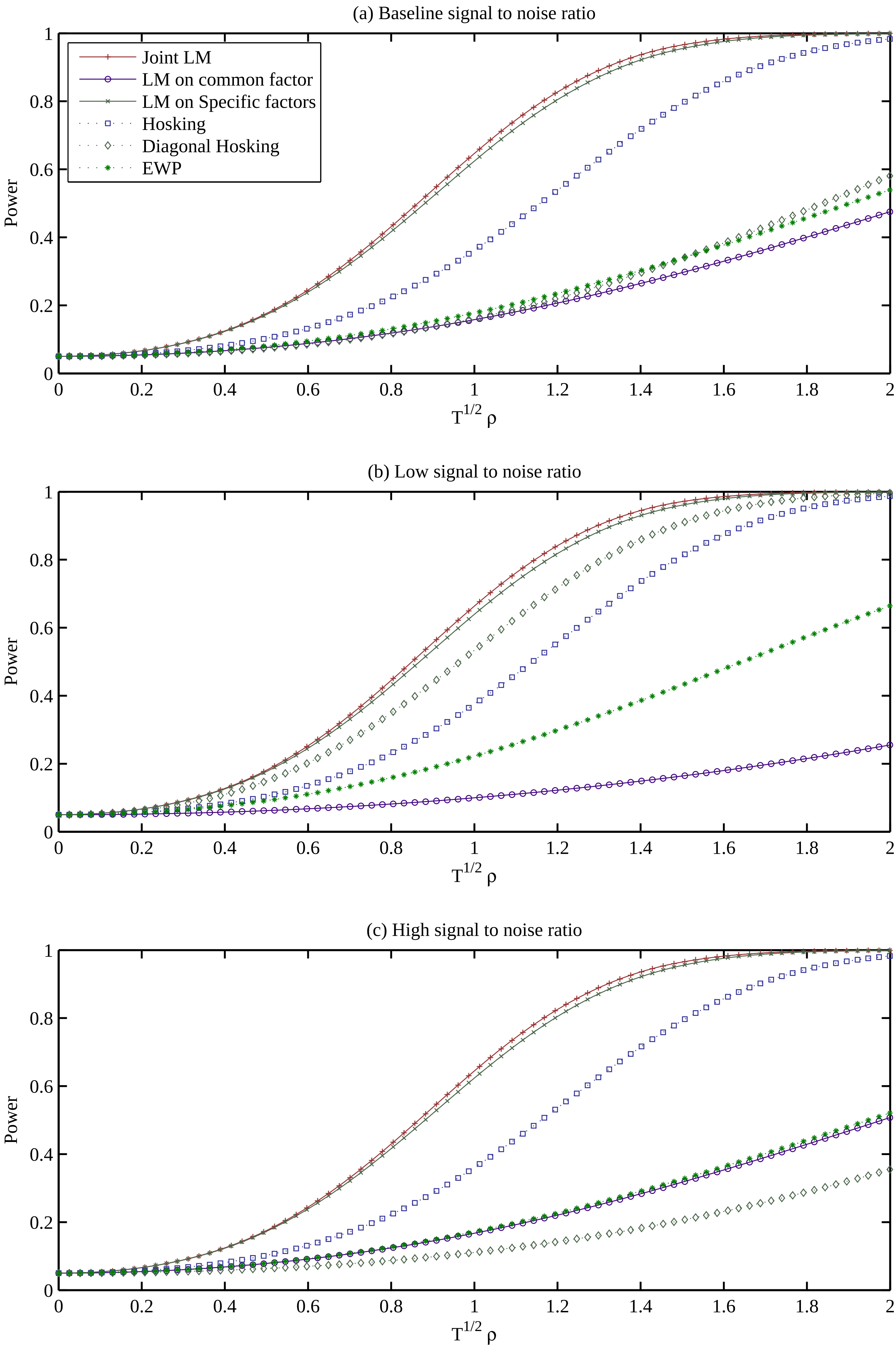


Figure 2: Power of mean dependence tests at 5% level against local alternatives

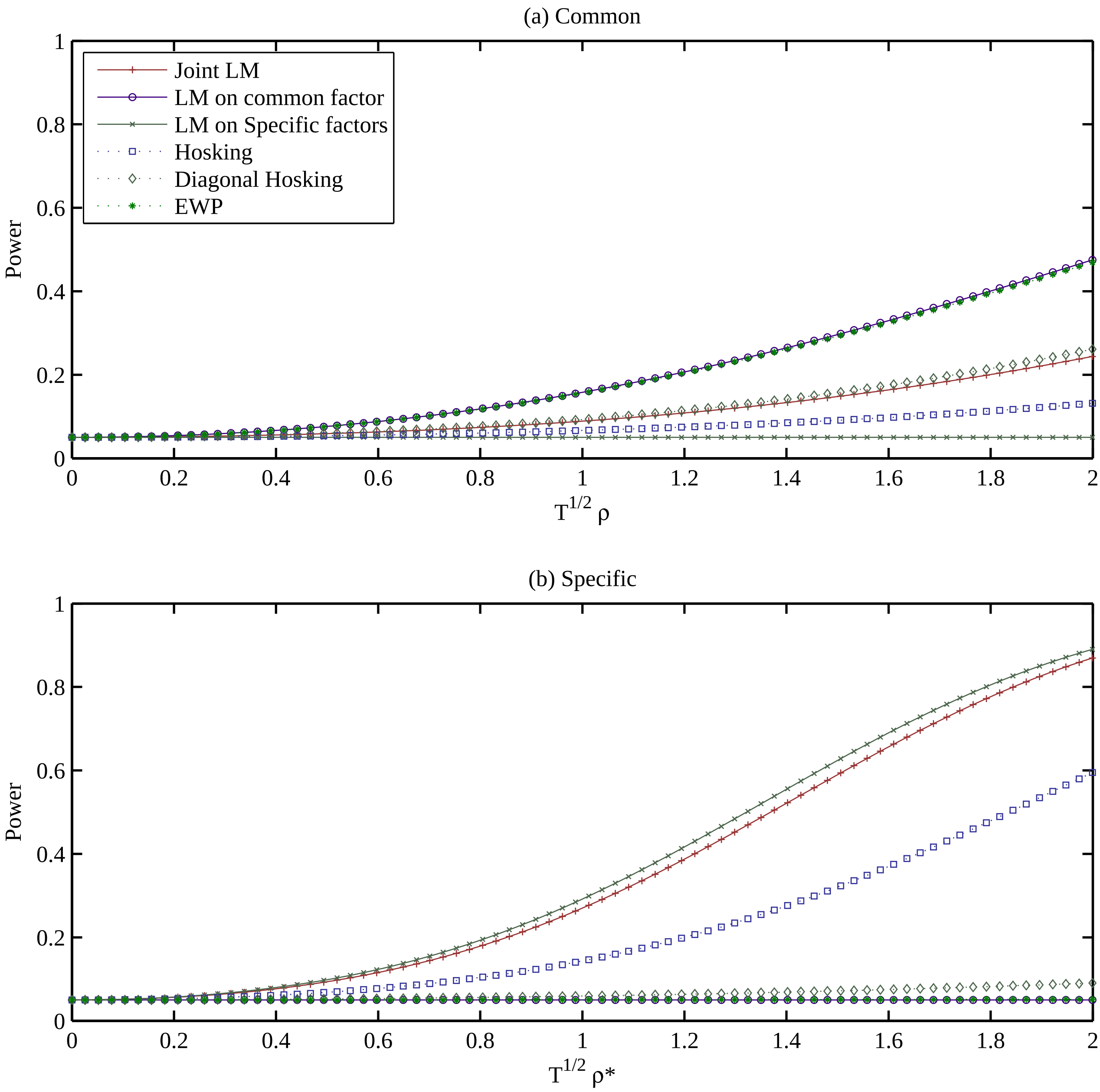


Figure 3: Power of variance dependence tests at 5% level against local alternatives

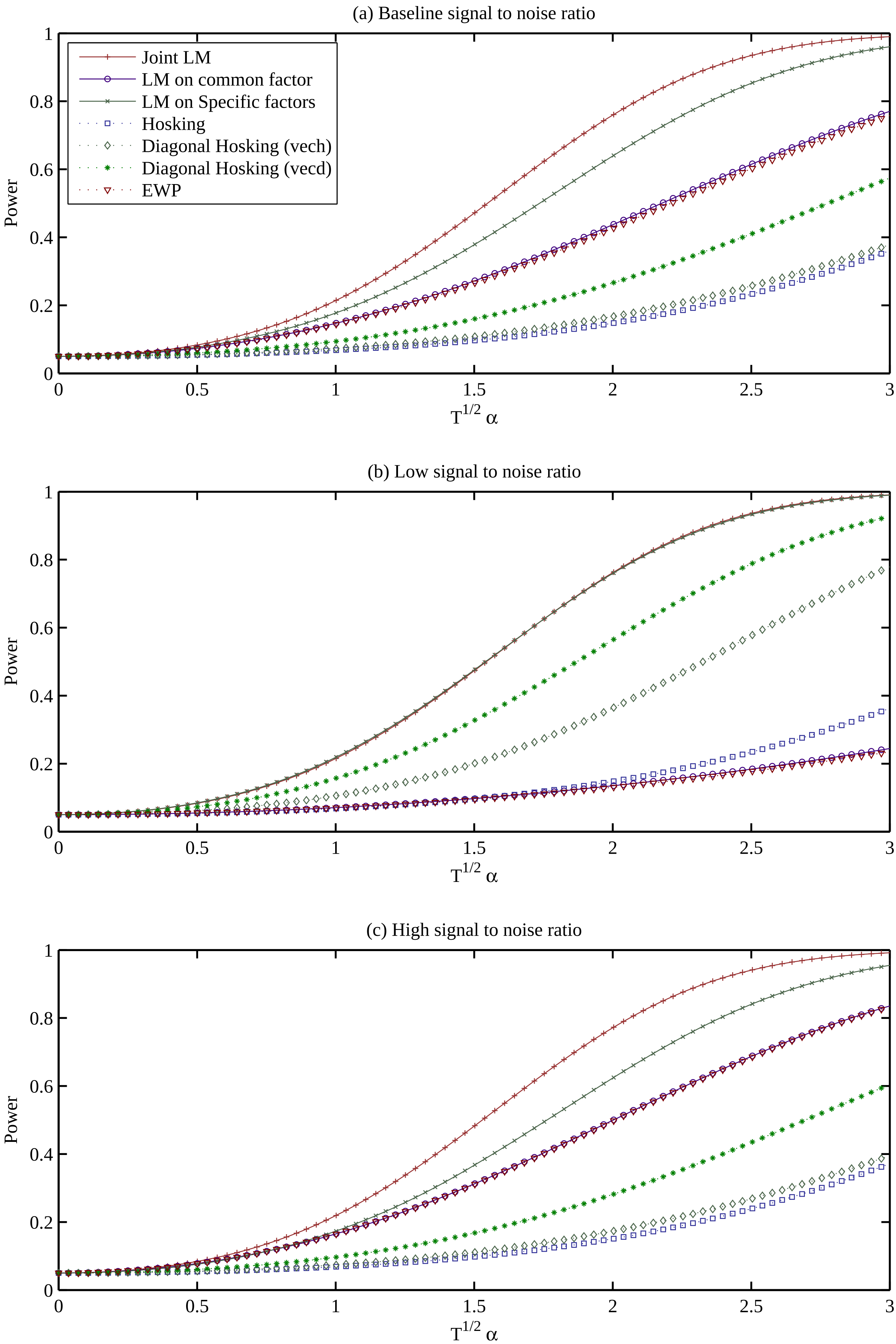
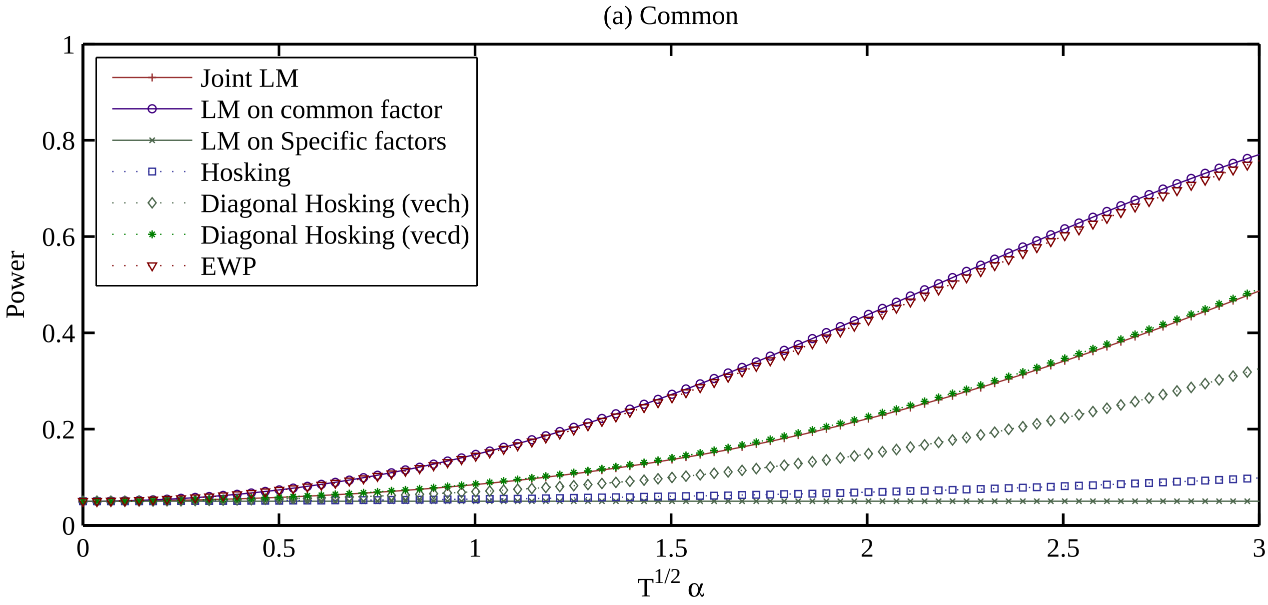


Figure 4: Power of variance dependence tests at 5% level against local alternatives



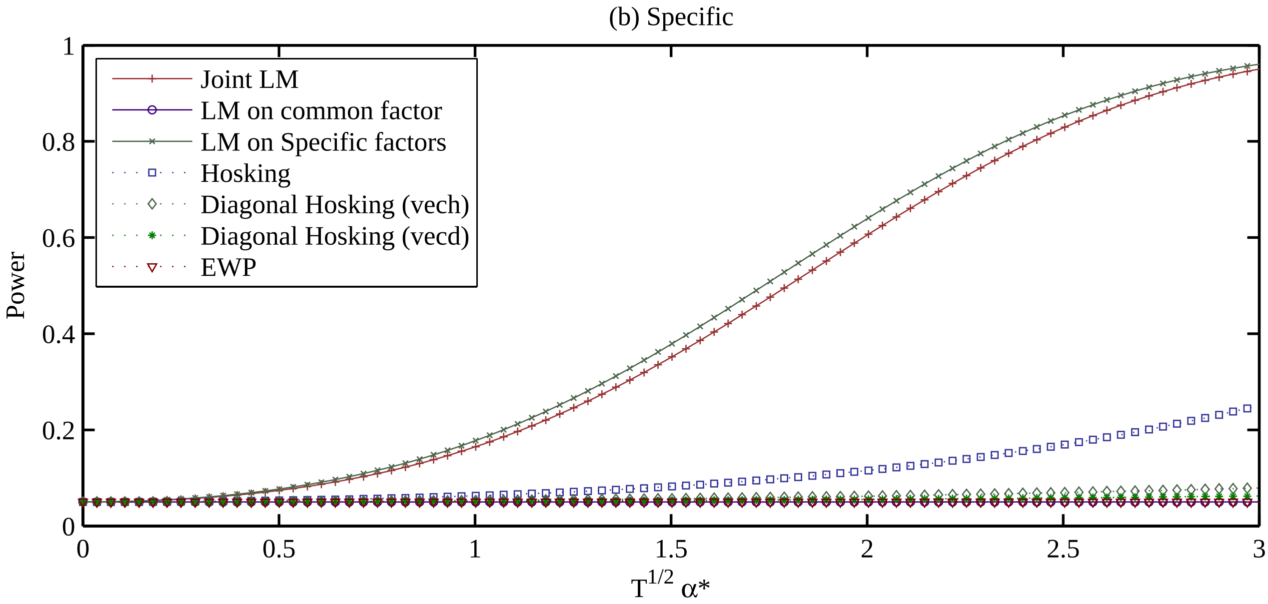
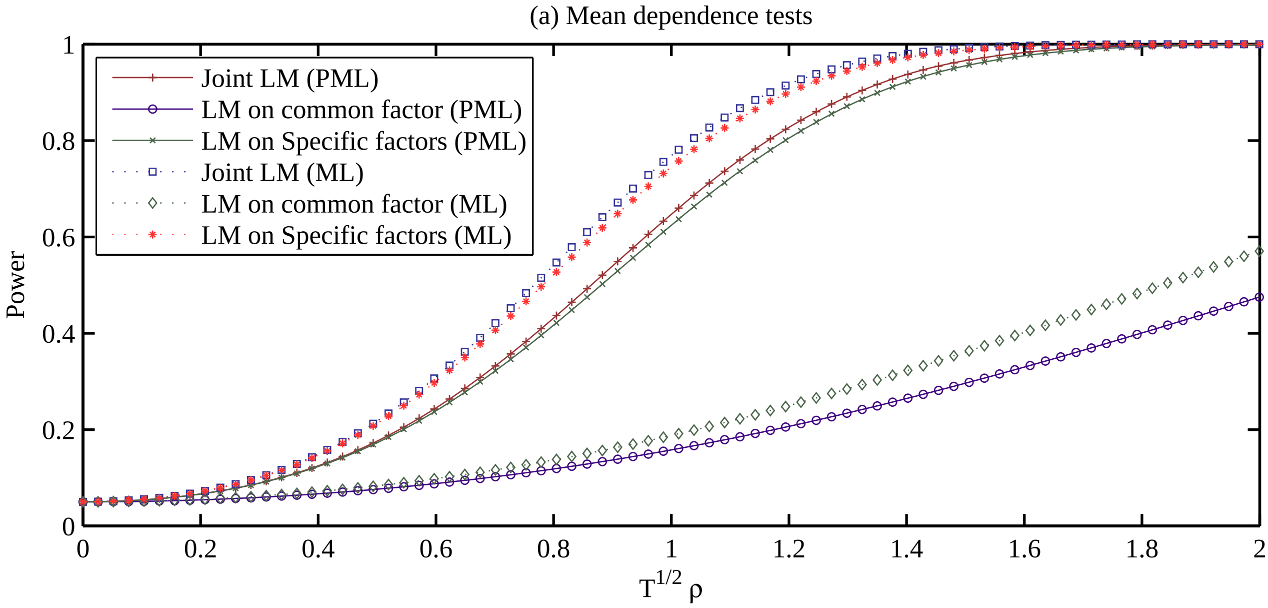


Figure 5: Power of serial dependence tests at 5% level against local alternatives
DGP: Student t with 6 df



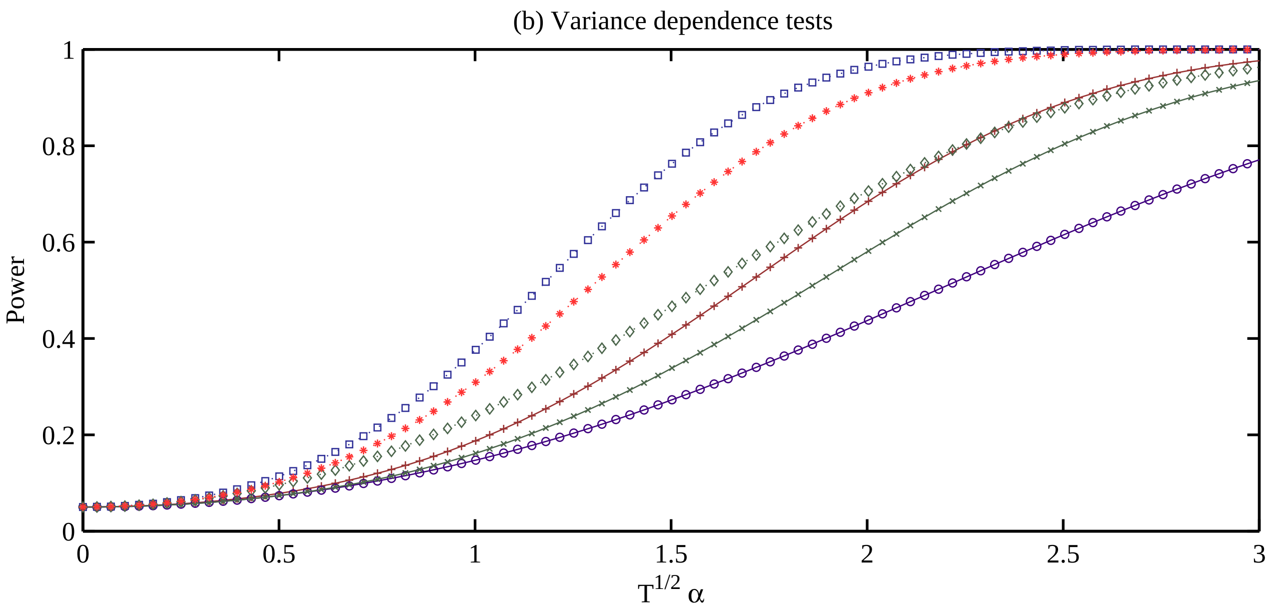


Figure 6: P−value discrepancy plots. Tests against AR(1) alternatives.

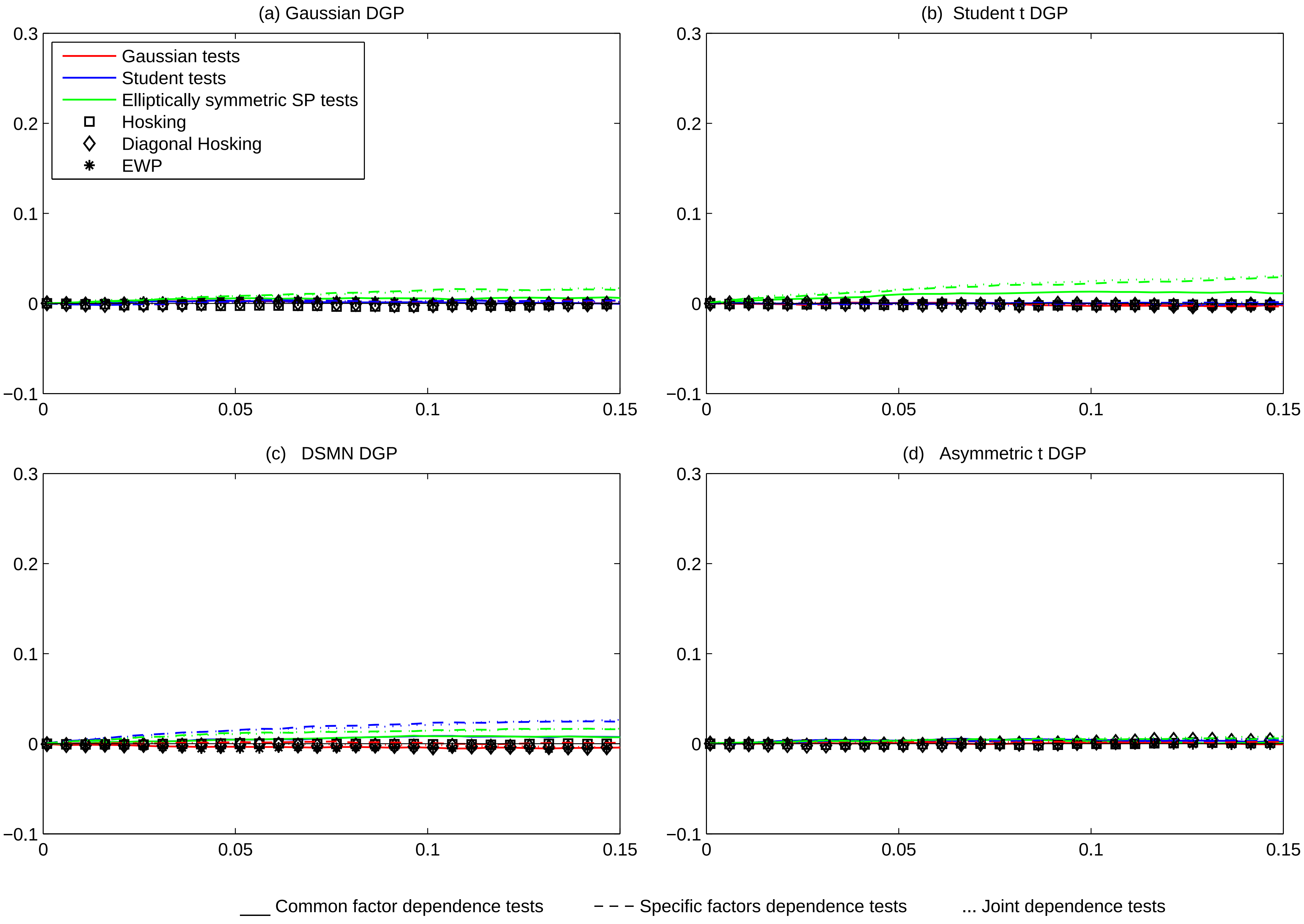


Figure 7: P−value discrepancy plots. Tests against ARCH(1) alternatives.

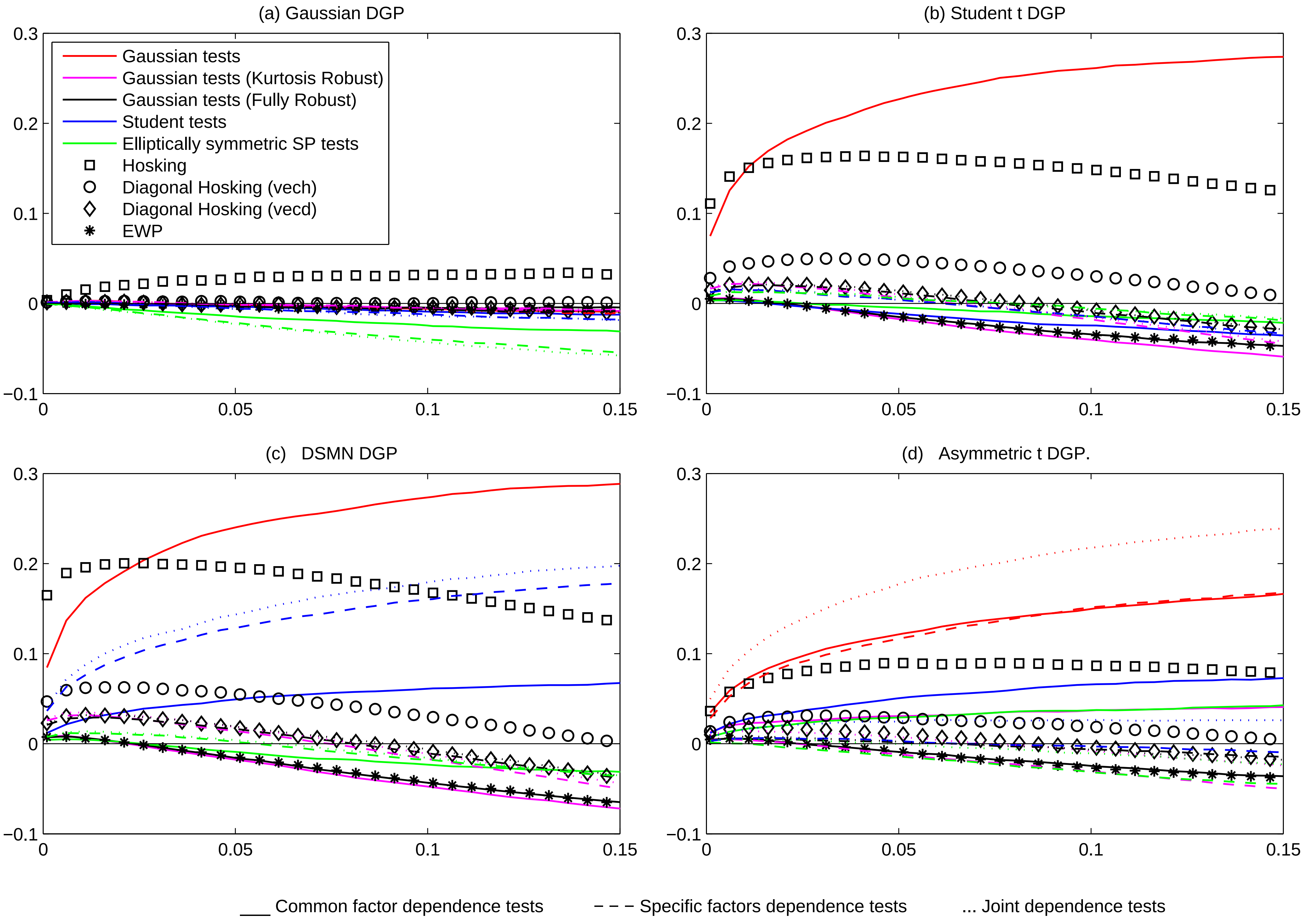


Figure 8: P−value discrepancy plots. Tests against GARCH(1,1) alternatives.