Normality tests for latent variables*

Tincho Almuzara
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<almuzara@cemfi.edu.es>

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

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

February 2016

Preliminary and incomplete

Abstract

We exploit the rationale behind the Expectation Maximisation algorithm to derive simple to implement and interpret score tests of normality in all or a subset of the innovations to the latent variables in state space models against Generalised Hyperbolic alternatives, including symmetric and asymmetric Student *t*. We decompose our tests into third and fourth moment components, and obtain one-sided Likelihood Ratio analogues, whose asymptotic distribution we provide. We perform a Monte Carlo study of the finite sample size and power of our procedures and previous proposals. Finally, we illustrate our tests in an application to US aggregate real output measurement.

Keywords: Kurtosis, Kuhn-Tucker test, LM tests, Skewness, Supremum test, Underidentified parameters, Wiener-Kolmogorov-Kalman smoother.

JEL: C32, C52, C12.

*Financial support from the Spanish Ministry of Economy and Competitiveness through grant ECO 2014-59262 is gratefully acknowledged.

1 Introduction

Latent variable models that relate a set of observed variables to a meaningful set of unobserved influences are widely used in many applied fields. Unlike in other social sciences, however, the majority of economic and financial applications maintain the assumption that the underlying state variables are Gaussian. Two prominent examples are factor analysis and dynamic state space models. The list of empirical studies that make use of those models is vast.

While sophisticated users will often look at several diagnostics, formal tests are hardly ever reported in empirical work. One particularly relevant issue is the normality of the underlying variables, which implies the normality of the observed variables, which in turn justifies the estimation of the unknown model parameters by maximum likelihood. Although there are many readibly available normality tests, they are designed to be directly applied to the observed variables in static models or their one-period ahead prediction errors in dynamic ones.

The objective of our paper is precisely to derive simple to implement and interpret tests for non-normality in all or a subset of (the innovations to) the state variables. We focus on Lagrange Multiplier (LM) tests, which only require estimation of the model under the null. As is well known, Likelihood ratio (LR), Wald and LM tests are asymptotically equivalent under the null and sequences of local alternatives, and therefore they share their optimality properties. Aside from non-trivial computational reasons, the advantage of LM tests is that rejections provide a clear indication of the directions along which modelling efforts should focus.

The complication that we face, though, is that the density function of the observed variables or their innovations is typically unknown when the distribution of the latent variables is not Gaussian, and in many cases it can only be approximated by simulation (see Durbin and Koopman (2012) for an extensive discussion in the context of dynamic models). As a result, the log-likelihood function under the alternative, its score and information matrix can seldom be obtained in closed form despite the fact that we can compute the true log-likelihood function under the Gaussian null. In this context, we use the Expectation Maximisation (EM) principle to obtain the scores of the parameters that characterise departures from normality. The EM algorithm studied in Dempster, Laird and Rubin (1977) is a well known procedure for obtaining maximum likelihood estimates in both static and dynamic latent variable models (see e.g. Rubin and Thayer (1982) or Watson and Engle (1983), respectively). However, to the best of our knowledge it has only been used for testing purposes by Fiorentini and Sentana (2015), who employ it to assess the dynamic mean and variance specification of factor models.

Our approach introduces a relatively minor complication: the influence functions that constitute the basis of our tests are serially correlated in dynamic models. In this regard, our methods are related to Bai and Ng (2005) and Bontemps and Meddahi (2005), who derive moment-based normality tests for a single observed variable or its innovations in potentially serially correlated contexts by relying on heteroskedastic and autocorrelation consistent estimators of the asymptotic variances. Nevertheless, we derive analytical expressions for the autocovariance matrices of the influence functions, which we would expect a priori to lead to more reliable finite sample sizes for our statistics than their non-parametric counterparts. For that reason, our approach is more closely related to Harvey and Koopman (1992), who apply standard univariate normality tests for observed variables to the smoothed values of the innovations in the underlying components of a univariate random walk plus noise model explicitly taking into account the serial correlation in those filtered estimates implied by the model. Unlike us, though, neither of those authors justify their procedures by appealing to the likelihood principle.

For most practical purposes, departures from normality can be attributed to two different sources: excess kurtosis and skewness. In this sense, we follow Mencía and Sentana (2012) in considering Generalised Hyperbolic (GH) alternatives, which include the symmetric and asymmetric Student t, normal-gamma mixtures, hyperbolic, normal inverse Gaussian and symmetric and asymmetric Laplace distributions. The main advantage of these GH alternatives is that the number of third and fourth moments which are effectively tested is proportional to the number of series involved, unlike tests against Hermite expansions of the multivariate normal density (see Amengual and Sentana (2015) for a comparison in the context of copulas).

Importantly, we show that our tests are not affected by the sampling variability in the model parameters estimated under the null, which implies that we can treat them as if they were known.

The rest of the paper is organised as follows. Section 2 describes the econometric model, as well as the GH alternatives. We derive the normality tests in section 3. Section 4 presents the results of our Monte Carlo experiments. Finally, we include our empirical application in section 5, followed by our conclusions. Proofs and auxiliary results can be found in appendices.

2 The model

2.1 Linear state space models

A linear, time-invariant, parametric state-space model for a finite dimensional vector of N observed series, $\mathbf{y}_t$, can be recursively defined in the time domain by the system of stochastic difference equations

$$\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\xi}_t \tag{1}$$

$$\boldsymbol{\xi}_t = \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} + \mathbf{M}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^* \tag{2}$$

$$\boldsymbol{\varepsilon}_t^*|\mathcal{I}_{t-1};\boldsymbol{\phi} \sim iid\ D(\mathbf{0},\mathbf{I},\boldsymbol{\eta}) \tag{3}$$

where $\boldsymbol{\phi} = (\boldsymbol{\pi}', \boldsymbol{\theta}', \boldsymbol{\eta}')'$, $\boldsymbol{\pi}$ is the mean vector of the observed series, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is a vector of p additional second moment parameters, $\mathbf{H} : \Theta \to \mathbb{R}^{N\times M}$, $\mathbf{F} : \Theta \to \mathbb{R}^{M\times M}$ and $\mathbf{M} : \Theta \to \mathbb{R}^{M\times K}$ are matrix valued functions of coefficients, many of whose elements will typically be either 0 or 1, $\boldsymbol{\xi}_t$ is an M-dimensional vector of state variables, $\boldsymbol{\varepsilon}_t^*$ is an K-dimensional vector of standardised structural *iid* innovations driving those variables whose distribution depends on a vector of shape parameters $\boldsymbol{\eta}$, and $\mathcal{I}_{t-1}$ is an information set that contains the values of $\mathbf{y}_t$ and $\boldsymbol{\xi}_t$ up to and including $t-1$.

We assume that $N \leq K \leq M$ to avoid dynamic singularities. We also assume that the model above is correctly specified, in the sense that there is some $\boldsymbol{\theta}_0$ for which (1) and (2) constitute the true data generating process of $\{\mathbf{y}_t, \boldsymbol{\xi}_t\}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In this context, static models will be such that $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}$ for all $\boldsymbol{\theta}$.

There are multiple alternative representations of state-space models,[1] but in this paper we follow the one in Harvey (1989), except that we have deliberately subsumed any possible error in the measurement equation (1) into the state vector so as to be able to test for normality not only in the minimal possible set of state variables but also in the measurement errors. For that reason, equations (1) and (2) closely resemble the usual state representation in the engineering literature, in which the elements of $\boldsymbol{\varepsilon}_t^*$ would be regarded as control variables (see Anderson and Moore (1979)). For ease of exposition, we do not look at models with exogeneous regressors or those in which some of the system matrices are deterministic functions of time or observable

[1] For example, Durbin and Koopman (2012) shift the transition equation (2) forward by one period, as in Anderson and Moore (1979), and include measurement errors in (1), which they assume are orthogonal to the innovations in the state variables. On the other hand, Komunjer and Ng (2011) substitute the transition equation (2) into the measurement equation (1), thereby creating an alternative measurement equation whose innovations are perfectly correlated with the innovations in the transition equation.

predermined variables.[2]

We also assume without loss of generality that the columns of the matrix $\mathbf{M}(\boldsymbol{\theta})$ are linearly independent so that there are no redundant elements in $\boldsymbol{\varepsilon}_t^*$. Typically, $\mathbf{M}(\boldsymbol{\theta})$ will be a selection matrix whose columns are (proportional to) vectors of the M dimensional canonical basis, but in principle they could be different. As a result, we can uniquely recover $\boldsymbol{\varepsilon}_t^*$ from $\boldsymbol{\xi}_t$ as

$$\boldsymbol{\varepsilon}_t^* = \mathbf{M}^+(\boldsymbol{\theta}_0)[\mathbf{I}_M - \mathbf{F}(\boldsymbol{\theta}_0)L]\boldsymbol{\xi}_t, \tag{4}$$

where $\mathbf{M}^+(\boldsymbol{\theta}) = [\mathbf{M}'(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})]^{-1}\mathbf{M}'(\boldsymbol{\theta})$ denotes the Moore-Penrose inverse of $\mathbf{M}(\boldsymbol{\theta})$.

Finally, we assume that the researcher makes sure that the model parameters $\boldsymbol{\theta}$ are identified before estimating the model, which often requires restrictions on the system matrices.

2.2 The reduced form

Assuming covariance stationarity, possibly after some suitable transformation, we can find the autocorrelation structure of the observed series generated by (1)-(2), as well as the corresponding Wold representation, which will typically resemble a VARMA model, with potentially long but finite AR and MA orders, but restricted coefficient matrices because $M \geq N$.

As a result, we will be able to write

$$(\mathbf{y}_t - \boldsymbol{\pi}) = \sum_{j=1}^{p_y} \mathbf{A}_j(\boldsymbol{\theta})(\mathbf{y}_{t-j} - \boldsymbol{\pi}) + \mathbf{w}_t + \sum_{j=1}^{q_y} \mathbf{B}_j(\boldsymbol{\theta})\mathbf{w}_{t-j},$$

where $\mathbf{w}_t$ is a serially uncorrelated sequence linearly unpredictable on the basis of lagged values of $\mathbf{y}_t$. In fact, assuming that the Wold representation is strictly invertible,

$$\mathbf{w}_t = \left[\mathbf{I}_N + \sum_{j=1}^{q_y} \mathbf{B}_j(\boldsymbol{\theta})L^j\right]^{-1} \left[\mathbf{I}_N - \sum_{j=1}^{q_y} \mathbf{A}_j(\boldsymbol{\theta})L^j\right] (\mathbf{y}_t - \boldsymbol{\pi}). \tag{5}$$

This relationship will prove useful when we compare our tests, which target the components in $\boldsymbol{\varepsilon}_t^*$ directly, to existing tests, which target $\mathbf{w}_t$ instead. If $\boldsymbol{\varepsilon}_t^*|\mathcal{I}_{t-1}$ is *iid* normal, then $\mathbf{y}_t$ will be a Gaussian process, and therefore $\mathbf{w}_t|\mathcal{I}_{t-1}$ will be *iid* normal too. However, if some elements of $\boldsymbol{\varepsilon}_t^*$ are not normal, then the unconditional distribution of $\mathbf{w}_t$ will typically be extremely complicated,[3] although we would expect it to be closer to a normal than $\boldsymbol{\varepsilon}_t^*$ because of the averaging implicit in (5). We will revisit this issue in section 3.5.2.

Next, we briefly outline three special cases that we will study in more detail in sections 4 and 5.

[2] Minor changes to the testing procedures we propose will render them applicable to those situations.
[3] In general $\mathbf{w}_t$ will be neither *iid* nor even a martingale difference sequence with respect to $\mathcal{I}_{t-1}$.

2.3 Three examples

2.3.1 Static factor models

We start by considering a single factor version of a traditional (i.e. static, conditionally homoskedastic and exact) factor model, which suffices to illustrate our main results. Specifically,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\pi} + \mathbf{c} f_t + \mathbf{v}_t, \\ \left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array} \right) \Bigg| \mathcal{I}_{t-1}; \boldsymbol{\phi} &\sim iid\ D\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array} \right), \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array} \right), \boldsymbol{\eta} \right], \end{aligned} \tag{6}$$

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 an $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to f_t, $\boldsymbol{\Gamma}$ is an $N \times N$ diagonal positive definite (p.d.) matrix of constant idiosyncratic variances, and $\boldsymbol{\theta} = (\mathbf{c}', \boldsymbol{\gamma}')'$, with $\boldsymbol{\gamma} = vecd(\boldsymbol{\Gamma})$.

A non-trivial advantage of static factor models is that they automatically guarantee a p.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 (6) 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).

We can easily express model (6) as in (1)–(2) with $\boldsymbol{\xi}_t = (f_t, \mathbf{v}_t')'$, $\mathbf{H}(\boldsymbol{\theta}) = (\mathbf{c}, \mathbf{I}_N)$, $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}$,

$$\mathbf{M}(\boldsymbol{\theta}) = \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & diag^{1/2}(\boldsymbol{\gamma}) \end{array} \right)$$

and $\boldsymbol{\varepsilon}_t^* = (f_t, \mathbf{v}_t^{*\prime})'$, with $\mathbf{v}_t^* = \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_t$. Notice that this specification trivially implies that

$$\mathbf{y}_t | \mathcal{I}_{t-1}; \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\eta} \sim iid\ D^*[\boldsymbol{\pi}, \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\eta}], \quad \text{with} \quad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{c}\mathbf{c}' + \boldsymbol{\Gamma},$$

so that $\mathbf{w}_t = \mathbf{y}_t - \boldsymbol{\pi}$. While the normality of $\boldsymbol{\xi}_t$ implies the normality of $\mathbf{y}_t$, in principle the distribution of $\mathbf{y}_t$ and $\boldsymbol{\xi}_t$ will be different under the alternative.

2.3.2 Univariate unobserved components models

A UCARIMA model for a univariate observed series, y_t can be defined by the equations:

$$\begin{aligned}
y_t &= \pi + x_t + \epsilon_t \\
\alpha_x(L)x_t &= \beta_x(L)f_t \\
\alpha_\epsilon(L)\epsilon_t &= \beta_\epsilon(L)v_t \\
\left(\begin{array}{c} f_t \\ v_t \end{array}\right)\Bigg|\mathcal{I}_{t-1};\boldsymbol{\phi} &\sim iid\ D\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma_f^2 & 0 \\ 0 & \sigma_v^2 \end{array}\right), \boldsymbol{\eta}\right],
\end{aligned}$$

where x_t is the "signal" component, v_t the orthogonal "non-signal" component, $\alpha_x(L)$ and $\alpha_\epsilon(L)$ are one-sided polynomials of orders p_x and p_ϵ, respectively, while $\beta_x(L)$ and $\beta_\epsilon(L)$ are one-sided polynomials of orders q_x and q_ϵ coprime with $\alpha_x(L)$ and $\alpha_\epsilon(L)$, respectively, and $\boldsymbol{\theta}$ refers to the model parameters that characterise the autocovariance structure of the observed series.[4]

In section 4 we consider two simple but popular examples: (i) the AR(1) signal plus noise and (ii) the so-called local-level model, both of which assume that

$$\alpha_x(L) = 1 - \alpha L, \text{ and } \alpha_\epsilon(L) = \beta_x(L) = \beta_\epsilon(L) = 1,$$

the only difference being that in the former $|\alpha| < 1$ while in the latter $\alpha = 1$.

Once again, we can write the basic UCARIMA model above as in (1)–(2) by defining $\boldsymbol{\xi}_t = (x_t, \epsilon_t)'$, $\mathbf{H}(\boldsymbol{\theta}) = (1, 1)$,

$$\mathbf{F}(\boldsymbol{\theta}) = \left(\begin{array}{cc} \alpha & 0 \\ 0 & 0 \end{array}\right), \quad \mathbf{M}(\boldsymbol{\theta}) = \left(\begin{array}{cc} \sigma_f & 0 \\ 0 & \sigma_v \end{array}\right)$$

and $\boldsymbol{\varepsilon}_t^* = (f_t^*, v_t^*)'$, with $f_t^* = f_t/\sigma_f$ and $v_t^* = v_t/\sigma_v$.

The reduced form model will be an ARMA process with maximum orders $p_y = p_x + p_\epsilon$ for the AR polynomial $\alpha_y(.) = \alpha_x(.)\alpha_\epsilon(.)$ and $q_y = \max(p_x + q_\epsilon, q_x + p_\epsilon)$ for the MA polynomial $\beta_y(.)$. Cancellation will trivially occur when $\alpha_x(.)$ and $\alpha_\epsilon(L)$ share c common roots, but there could also be other cases (see Granger and Morris (1976) for further details). The coefficients of $\beta_y(L)$, as well as σ_w^2, which is the variance of the univariate Wold innovations, w_t, are obtained by matching autocovariances (see Fiorentini and Planas (1998) for a comparison of numerical methods). Assuming strict invertibility of the MA part, we could then obtain the reduced form innovations by the inverse ARMA represention (5) or from the prediction equations of the Kalman filter without making use of the expressions for $\alpha_y(.)$ or $\beta_y(.)$.

[4] See Fiorentini and Sentana (2015) and the references therein for the identification conditions for this model.

2.3.3 A cointegrated, dynamic single factor model

Consider the following bivariate model

$$
\begin{aligned}
\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} x_t + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \\
(1-\alpha_x L)(\Delta x_t - \mu) &= f_t \\
(1-\alpha_{\epsilon_1} L)(\epsilon_{1,t} - \delta/2) &= v_{1,t} \\
(1-\alpha_{\epsilon_2} L)(\epsilon_{2,t} + \delta/2) &= v_{2,t} \\
\left.\begin{pmatrix} f_t \\ v_{1,t} \\ v_{2,t} \end{pmatrix}\right| \mathcal{I}_{t-1};\boldsymbol{\phi} &\sim iid\ D\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & 0 & 0 \\ 0 & \sigma_{v_1}^2 & 0 \\ 0 & 0 & \sigma_{v_2}^2 \end{pmatrix}, \boldsymbol{\eta}\right].
\end{aligned}
$$

where x_t is the common factor, which follows an Arima(1,1,0) with autoregressive coefficient α_x, drift $\mu(1-\alpha_x)$ and innovation variance σ_f^2, $\epsilon_{i,t}$ for $i = 1,2$ are the idiosyncratic shocks, which follow covariance stationary Ar(1) processes with autoregressive coefficients α_{ϵ_1} and α_{ϵ_2}, unconditional means $\delta/2$ and $-\delta/2$, respectively, and innovation variances $\sigma_{v_1}^2$ and $\sigma_{v_2}^2$.[5]

This model, which will form the basis of our empirical application in section 5, has two important features: (i) $y_{1,t}$ and $y_{2,t}$ are cointegrated both with x_t and between themselves, with cointegrating vector (1,-1); and (ii) all the state variables are serially correlated.

In terms of the formulation (1)–(2), we have that $\boldsymbol{\pi} = (\delta/2, -\delta/2)'$, $\boldsymbol{\xi}_t = (1, x_t, x_{t-1}, \epsilon_{1,t}, \epsilon_{2,t})'$,

$$
\mathbf{H}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mu(1-\alpha_x) & 1+\alpha_x & -\alpha_x & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{\epsilon_1} & 0 \\ 0 & 0 & 0 & 0 & \alpha_{\epsilon_2} \end{pmatrix},
$$

$$
\mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 \\ \sigma_f & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_{v_1} & 0 \\ 0 & 0 & \sigma_{v_2} \end{pmatrix}
$$

and $\boldsymbol{\varepsilon}_t^* = (f_t^*, v_{1,t}^*, v_{2,t}^*)'$, with $f_t^* = f_t/\sigma_f$, $v_{1,t}^* = v_{1,t}/\sigma_{v_1}$ and $v_{2,t}^* = v_{2,t}/\sigma_{v_2}$.

To obtain the reduced form of the model, it is convenient to work with a stationarity inducing transformation that premultiplies the system by the matrix

$$
\mathbf{A}(L) = \begin{pmatrix} 1-L & 0 \\ 1 & -1 \end{pmatrix}.
$$

[5] The restrictions that the means of the idiosyncratic shocks coincide in magnitude but not in sign simply reflects the fact that we can only identify δ from the mean of the error correction term $y_{1t} - y_{2t}$.

Thus, we end up with

$$\begin{pmatrix} \Delta y_{1t} \\ y_{1t} - y_{2t} \end{pmatrix} = \begin{pmatrix} \Delta x_t \\ 0 \end{pmatrix} + \begin{pmatrix} \Delta\epsilon_{1t} \\ \Delta\epsilon_{1t} - \Delta\epsilon_{2t} \end{pmatrix}.$$

It is then easy to see that Δy_{1t} follows a UCARIMA model like the one in section 2.3.2, while $y_{1t} - y_{2t}$ is the linear combination of two ARMA(1,1) processes, and therefore an ARMA process itself. As a result, the reduced form will be a VARMA process.

2.4 Null and alternative hypotheses

As we mentioned in the introduction, the most common assumption in the literature by far is that the innovations in the state variables are normally distributed, which gives rise to a Gaussian process for $\mathbf{y}_t$. In that case, the Kalman filter yields the optimal filter for the state variables and Gaussian maximum likelihood is fully efficient. In contrast, if the innovations are not Gaussian, the Kalman filter only provides the best linear filter, and Gaussian Pseudo ML provides at best consistent estimators of the model parameters.

In this paper we derive computationally simple tests of the null hypothesis that all the structural innovations are Gaussian against the alternative that they follow a member of the Generalised Hyperbolic family of distributions introduced by Barndorff-Nielsen (1977) and studied in detail by Blæsild (1981). This is a rather flexible family of multivariate distributions that nests not only the normal and Student t but also many other examples such as the asymmetric Student t, the hyperbolic and normal inverse Gaussian distributions, as well as symmetric and asymmetric versions of the normal-gamma mixture and Laplace.

We can gain some intuition about the *GH* distribution by considering its interpretation as a location-scale mixture of normals in which the mixing variable is a Generalised Inverse Gaussian (*GIG*). If $\boldsymbol{\varepsilon}^*$ is a *GH* vector, then it can be expressed as

$$\boldsymbol{\varepsilon}^* = \boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\zeta^{-1} + \zeta^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\boldsymbol{\varepsilon}^\circ, \tag{7}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^K$, $\boldsymbol{\Upsilon}$ is a symmetric positive definite matrix of order K, $\boldsymbol{\varepsilon}^\circ \sim iid\, N(\mathbf{0}, \mathbf{I}_K)$ and the positive mixing variable ζ is an independent *iid GIG* with parameters $-\nu$, γ and δ, or $\zeta \sim GIG(-\nu, \gamma, \delta)$ for short, where $\nu \in \mathbb{R}$ and $\gamma, \delta \in \mathbb{R}^+$ (see Jørgensen (1982) and Johnson, Kotz, & Balakrishnan (1994) for details). Since $\boldsymbol{\varepsilon}^*$ given ζ is Gaussian with conditional mean $\boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\zeta^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\zeta^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. There is a further scale parameter, δ, two other scalars, ν

and γ, to allow for flexible tail modelling, and the vector $\boldsymbol{\beta}$, which introduces skewness in this distribution. In this sense, the distribution of $\boldsymbol{\varepsilon}^*$ becomes a simple scale mixture of normals, and thereby spherical, when $\boldsymbol{\beta}$ is zero. Mencía and Sentana (2012) set $\delta = 1$ and derive restrictions on $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ which ensure that the elements of $\boldsymbol{\varepsilon}^*$ are uncorrelated with zero means and unit variances. They also find it analytically convenient to replace ν and γ by η and ψ, where $\eta = -.5\nu^{-1}$ and $\psi = (1+\gamma)^{-1}$, although we continue to use ν and γ in some equations for notational simplicity. Thus, in what follows we identify the vector of shape parameters $\boldsymbol{\eta}$ with $(\eta, \psi, \boldsymbol{\beta}')'$, so that $\boldsymbol{\eta} = \mathbf{0}$ corresponds to the Gaussian null.

In many applications, though, the researcher may only be interested in testing whether the source of non-normality comes from a subset of the underlying components, which have some meaningful interpretation. Given that we can always re-order the vector of structural innovations $\boldsymbol{\varepsilon}_t^*$ and postmultiply the matrix $\mathbf{M}(\boldsymbol{\theta})$ by a permutation matrix, without loss of generality we can assume that the non-Gaussian distribution is confined to the first $R \leq K$ innovations under the alternative. Henceforth, we refer to the relevant components as $\boldsymbol{\varepsilon}_t^{*\mathrm{GH}} = \mathbf{E}_{RK}\boldsymbol{\varepsilon}_t^*$, with $\mathbf{E}_{RK} = (\mathbf{I}_R, \mathbf{0}_{R\times(K-R)})$, and to the remaining component as $\boldsymbol{\varepsilon}_t^{*\mathrm{N}}$.

Finally, we may also envisage an alternative situation in which the elements of $\boldsymbol{\varepsilon}_t^*$ are cross-sectionally independent but non-Gaussian.

As a result, we explicitly consider the following three alternative hypotheses:

1. The joint distribution of all structural innovations is *GH*: $H_J : \boldsymbol{\varepsilon}_t^* \sim GH_K(\eta, \psi, \boldsymbol{\beta})$;

2. The joint distribution of the first R structural innovations is *GH* while the rest are Gaussian: $H_S : \boldsymbol{\varepsilon}_t^{*\mathrm{GH}} \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, $\boldsymbol{\varepsilon}_t^{*\mathrm{N}} \sim N_{K-R}(\mathbf{0}; \mathbf{I})$;

3. Each structural innovation is independently distributed as a univariate *GH*: $H_I : \varepsilon_{it}^* \sim GH(\eta_i, \psi_i, \beta_i)$, for $i = 1, ..., K$.

3 Normality tests for latent variables

Before presenting our main results, we introduce some additional notation. We define $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ as the value of $\boldsymbol{\varepsilon}_t^*$ generated by the right hand side of equation (4) evaluated at $\boldsymbol{\theta} \in \Theta$. Similarly, $\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})$ denotes the smoothed (filtered) values of the innovations at t given Y_T, which contains past, present and future values of the observed series, and $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})$ the corresponding mean-square error, so that $\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}) = \mathbf{E}_{RK}\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}) = \mathbf{E}_{RK}\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})\mathbf{E}_{RK}'$. Finally, we define

$$
\begin{aligned}
\mathbf{m}_{1,t|T}(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}), \\
\mathbf{m}_{2,t|T}(\boldsymbol{\theta}) &= vec\left[\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})'\right], \\
\mathbf{m}_{3,t|T}(\boldsymbol{\theta}) &= vec\left\{\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})]'\right\}, \text{ and} \\
\mathbf{m}_{4,t|T}(\boldsymbol{\theta}) &= vec\left\{[\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})][\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})]'\right\}.
\end{aligned}
\tag{8}
$$

3.1 The score under Gaussianity

LM tests are usually obtained from the score associated to the (marginal) likelihood function of the observed variables, $f_T(Y_T; \boldsymbol{\phi})$ say, evaluated under the Gaussian null. Unfortunately, the functional form of $f_T(Y_T; \boldsymbol{\phi})$ is generally unknown under the alternative, and consequently the same is true of its score vector under the null despite the fact that we can easily compute the Gaussian likelihood function. For that reason, we rely on the EM principle.

Let $f_T(Y_T, \Xi_T; \boldsymbol{\phi})$ denote the (joint) likelihood function for both observed and state variables of model (1)–(2) for a sample of size T and $f_T(\Xi_T|Y_T; \boldsymbol{\phi})$ the conditional likelihood function of the latent variables given the observed ones. Since the Kullback inequality implies that $E\left[\ln f_T(\Xi_T|Y_T; \boldsymbol{\phi})|Y_T; \boldsymbol{\phi}\right] = 0$, it follows that we can obtain $\partial \ln f_T(Y_T; \boldsymbol{\phi})/\partial \boldsymbol{\phi}$ as the expected value (given Y_T and $\boldsymbol{\phi}$) of the the unobservable score corresponding to $f_T(\Xi_T|Y_T; \boldsymbol{\phi})$. Specifically,

$$
\frac{\partial \ln f_T(Y_T; \boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = E\left[\left.\frac{\partial \ln f_T(Y_T, \Xi_T; \boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right| Y_T\right]. \tag{9}
$$

This result was first noted by Louis (1982); see also Ruud (1991) and Tanner (1996).

But we still face two additional difficulties. First, as discussed by Mencía and Sentana (2012), there are three different paths along which a symmetric GH distribution converges to a Gaussian distribution. Specifically, the normal distribution can be achieved when (i) $\nu \to -\infty$ or (ii) $\nu \to +\infty$, regardless of the value of γ; and (iii) $\gamma \to \infty$ irrespective of the value of ν. In addition, one of the shape parameters becomes increasingly underidentified when the other one is on a normality path. Fortunately, Mencía and Sentana (2012) showed that the score of the remaining identified parameters evaluated under the null of normality are proportional amongh them along those three paths. As a result, we can focus on deriving the relevant quantities as $\tau = \eta \cdot \psi \to 0$.

Second, the derivation of the Lagrange multiplier (LM) test for a multivariate normal versus an asymmetric GH is complicated by the fact that $\boldsymbol{\beta}$ drops out from both the joint and marginal distributions when $\tau \to 0$ (see Mencía and Sentana (2012)).

One standard solution in the literature to deal with testing situations with underidentified parameters under the null involves fixing those parameters to some arbitrary values, and then computing the appropriate test statistic for the chosen values. Thus, we obtain:

Proposition 1 *The score of the asymmetric GH with respect to the parameter τ when $\tau = 0$ for fixed values of the skewness parameters $\boldsymbol{\beta}$ is given by*

$$\begin{aligned} \bar{s}_{\mathrm{GH},T}(\boldsymbol{\theta},\boldsymbol{\beta}) &= \frac{1}{T}\sum_{t=1}^{T}\left[s_{\mathrm{k},t|T}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})\right], \qquad (10)\\ s_{\mathrm{k},t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{k},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{k},t|T}(\boldsymbol{\theta}),\\ \mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{s},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{s},t|T}(\boldsymbol{\theta}), \end{aligned}$$

where

$$\mathbf{m}_{\mathrm{k},t|T}(\boldsymbol{\theta}) = \begin{pmatrix} 1 \\ \mathbf{m}_{2,t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{4,t|T}(\boldsymbol{\theta}) \end{pmatrix}, \quad \mathbf{b}_{\mathrm{k},t|T}(\boldsymbol{\theta}) = \begin{pmatrix} b_{0,t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{2,t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{4,t|T}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{m}_{\mathrm{s},t|T}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{m}_{1,t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{3,t|T}(\boldsymbol{\theta}) \end{pmatrix}, \quad \mathbf{b}_{\mathrm{s},t|T}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{b}_{1,t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{3,t|T}(\boldsymbol{\theta}) \end{pmatrix},$$

$$b_{0,t|T}(\boldsymbol{\theta}) = c_0 + \{c_1 + c_2 tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\} tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})] + 2c_2 tr\{[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]^2\},$$

$$\mathbf{b}_{1,t|T}(\boldsymbol{\theta}) = [c_3 + tr(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\mathbf{E}'_{RK} + 2\mathbf{E}'_{RK}\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}),$$

$$\mathbf{b}_{2,t|T}(\boldsymbol{\theta}) = \{c_1 + 2c_2 tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\}\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right) vec(\mathbf{I}_R) + 4c_2\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right) vec[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})],$$

$$\mathbf{b}_{3,t|T}(\boldsymbol{\theta}) = \mathbf{E}'_{RK}\boldsymbol{\iota}_R \otimes \mathbf{E}'_{RK},$$

$$\mathbf{b}_{4,t|T}(\boldsymbol{\theta}) = c_2\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right)\boldsymbol{\iota}_{R^2},$$

with $c_0 = R(R+2)/4$, $c_1 = -(R+2)/2$, $c_2 = 1/4$, $c_3 = -(R+2)$ and $\boldsymbol{\iota}_H$ a vector of H ones.

This result provides an intuitive interpretation for $s_{\mathrm{GH},t|T}(\boldsymbol{\theta},\boldsymbol{\beta})$ as a linear combination of a kurtosis component, $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$, and N skewness components, $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$.

We could then develop the associated test statistic, $LM_T(\boldsymbol{\theta},\boldsymbol{\beta})$ by finding the asymptotic variance for (10). However, while such an approach is plausible in situations where there are values of the underidentified parameters that make sense from an economic or statistical point of view, it is not at all clear a priori what values of $\boldsymbol{\beta}$ are likely to prevail under the alternative of *GH* innovations. For that reason, we will also consider instead a second approach, which consists in computing the LM test for all possible values of $\boldsymbol{\beta}$, and then take the supremum over those parameter values.

As we shall see below, either way we require the asymptotic variance of $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$ and $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$.

3.2 Asymptotic covariance matrix of the score under Gaussianity

Initially, we assume that $s_{k,t|T}(\boldsymbol{\theta})$ and $\mathbf{s}_{s,t|T}(\boldsymbol{\theta})$ are evaluated at the true parameter value $\boldsymbol{\theta}_0$. As is well known, the smoothed process $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}_0)$ will typically be serially correlated in spite of ε_t^* being *iid*. Consequently, the same will be true of $s_{k,t|T}(\boldsymbol{\theta}_0)$ and $\mathbf{s}_{s,t|T}(\boldsymbol{\theta}_0)$. In addition, the autocovariances of $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}_0)$ change with both t and T. However, since we are interested in the asymptotic variance of $\bar{s}_{k,t|T}(\boldsymbol{\theta})$ and $\bar{\mathbf{s}}_{s,t|T}(\boldsymbol{\theta})$ as $T \to \infty$, where the overbar denotes a sample mean, it suffices to compute the autocovariances of powers of the Wiener-Kolmogorov filter of ε_t^* based on a double-infinite sample of the observable vector $\mathbf{y}_t$, which we denote by $\boldsymbol{\varepsilon}^*_{t|\infty}(\boldsymbol{\theta})$. Likewise, we define $\mathbf{m}_{j,t|\infty}(\boldsymbol{\theta})$ for $j = 1, ..., 4$ as $\mathbf{m}_{j,t|T}(\boldsymbol{\theta})$ in (8) with $\boldsymbol{\varepsilon}^*_{t|\infty}(\boldsymbol{\theta})$ in place of $\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})$, and $\mathbf{b}_j(\boldsymbol{\theta})$ for $j = 1, ..., 4$ as $\mathbf{b}_{j,t|T}(\boldsymbol{\theta})$ in Proposition 1 evaluated after replacing $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})$ by $\boldsymbol{\Omega}_\infty(\boldsymbol{\theta})$.[6]

We can then prove the following result:

Proposition 2 *$\sqrt{T}\bar{s}_{k,t|T}(\boldsymbol{\theta}_0)$ and $\sqrt{T}\bar{\mathbf{s}}_{s,t|T}(\boldsymbol{\theta}_0)$ are asymptotically independent, with asymptotic variances given by*

$$\mathcal{C}_{k|\infty}(\boldsymbol{\theta}_0) = \mathbf{b}_4'(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_4(\boldsymbol{\theta}_0)\mathbf{b}_4'(\boldsymbol{\theta}_0) - \mathbf{b}_2'(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_2(\boldsymbol{\theta}_0)\mathbf{b}_2'(\boldsymbol{\theta}_0),$$

and

$$\mathcal{C}_{s|\infty}(\boldsymbol{\theta}_0) = \mathbf{b}_3'(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_3(\boldsymbol{\theta}_0)\mathbf{b}_3'(\boldsymbol{\theta}_0) - \mathbf{b}_1'(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_1(\boldsymbol{\theta}_0)\mathbf{b}_1'(\boldsymbol{\theta}_0),$$

respectively, with

$$\boldsymbol{\kappa}_i(\boldsymbol{\theta}) = \sum_{j=-\infty}^{\infty} cov[\mathbf{m}_{i,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{i,t-j|\infty}(\boldsymbol{\theta})],$$

denoting the autocovariance generating function of $\mathbf{m}_{i,t|\infty}(\boldsymbol{\theta})$ evaluated at one.

In Appendix B we develop a numerically reliable algorithm for computing these asymptotic variances for any state space model.

3.2.1 Parameter uncertainty

In practice, we do not generally know $\boldsymbol{\theta}_0$. Given that the model is normal under the null, it seems natural to study the asymptotic variance of $\bar{s}_{k,t|T}(\hat{\boldsymbol{\theta}}_T)$ and $\bar{\mathbf{s}}_{s,t|T}(\hat{\boldsymbol{\theta}}_T)$, where $\hat{\boldsymbol{\theta}}_T$ is the Gaussian Maximum Likelihood estimator of $\boldsymbol{\theta}$. Importantly, the following proposition shows that

[6] Under fairly general conditions $\boldsymbol{\Omega}_{t|\infty}(\boldsymbol{\theta})$ will not depend on t, so we can write $\boldsymbol{\Omega}_\infty(\boldsymbol{\theta}) = \boldsymbol{\Omega}_{t|\infty}(\boldsymbol{\theta})$ under the assumption that those conditions hold.

the sampling variability of the Gaussian PML estimators of $\boldsymbol{\theta}$ does not affect the asymptotic variance of the skewness and kurtosis components of the tests:

Proposition 3 *Let $\bar{\mathbf{s}}_{\mathrm{MV},T}(\boldsymbol{\theta})$ denote the Gaussian ML score with respect to the conditional mean and variance parameters $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$. Then,*

$$ACov\left[\bar{\mathbf{s}}_{\mathrm{MV},T}(\boldsymbol{\theta}_0), s_{\mathrm{k},t|\infty}(\boldsymbol{\theta}_0)\middle|\, \boldsymbol{\theta}_0\right] = \mathbf{0}$$

and

$$ACov\left[\bar{\mathbf{s}}_{\mathrm{MV},T}(\boldsymbol{\theta}_0), \mathbf{s}_{\mathrm{s},t|\infty}(\boldsymbol{\theta}_0)\middle|\, \boldsymbol{\theta}_0\right] = \mathbf{0}.$$

This result generalises Proposition 3 in Fiorentini and Sentana (2007), which contains an analogous result for dynamic location-scale multivariate models. It is also related to Bontemps and Meddahi (2005), who show that normality tests based on Hermite polynomials of univariate variables are insensitive to parameter uncertainty.

3.3 Test statistics

3.3.1 Multivariate Normality versus multivariate Student t innovations

The multivariate Student t distribution depends on a single shape parameter, which reflects the degrees of freedom of the underlying distribution. We work with its reciprocal, which we denote as η. We can easily compute an LM test for multivariate normality versus multivariate Student t distributed innovations on the basis of the value of the score of the log-likelihood function corresponding to η evaluated at the PML estimates $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\theta}}',0)'$.

Proposition 4 *The multivariate Student t-based LM test of normality can be expressed as:*

$$LM_T^{Student}(\boldsymbol{\theta}_0) = \frac{T}{\mathcal{C}_{\mathrm{k}|\infty}(\boldsymbol{\theta}_0)}\left[\frac{1}{T}\sum_{t=1}^{T} s_{\mathrm{k},t|T}(\boldsymbol{\theta}_0)\right]^2,$$

which is asymptotically distributed as a χ^2 distribution with one degree of freedom under the null.

It is important to mention that the fact that $\eta = 0$ lies at the boundary of the admissible parameter space invalidates the usual χ_1^2 distribution of the likelihood ratio (LR) and Wald (W) tests, which under the null will be more concentrated towards the origin (see Andrews (2001) and the references therein). The intuition can be perhaps more easily obtained in terms of the W test. Given that $\hat{\eta}_T$ cannot be negative, $\sqrt{T}\hat{\eta}_T$ will have a half-normal asymptotic distribution under the null (Andrews (1999)). As a result, the W test will be an equally weighted mixture of a chi-square distribution with 0 degrees of freedom (by convention, χ_0^2 is a degenerate random variable that equals zero with probability 1), and a chi-square distribution with 1 degree of freedom. In practice, we simply need to compare the appropriate t-statistic with the appropriate one-sided critical value from the normal tables. For analogous reasons, the asymptotic distribution of the LR test will also be degenerate half the time, and a χ_1^2 the other half.

Although the above argument does not invalidate the distribution of the LM test statistic, intuition suggests that the one-sided nature of the alternative hypothesis should be taken into account to obtain a more powerful test. For that reason, we also propose a simple one-sided version of the LM test for multivariate normality. In particular, since $E\left[\bar{s}_{\mathrm{k},T}(\boldsymbol{\theta})|\boldsymbol{\phi}_0\right] > 0$ when $\eta_0 > 0$, we suggest to use the LM test statistic when the sample average of the score is positive and 0 otherwise as our one-sided LM test, and to compare it to the same 50:50 mixture of chi-squares 0 and 1. In this context, we would reject H_0 at the $100\varkappa\%$ significance level if the average score with respect to η evaluated under the Gaussian null is positive *and* the LM statistic exceeds the $100(1-2\varkappa)$ percentile of a χ_1^2 distribution. Since the Kuhn-Tucker (KT) multiplier associated with the inequality restriction $\eta \geq 0$ is equal to $\max[-T^{-1}\sum_{t=1}^{T} s_{\mathrm{k},t|T}(\hat{\boldsymbol{\theta}}_T, 0), 0]$, our proposed one-sided LM test is equivalent to the KT multiplier test introduced by Gourieroux, Holly and Monfort (1980), which in turn is equivalent in large samples to the LR and W tests. As we argued before, the reason is that those tests are implicitly one-sided in our context. In this respect, it is important to mention that when there is a single restriction, such as in our case, those one-sided tests would be asymptotically locally more powerful (Andrews (2001)).

3.3.2 Multivariate Normality versus asymmetric GH innovations

If we combine Propositions 1 and 2, we can easily show that the LM test statistic for a given value of $\boldsymbol{\beta}$ will be given by

$$LM_T^{GH}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{T}{\mathcal{C}_{\mathrm{k}|\infty} + \boldsymbol{\beta}'\mathcal{C}_{\mathrm{s}|\infty}\boldsymbol{\beta}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \left[s_{\mathrm{k},t|T}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) \right] \right\}^2,$$

which will also follow an asymptotic χ_1^2 distribution under H_0.

But we can maximise $LM_T^{GH}(\boldsymbol{\theta}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ in closed form, and also obtain the asymptotic distribution of the resulting sup test statistic:

Proposition 5 *The supremum with respect to $\boldsymbol{\beta}$ of the LM tests based on (10) is equal to*

$$\sup_{\boldsymbol{\beta}} LM_T^{GH}(\boldsymbol{\theta},\boldsymbol{\beta}) = LM_T^{Student}(\boldsymbol{\theta}) + T\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})\right]' \mathcal{C}_{\mathrm{s}|\infty}^{-1}(\boldsymbol{\theta})\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})\right]$$

which is asymptotically distributed as a χ^2 distribution with $R+1$ degrees of freedom under the null.

Given that $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$ is asymptotically orthogonal to the other R moment conditions in $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$, we can conduct a partially one-sided test by combining the KT one-sided version of the symmetric GH test and the moment test based on $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$. By analogy with Mencía and Sentana (2012), this one-sided version should be equivalent in large samples to the corresponding LR test. The asymptotic distribution of the joint test under the null will be a 50:50 mixture of χ^2_R and χ^2_{R+1}.

3.4 Two illustrative examples

We illustrate our procedures with a static model and a dynamic one.

3.4.1 Static factor models (to be completed)

Let $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})$, with $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{M}(\boldsymbol{\theta})$ defined in section 2.3.1. We show in the appendix that

$$\boldsymbol{\varepsilon}^*_{t|\infty}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^*_{t|t}(\boldsymbol{\theta}) = \mathbf{G}'(\boldsymbol{\theta})[\mathbf{G}(\boldsymbol{\theta})\mathbf{G}'(\boldsymbol{\theta})]^{-1}\mathbf{G}(\boldsymbol{\theta})(\mathbf{y}_t - \boldsymbol{\pi}),$$

which confirms that like in any other static model, $\boldsymbol{\varepsilon}^*_{t|\infty}$ will be white noise, with covariance matrix

$$\boldsymbol{\Gamma}(\mathbf{0}) = \mathbf{G}'(\boldsymbol{\theta})[\mathbf{G}(\boldsymbol{\theta})\mathbf{G}'(\boldsymbol{\theta})]^{-1}\mathbf{G}(\boldsymbol{\theta}).$$

In addition,

$$\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta}) = \boldsymbol{\Omega}_{t|\infty}(\boldsymbol{\theta}) = \boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta}) = \mathbf{I}_K - \mathbf{G}'(\boldsymbol{\theta})[\mathbf{G}(\boldsymbol{\theta})\mathbf{G}'(\boldsymbol{\theta})]^{-1}\mathbf{G}(\boldsymbol{\theta}),$$

which has rank N rather than K. Hence, we will have that under the null,

$$\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}_0)|Y_T \sim N\left[\boldsymbol{\varepsilon}^*_{t|t}(\boldsymbol{\theta}_0), \boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})\right],$$

which contains all the information we need to compute the normality tests.

To provide some intuition, though, it is convenient to focus on tests that focus exclusively on the common factor. If we could observe f_t, then we could write the joint log-likelihood function of $\mathbf{y}_t$ and f_t as the sum of the marginal log-likelihood function of f_t and the log-likelihood function of $\mathbf{y}_t$ conditional on f_t, which would coincide with the marginal log-likelihood function of the idiosyncratic terms $\mathbf{v}_t$. If we maintained the assumption that this conditional distribution was Gaussian, and confined the non-normality to the marginal distribution of f_t, the results in Mencía and Sentana (2012) would imply that the LM test of the null hypothesis that f_t is Gaussian versus the alternative that it follows an asymmetric Student t would be based on the following influence conditions:

$$\begin{array}{c} H_3(f_t) = f_t^3 - 3f_t, \\ H_4(f_t) = f_t^4 - 6f_t^2 + 3, \end{array} \tag{11}$$

which coincide with the third and fourth Hermite polynomials for f_t.

Unfortunately, f_t is unknown. But we can easily compute the expected values of these expressions conditional on $\mathbf{y}_t$, which under normality are simple functions of

$$f_{t|t}(\boldsymbol{\theta}) = E(f_t|\mathbf{y}_t) = \omega_f(\boldsymbol{\theta})\mathbf{c}'\boldsymbol{\Gamma}^{-1}(\mathbf{y}_t - \boldsymbol{\pi})$$

and

$$\omega_f(\boldsymbol{\theta}) = V(f_t|\mathbf{y}_t) = \frac{1}{\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c} + 1}.$$

In particular, we can show that the expected value of the elements of (11) is proportional to $H_3[f_{t|t}(\boldsymbol{\theta})/\sqrt{1-\omega_f(\boldsymbol{\theta})}]$ and $H_4[f_{t|t}(\boldsymbol{\theta})/\sqrt{1-\omega_f(\boldsymbol{\theta})}]$, respectively, where $V[f_{t|t}(\boldsymbol{\theta})] = 1 - \omega_f(\boldsymbol{\theta})$ by virtue of the fact that

$$V(f_t) = E[V(f_t|\mathbf{y}_t)] + V[E(f_t|\mathbf{y}_t)].$$

Somewhat remarkably, therefore, the LM test for the normality of the latent common factor will numerically coincide with the usual LM test for the normality of its best estimator in the mean square error sense. Obviously, analogous calculations apply to each element of $\mathbf{v}_t^*$.[7]

3.4.2 UCARIMA models (to be completed)

Consider now the AR(1) plus noise model of section 2.3.2. Since there are only two shocks, we should at least look at (i) a joint test of normality, (ii) a test of normality of the "signal" with the maintained hypothesis of normality for the "non-signal", and (iii) vice versa.

In the case of the non-signal component, Proposition 1 implies that for symmetric Student t alternatives, the score with respect to the reciprocal of the degrees of freedom parameter

[7] If we fix $c = 1$ for simplicity in the univariate case, $y_t = f_t + \sqrt{\gamma}v_t^*$ would be the sum of two white noise processes. Straigthforward calculations show that $LM_T^{Student}(\boldsymbol{\theta})$ is numerically identical irrespective of whether we are testing normality against H_J, H_S or H_I. The same applies to $\sup LM_T^{GH}(\boldsymbol{\theta})$.

evaluated under the null will be given by

$$E\left[s_{\eta t}^{S_v}(\boldsymbol{\theta},\mathbf{0})|Y_T\right]=\frac{1}{2}\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta})\right]^2-\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta})\right]v_{t|T}^{*2}(\boldsymbol{\theta})+\frac{1}{2}\sqrt{\frac{1}{6}}v_{t|T}^{*4}(\boldsymbol{\theta}). \quad (12)$$

But the optimality of the Wiener-Kolmogorov-Kalman filter under Gaussianity implies that

$$V(v_t^*)=V[v_{t|T}^*(\boldsymbol{\theta})]+V[v_t^*-v_{t|T}^*(\boldsymbol{\theta})],$$

which in turns means that

$$V[v_{t|T}^*(\boldsymbol{\theta})]=1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta}_0).$$

Hence, expression (12) is proportional to the fourth order Hermite polynomial of the standardised variable $v_{t|T}^*(\boldsymbol{\theta})/\sqrt{1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta}_0)}$. Therefore, for this model our proposed LM test also gives exactly the same influence function as an LM test of normal versus Student t that would treat $v_{t|T}^*(\boldsymbol{\theta})$ as an *iid* series. Unlike in the static model considered in section 3.4.1, though, the elements of (12) are serially correlated.

To compute the asymptotic variance of the sample average of this score, we can exploit the fact that in this case $v_{t|\infty}^*$ has the autocorrelation structure of an ARMA(1,1) process with moving average coefficient $-\rho$, autoregressive coefficient

$$\theta=\frac{1+(1+\rho^2)\gamma}{2\rho\gamma}-\frac{\rho}{|\rho|}\sqrt{\left[\frac{1+(1+\rho^2)\gamma}{2\rho\gamma}\right]^2-1}$$

and residual variance

$$\sigma_w^2=\frac{1+(1+\rho^2)\gamma}{1+\theta^2}$$

(see appendix C for details). Then, tedious but straightforward calculations based on Proposition 2 deliver

$$\mathcal{C}_{\mathrm{k}|\infty}^{S_v}(\boldsymbol{\theta}_0)=\left[1+\frac{(\theta_0+\rho_0)^2}{1-\theta_0^2}\right]^4\frac{\gamma_0^4}{\sigma_0^8}\left[1+2\frac{(1+\rho_0\theta_0)^4(\theta_0+\rho_0)^4}{(1+\rho_0^2+2\rho_0\theta_0)^4}\frac{1}{1-\theta_0^4}\right].$$

Not surprisingly, if we follow the same steps for the signal component we end up with

$$E\left[s_{\eta t}^{S_f}(\boldsymbol{\theta},\mathbf{0})|Y_T\right]=\frac{1}{2}\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{f},t|T}(\boldsymbol{\theta})\right]^2-\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{f},t|T}(\boldsymbol{\theta})\right]f_{t|T}^{*2}(\boldsymbol{\theta})+\frac{1}{2}\sqrt{\frac{1}{6}}f_{t|T}^{*4}(\boldsymbol{\theta}), \quad (13)$$

whose sample average has asymptotic variance

$$\mathcal{C}_{\mathrm{k}|\infty}^{S_f}(\boldsymbol{\theta}_0)=\frac{1}{\sigma_{w0}^8(1-\theta_0^2)^4}\left(\frac{1+\theta_0^4}{1-\theta_0^4}\right)$$

because $f_{t|T}^*$ has the autocorrelation structure of an AR(1) process whose autoregressive coefficient is θ.

3.5 Comparison with alternative approaches

3.5.1 Univariate tests applied to the smoothed innovations (to be completed)

As we mentioned in the introduction, Harvey and Koopman (1992) applied standard univariate normality tests for observed variables to the smoothed values of the innovations in the underlying components of a local level model explicitly taking into account the serial correlation in those filtered estimates implied by the model.

To illustrate their procedure, we consider again the Ar(1) plus noise of section 2.3.2, which includes as a particular case random walk plus noise model for $\alpha = 1$. Their asymmetry test is based on the skewness coefficient

$$s_{\varepsilon_i^*} = m_{\varepsilon_i^*,3}/m_{\varepsilon_i^*,2}^{3/2},$$

where

$$m_{\varepsilon_i^*,j} = T^{-1}\sum_{t=1}^{T}(\varepsilon_{i,t|T}^* - \bar{\varepsilon}^*)^j$$

is the j^{th} centred sample moment of the smoothed innovations. Under normality, the asymptotic variance of $s_{\varepsilon_i^*}$ will be given by $6\zeta_{\varepsilon_i^*}(\boldsymbol{\theta}_0, 4)$, where

$$\zeta_{\varepsilon_i^*}(\boldsymbol{\theta}_0, \lambda) = \sum_{j=-\infty}^{\infty}[\rho_{\varepsilon_i^*}(j)]^\lambda$$

provides the sum of powers of the autocorrelations, which are the autocorrelations of the powers of the original Gaussian series.

Similarly, their excess kurtosis test is based on the sample excess kurtosis coefficient

$$k_{\varepsilon_i^*} = m_{\varepsilon_i^*,4}/m_{\varepsilon_i^*,2}^2 - 3,$$

whose asymptotic variance under normality will be given by $24\zeta_{\varepsilon_i^*}(\boldsymbol{\theta}_0, 4)$.

Using the results in section 3.4.2 on the autocorrelation structure of $f_{t|T}^*$ and $v_{t|T}^*$, it is straightforward to show that

$$\begin{aligned}
\zeta_f(\boldsymbol{\theta}_0, 4) &= \frac{1+\theta_0^4}{1-\theta_0^4}, \\
\zeta_v(\boldsymbol{\theta}_0, 4) &= 1 + 2\frac{(1+\rho_0\theta_0)^4(\theta_0+\rho_0)^4}{(1+\rho_0^2+2\rho_0\theta_0)^4}\frac{1}{1-\theta_0^4}.
\end{aligned}$$

It is interesting to compare these tests to our LM tests based on (13) and (12). In this regard, the procedures proposed by Harvey and Koopman (1992) can be regarded as moment tests of

$$E[f^{*3}_{t|T}(\boldsymbol{\theta})] = 0 \quad E[f^{*4}_{t|T}(\boldsymbol{\theta}) - 3] = 0,$$
$$E[v^{*3}_{t|T}(\boldsymbol{\theta})] = 0 \quad E[v^{*4}_{t|T}(\boldsymbol{\theta}) - 3] = 0.$$

Therefore, the main difference is that they look at third and fourth moments, while we use the log-likelihood score, which coincides with the third and fourth Hermite polynomials. As shown by Bontemps and Meddahi (2005), the advantage of the latter is that they are not affected by the sampling variability introduced by replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}_T$. Nevertheless, Harvey and Koopman (1992) indicate that their tests are also insensitive to parameter uncertainty. In that regard, the situation seems analogous to the Jarque and Bera (1980) tests, whose distribution is insensitive to parameter uncertainty for many models (see Fiorentini, Sentana and Calzolari (2004)).

3.5.2 Reduced form tests (to be completed)

As we mentioned in section 2.2, if the innovations to the structural model $\boldsymbol{\varepsilon}^*_t$ are *iid* Gaussian, then the reduced form innovations $\mathbf{w}_t$ will also be so. As a result, checking the normality of the latter provides an indirect way of checking the normality of the former. Nevertheless, if some elements of $\boldsymbol{\varepsilon}^*_t$ are not normal, then the conditional distribution of the reduced form innovations will typically be extremely complicated, especially taking into account that they are unlikely to follow a martingale difference sequence in dynamic contexts. The problem is that the conditional mean of the observed variables given their past alone will no longer be given by the one-period ahead linear prediction generated by the Kalman filter recursions, $\mathbf{y}_{t|t-1}(\boldsymbol{\theta})$. Similarly, the conditional variance will not usually coincide with the associated mean-square error matrix $\boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\theta})$.

Still, it may be worth considering tests against the following alternative model

$$\mathbf{y}_t|\mathbf{y}_{t-1}, ..., \mathbf{y}_1; \boldsymbol{\theta} \sim GH[\mathbf{y}_{t|t-1}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\theta}), \eta, \psi, \boldsymbol{\beta}],$$

which maintains the assumption that the conditional mean and variance coincide with their values under normality, but allows for a non-Gaussian distribution. The assumption that the distribution of $\mathbf{y}_t$ conditional on Y_{t-1} is GH but with a mean vector and covariance matrix given by the usual Gaussian Kalman filter recursions may be regarded as a way of constructing a convenient auxiliary model that coincides with the model of interest for $\boldsymbol{\eta} = \mathbf{0}$, but whose log-likelihood function and score we can obtain in closed form for every possible value of $\boldsymbol{\theta}$ when $\boldsymbol{\eta} \neq \mathbf{0}$.

The pay-off is that the resulting model falls within the framework studied by Mencía and Sentana (2012). Specifically, if we define the standardised reduced form innovations

$$\mathbf{w}^*_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-\frac{1}{2}}_{t|t-1}(\boldsymbol{\theta})[\mathbf{y}_t - \mathbf{y}_{t|t-1}(\boldsymbol{\theta})],$$

and their (square) Euclidean norm as

$$\varsigma_{t|t-1}(\boldsymbol{\theta}) = \mathbf{w}^*_{t|t-1}(\boldsymbol{\theta})'\mathbf{w}^*_{t|t-1}(\boldsymbol{\theta}) = [\mathbf{y}_t - \mathbf{y}_{t|t-1}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{-1}_{t|t-1}(\boldsymbol{\theta})[\mathbf{y}_t - \mathbf{y}_{t|t-1}(\boldsymbol{\theta})],$$

we can write the influence functions underling their test as

$$\begin{aligned} s^{MS}_{\mathrm{k},t|t-1}(\boldsymbol{\theta}) &= \frac{1}{4}\varsigma^2_{t|t-1}(\boldsymbol{\theta}) - \frac{N+2}{2}\varsigma_{t|t-1}(\boldsymbol{\theta}) + \frac{N(N+2)}{4}, \\ \mathbf{s}^{MS}_{\mathrm{s},t|t-1}(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|t-1}(\theta)\mathbf{w}^*_{t|t-1}(\boldsymbol{\theta})\left[\varsigma_{t|t-1}(\boldsymbol{\theta}) - (N+2)\right]. \end{aligned}$$

Propositions 3 and 5 in Mencía and Sentana (2012) provide the asymptotic variance for the sample average of those influence functions, which depend on $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = V(\mathbf{w}_t)$, which typically coincides with the steady state value of $\boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\theta})$.

Interestingly, we can show that in static factor models, tests of the null hypothesis that $\mathbf{w}^*_{t|t-1}(\boldsymbol{\theta})$ is Gaussian against the alternative hypothesis that it follows a *GH* distribution are numerically identical to the analogous tests for the latent variables $\boldsymbol{\varepsilon}^*_t$. The intuition is as follows. A well known property of the *GH* distribution is that the distribution of linear combinations (including the individual components) also follow *GH* distributions (see Blæsild (1981)). Therefore, in this case the relationship between the non-normality of $\boldsymbol{\varepsilon}^*_t$ and $\mathbf{w}^*_t$ is exact.

4 Monte Carlo simulations

In this section, we assess the finite sample size and power properties of the testing procedures discussed above by means of several extensive Monte Carlo exercises.

4.1 Design and estimation details

In addition to Gaussian innovations, to assess the power properties of the tests we also consider two different distributional assumptions for the innovations under the alternative, Student t with 8 degrees of freedom and asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\iota}_{K\times 1}$, for the several subsets of (the innovations to) the state variables involved by the different alternative hypotheses. Since we consider the four alternative hypotheses of section 2.4, we end up with nine different specifications in total.

For each distributional assumption on $\boldsymbol{\varepsilon}^*$, we generate 10,000 samples of size T equal to 100 and 250 observations each for:

1. The trivariate static factor model of section (6) (i.e. $N = 3$ and $K = 4$) with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} =(1,1,1)'$ and $\boldsymbol{\gamma} = q^{-1}(1,1,1)'$, where q trivially reflects the signal-to-noise ratio, which we set to 2, 1, and .5.

2. An Ar(1) signal plus noise stationary Ucarima model as the one described in section (2.3.2) in which $\alpha_x = .75$ and the signal-to-noise ratio, which is now given by $q = \sigma_x^2/\sigma_\epsilon^2 = \sigma_f^2/[(1-\alpha_x^2)\sigma_v^2]$, takes the values 2 and 1.

3. A local-level model in which the signal-to-noise ratio, which is given by $q = \sigma_f^2/\sigma_v^2$, takes the values 2 and .5, as in Harvey and Koopman (1992).

4. The bivariate cointegrated, dynamic single factor model of section 2.3.3, with $\alpha_x = .5$, $\alpha_{\epsilon_1} = .2$, $\alpha_{\epsilon_2} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_1 = 2$ and $q_2 = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2 = [\sigma_f^2(1-\alpha_{\epsilon_i}^2)]/[(1-\alpha_x^2)\sigma_v^2]$ represents the signal-to-noise ratio for y_{it}.

We use standard Matlab routines for estimation. In the case of Ucarima models 2 and 3, we use the reduced form representation of section 2 to improve the computational efficiency of the algorithm. Finally, for computation of the variances of the test statistics we truncate the infinite sums of Proposition 2 whenever the difference between two consecutive iterations is less than 10^{-5}.

4.2 Size properties (to be completed)

The first question that we need to address is whether the asymptotic distribution under the null attributed to the test statistics introduced in Section 3 is reliable in finite samples. In Table 1 we report the rejection rates under the null of the test statistics at the 1%, 5% and 10% nominal level. Interestingly, most of the tests are oversized (undersized) at the 1% (10%) level irrespective of the DGP. In contrast, they only present a moderate (<1%) distortion at the 5% level.

4.3 Power properties (to be completed)

As expected, for a given alternative, the most powerful tests are typically the score tests which we have designed against that particular alternative. Nevertheless, the reduced form tests seem to have non-trivial power.

5 Empirical application

In theory the expenditure and income measures of aggregate (real) production should be equal, but in practice they differ because they are calculated from different sources (see Landefeld, Seskin and Fraumeni (2008) for a review). Traditionally, the difference between the two, officially known as the "statistical discrepancy" (see Grimm (2007)), was regarded by many academic economists as a curiosity in the US National Input and Product Accounts (NIPA) elaborated by the Bureau of Economic Analysis (BEA) of the US Department of Commerce. However, the Great Recession that began in the second half of 2007 and ended in June 2009 substantially renewed interest in the possibility of obtaining more reliable GDP growth figures by combining the two measures (see e.g. Nalewaik (2010, 2011), Greenaway-McGrevy (2011) and Aruoba, Diebold, Naleweik and Schorfheide (2015)). Some national statistical offices compute a simple equally weighted average of the different aggregate series, and in fact, BEA began providing such an average in July 2015. More sophisticated combination methods would give higher weights to the more precise GDP measures, as argued by Stone, Champernowne and Meade (1942) (see Weale (1992) for an account of the earlier literature). As emphasised by Smith, Weale and Satchell (1998), though, dynamic considerations matter. There are at least two important reasons:

1. The expenditure and income measures should be cointegrated with the true GDP, and therefore between themselves, with cointegrating vector (1,-1).

2. The associated measurement errors should be stationary but they may well be serially correlated.

For those reasons, the single factor model with unit loadings on an I(1) common factor and covariance stationary specific factors described in section 2.3.3 provides a rather natural way of capturing the dynamics of the log GDP and GDI series. Importantly, such a model allows for systematic biases in the measurement errors through δ, the difference between those biases determining the mean of the statistical discrepancy while their levels fixing the initial conditions.[8]

[8] The specification of the serial correlation structure for the underlying GDP and measurement error follows from the empirical analysis in earlier versions of Fiorentini and Sentana (2013), who found evidence in favour

The normality of the underlying GDP series, as opposed to its two noisy measurements, is particularly relevant in the design of macroeconomic stability policies because the quadratic loss function embeded in a Taylor rule implicitly assumes that short run deviations of income from its trend growth path are Gaussian.

We initially look at US data from 1984Q2 to 2007Q2. We chose the final date to exclude the Great Recession from the sample. As for the start date, we chose it because it marks the beginning of the so-called Great Moderation, as in Nalewaik (2010).

Panel A of Table 6 presents the Gaussian maximum likelihood estimates of the model parameters and their corresponding standard errors based on the average Hessian. Those estimates suggest that GDP provides a better measure of output than GDI, in the sense that GDP measurement errors have both a smaller autoregressive coefficent –in absolute value– and a smaller conditional variance parameter.[9] As for the common factor, the estimates suggest that the growth rate of the "true" aggregate real output series is quite persistent.

Moving on to the normality tests, which are reported in Panel B of Table 6, the null of Gaussian innovations is rejected for the measurement errors. Specifically, we reject at conventional levels not only when we use the joint test but also when we separately look at the skewness component (p-value of .012) or the kurtosis one (p-value of .016). In contrast, the normality of the innovations to the underlying GDP growth rates is not rejected, which suggests that the soothing effects of the so–called Great Moderation propagated beyond second moments. Nevertheless, it should be remembered that a sample size of 90 observations is relatively small.

In Table 7 we present analogous results for a slightly larger sample that includes the Great Recession (1984Q2-2015Q3). As can be seen from Panel A, there are no dramatic changes in the parameter estimates, except perhaps for a higher persistence of the common factor, whose innovations have unsurprinsingly a larger variance too. In contrast, the assumption of normality is massively rejected at the 5% level for all the alternative hypothesis we consider.

To provide some further insight, in Figure 1 we plot the temporal evolution of the smoothed innovations (top panels), as well as the influence functions underlying the kurtosis tests (middle panels) and skewness tests (bottom panels) for both common factor (left panels) and measurement errors (right panels) for the sample 1984Q2 to 2007Q2.

---

of AR(1) dynamics for both (i) the first difference of the common factor, and (ii) the levels of the measurement errors.

[9] The negative serial correlation coefficient suggests a tendency to compensate prior measurement errors.

As can be seen from the left panels of this figure, the recession of 1991 constitutes the most atypical feature driving the underlying factor innovations, although its contribution to $\bar{s}_{\mathrm{k},T}(\boldsymbol{\theta})$ and $\bar{\mathbf{s}}_{\mathrm{s},T}(\boldsymbol{\theta})$ is not sufficiently large for the skewness and kurtosis test statistics to reject the null of normality.

In contrast, Figures 1d and 1f suggest that there was an unusual measurement issue around 2000, which leads to a massive rejection of the Gaussian null for the measurement errors.

In this regard, we would like to emphasise that plots of the influence functions $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$ and $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$'s seem to be more informative than plots of the smoothed innovations for the purposes of detecting non-normality. For example, Figure 2, which includes the Great Recession, shows that the dramatic drop experienced in 2008Q4 leads to an impact on those scores which is at least five times larger, resulting in a strong rejection of the null.

Finally, we also report the results for a much longer sample covering 1952Q1 to 2015Q3. We chose this alternative start date so that our sample begins soon after the Treasury - Federal Reserve Accord whereby the Fed stopped its wartime pegging of interest rates. Once again, the Gaussian null is systematically rejected, which suggests that the unusual shocks observed during the Great Recession are not that unsual when one takes a longer historical perspective. The same conclusion can be reached by looking at the time series plots in Figure 3: the Great Recession is not that great when compared with the turmoil in the early 80's.

6 Conclusions (to be completed)

We exploit the rationale behind the Expectation Maximisation algorithm to derive simple to implement and interpret score tests of normality in all or a subset of the innovations to the latent variables in state space models against Generalised Hyperbolic alternatives, including symmetric and asymmetric Student t. We decompose our tests into third and fourth moment components, and obtain one-sided Likelihood Ratio analogues, whose asymptotic distribution we provide. We perform a Monte Carlo study of the finite sample size and power of our procedures and previous proposals. Finally, we illustrate our tests in an application to US aggregate real output measurement.

References

Amengual, D. and E. Sentana (2015): "Is a normal copula the right copula?" CEMFI Working Paper No. 1504.

Anderson, B.D.O. and J.B. Moore (1979): *Optimal filtering*, Prentice-Hall, New Jersey.

Andrews, D.W.K. (1999): "Estimation when a parameter in on the boundary," *Econometrica*, 67, pp.1341–1383.

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

Aruoba, S. B., F.X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2015): "Improving GDP measurement: A measurement-error perspective," NBER Working Papers 18954.

Bai, J. and S. Ng (2005): "Tests for skewness, kurtosis, and normality for time Sseries data," *Journal of Business Economics and Statistics,* 23, pp. 49–60.

Barndorff-Nielsen, O. (1977): "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society,* 353, pp. 401–419.

Blæsild, P. (1981): "The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data," *Biometrika* 68, pp. 251–263.

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

Bontemps, C. and N. Meddahi (2005): "Testing Normality: A GMM approach," *Journal of Econometrics,* 124, pp. 149–186.

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

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

Durbin, J. and S.J. Koopman (2012): *Time series analtsis by state space methods*, 2nd ed., Oxford University Press, Oxford.

Fiorentini, G. and C. Planas (1998): "From autocovariances to moving average: An algorithm comparison," *Computational Statistics,* 13, pp. 477–484.

Fiorentini G., E. Sentana, and G. Calzolari (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations," *Journal of Business and Economic Statistics,* 21(4), pp. 532–546.

Fiorentini G., and E. Sentana (2012): "Tests for serial dependence in static non-Gaussian factor models," CEMFI Working Paper No.1211.

Fixler D., R. Greenaway-McGrevy, and B. Grimm (2011): "Revisions to GDP, GDI, and their major components," *Survey of Current Business*, 91(7), pp. 1–23.

Gourieroux C., A. Holly, and A. Monfort (1980): "Kühn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters," Discussion Paper 770, Harvard Institute of Economic Research.

Granger, C.W.J., and M.J. Morris (1976): "Time series modelling and interpretation," *Journal of the Royal Statistical Society* Series A, 139, pp. 246–257.

Grimm, B.T. (2007): "The statistical discrepancy," Bureau of Economic Analysis, Washington D.C.

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

Harvey A., and S.J. Koopman (1992): "Diagnostic checking of unobserved-components time series models," *Journal of Business and Economic Statistics,* 10(4), pp. 377–389.

Johnson, N., S. Kotz, and N. Balakrishnan (1994): *Continuous univariate distributions*, Wiley, New York.

Jørgensen, B. (1982): *Statistical properties of the Generalized Inverse Gaussian distribution*, Springer-Verlag, New York.

Komunjer I. and S. Ng (2011): "Dynamic identification of dynamic stochastic general equilibrium models," *Econometrica,* 79, pp. 1995–2032.

Landefeld, J.S., E.P. Seskin, and B.M. Fraumeni (2008): "Taking the pulse of the economy: measuring GDP," *Journal of Economic Perspectives,* 22, pp. 193–216.

Lehmann, B.N., and D.M. Modest (1989): "The empirical foundations of the arbitrage pricing theory," *Journal of Financial Economics,* 21, pp. 213–254.

Lomnicki Z. (1961): "Tests for departure from normality in the base of linear stochastic processes," *Metrika,* 4, pp. 37–62.

Louis, T. A. (1982): "Finding observed information using the EM algorithm," *Journal of the Royal Statistical Society* Series B, 44, pp. 98–103.

Mencía J., and E. Sentana (2012): "Distributional tests in multivariate dynamic models with normal and Student-t innovations," *The Review of Economics and Statistics,* 94(1), pp. 133–152.

Nalewaik, Jeremy J. (2010): "The income- and expenditure-side measures of output growth," *Brookings Papers on Economic Activity,* 1, pp. 71–106.

Nalewaik, Jeremy J. (2011): "The income- and expenditure-side measures of output growth–An update through 2011Q2," *Brookings Papers on Economic Activity,* 2, pp. 385–402.

Rubin, D. and D. Thayer (1982): "EM algorithms for ML factor analysis," *Psychometrika,* 47, pp. 69–76.

Ruud, P. (1991): "Extensions of estimation methods using the EM algorithm," *Journal of Econometrics,* 49, pp. 305–341.

Smith R., M. Weale, and S. Satchell (1998): "Measurement error with accounting constraints point and interval estimation for latent data with an application to UK gross domestic product," *Review of Economic Studies,* 65(1), pp. 109–134.

Stone, R., D.G. Champernowne, and J.E. Meade (1942): "The precision of national income estimates," *Review of Economic Studies,* 9, pp. 111–125.

Tanner, M. (1996): *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*, (3rd ed.), Springer-Verlag.

Watson, M.W., and R.F. Engle (1983): "Alternative algorithms for estimation of dynamic MIMIC, factor, and time varying coefficient regression models," *Journal of Econometrics,* 23, pp. 385–400.

Weale, M. (1992): "Estimation of Data Measured with Error and Subject to Linear Restrictions," *Journal of Applied Econometrics,* 7, pp. 167–174.

Appendix

A Proofs

A.1 Technical Prelude

A number of formal results is to be invoked in order to establish the main propositions of the paper. Also, a set of techniques is required to implement the testing procedure in practice. For the reader's own convenience we have collected them all under a single section.

A.1.1 Some properties of Gaussian random vectors

Lemma 1 *Let $\boldsymbol{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a n_z-dimensional real Gaussian random vector. Then,*

1. *Expectation of second powers:*

$$E(\boldsymbol{z}\boldsymbol{z}') = \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}$$

2. *Expectation of third powers:*

$$E[\boldsymbol{z}(\boldsymbol{z} \odot \boldsymbol{z})'] = \boldsymbol{\mu}(\boldsymbol{\mu} \odot \boldsymbol{\mu})' + 2(\boldsymbol{\Sigma} \odot \boldsymbol{\ell}_{n_z}\boldsymbol{\mu}') + \boldsymbol{\mu}\mathrm{vecd}'(\boldsymbol{\Sigma})$$

3. *Expectation of fourth powers:*

$$\begin{aligned} E((\boldsymbol{z} \odot \boldsymbol{z})(\boldsymbol{z} \odot \boldsymbol{z})') = {} & (\boldsymbol{\mu} \odot \boldsymbol{\mu})(\boldsymbol{\mu} \odot \boldsymbol{\mu})' + 2(\boldsymbol{\Sigma} \odot \boldsymbol{\Sigma}) + \mathrm{vecd}(\boldsymbol{\Sigma})\mathrm{vecd}'(\boldsymbol{\Sigma}) \\ & + 4(\boldsymbol{\Sigma} \odot \boldsymbol{\mu}\boldsymbol{\mu}') + \mathrm{vecd}(\boldsymbol{\mu}\boldsymbol{\mu}')\mathrm{vecd}'(\boldsymbol{\Sigma}) + \mathrm{vecd}(\boldsymbol{\Sigma})\mathrm{vecd}'(\boldsymbol{\mu}\boldsymbol{\mu}') \end{aligned}$$

where $\odot$ denotes Hadamard (or Schur) element-wise product, $\mathrm{vecd}(\cdot)$ *is the diagonal linear transformation and $\boldsymbol{\ell}_{n_z}$ is the vector of ones in $\mathbb{R}^{n_z}$.*

Proof. The proof is straightforward –rather tedious– algebra. ■

Lemma 2 *Define $\mathbf{m}_h : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_1 n_2}$ for $n_1, n_2 \in \mathbb{Z}_{++}$ and $h \in \{2, 3, 4\}$ by*

$$\mathbf{m}_2(\mathbf{w}_1, \mathbf{w}_2) = vec(\mathbf{w}_1\mathbf{w}_2')$$

$$\mathbf{m}_3(\mathbf{w}_1, \mathbf{w}_2) = vec[\mathbf{w}_1(\mathbf{w}_2 \odot \mathbf{w}_2)']$$

$$\mathbf{m}_4(\mathbf{w}_1, \mathbf{w}_2) = vec[(\mathbf{w}_1 \odot \mathbf{w}_1)(\mathbf{w}_2 \odot \mathbf{w}_2)']$$

where $\mathbf{w}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{w}_2 \in \mathbb{R}^{n_2}$ and $\mathrm{vec}(\cdot)$ *is the vectorization (by columns) operator. Consider the real Gaussian random vector*

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}'_{xy} & \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}'_{xz} & \boldsymbol{\Sigma}'_{yz} & \boldsymbol{\Sigma}_{zz} \end{bmatrix} \right)$$

where $\boldsymbol{x}$ is n_x-dimensional, $\boldsymbol{y}$ is n_y-dimensional and $\boldsymbol{z}$ is n_z-dimensional. Let $\otimes$ denote Kronecker product. Then,

i) Covariance with the first power:

$$\begin{aligned}
\operatorname{cov}(\boldsymbol{x}, \mathbf{m}_2(\boldsymbol{y},\boldsymbol{z})) &= 0_{(n_x, n_y n_z)} \\
\operatorname{cov}(\boldsymbol{x}, \mathbf{m}_3(\boldsymbol{y},\boldsymbol{z})) &= 2\left(\boldsymbol{\ell}_{n_x} \otimes \operatorname{vec}'(\boldsymbol{\Sigma}_{yz})\right) \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}'_{n_y}\right) \\
&\quad :::: + \left(\operatorname{vecd}'(\boldsymbol{\Sigma}_{zz}) \otimes \mathbf{1}_{n_x \times n_y}\right) \odot \left(\boldsymbol{\ell}'_{n_z} \otimes \boldsymbol{\Sigma}_{xy}\right) \\
\operatorname{cov}(\boldsymbol{x}, \mathbf{m}_4(\boldsymbol{y},\boldsymbol{z})) &= 0_{(n_x, n_y n_z)}
\end{aligned}$$

ii) Covariance with the second power:

$$\begin{aligned}
\operatorname{cov}(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_2(\boldsymbol{y},\boldsymbol{z})) &= \left(\mathbf{1}_{n_x \times n_z} \otimes \boldsymbol{\Sigma}_{xy}\right) \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\
&\quad + \left(\boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}'_{n_y}\right) \odot \left(\boldsymbol{\ell}'_{n_x} \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_z}\right) \\
\operatorname{cov}(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_3(\boldsymbol{y},\boldsymbol{z})) &= 0_{(n_x^2, n_y n_z)} \\
\operatorname{cov}(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_4(\boldsymbol{y},\boldsymbol{z})) &= 4\left(\boldsymbol{\ell}_{n_x^2} \otimes \operatorname{vec}'(\boldsymbol{\Sigma}_{yz})\right) \odot \operatorname{cov}(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_2(\boldsymbol{y},\boldsymbol{z})) \\
&\quad + 2\left(\boldsymbol{\ell}_{n_x^2} \otimes \boldsymbol{\ell}'_{n_z} \otimes \operatorname{vecd}'(\boldsymbol{\Sigma}_{yy})\right) \\
&\quad \odot \left(\boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}'_{n_y}\right) \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\
&\quad + 2\left(\boldsymbol{\ell}_{n_x^2} \otimes \operatorname{vecd}'(\boldsymbol{\Sigma}_{zz}) \otimes \boldsymbol{\ell}'_{n_y}\right) \\
&\quad \odot \left(\mathbf{1}_{n_x \times n_z} \otimes \boldsymbol{\Sigma}_{xy}\right) \odot \left(\boldsymbol{\ell}'_{n_y} \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_x}\right)
\end{aligned}$$

iii) Covariance with the third power:

$$\begin{aligned}
\operatorname{cov}(\mathbf{m}_3(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_3(\boldsymbol{y},\boldsymbol{z})) &= \left(\operatorname{vecd}(\boldsymbol{\Sigma}_{xx}) \otimes \boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\ell}'_{n_y n_z}\right) \odot \left(\boldsymbol{\ell}_{n_x} \otimes \operatorname{cov}(\boldsymbol{x}, \mathbf{m}_3(\boldsymbol{y},\boldsymbol{z}))\right) \\
&\quad + 2\left(\mathbf{1}_{n_x \times n_z} \otimes \boldsymbol{\Sigma}_{xy}\right) \odot \left((\boldsymbol{\Sigma}_{xz} \odot \boldsymbol{\Sigma}_{xz}) \otimes \mathbf{1}_{n_x \times n_y}\right) \\
&\quad + 2\left(\operatorname{vec}(\boldsymbol{\Sigma}_{xx}) \otimes \mathbf{1}_{1 \times n_y n_z}\right) \odot \left(\boldsymbol{\ell}_{n_x^2} \otimes \operatorname{vecd}'(\boldsymbol{\Sigma}_{zz}) \otimes \boldsymbol{\ell}'_{n_y}\right) \\
&\quad \odot \left(\boldsymbol{\ell}'_{n_y} \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_x}\right) \\
&\quad + 4\left(\operatorname{vec}(\boldsymbol{\Sigma}_{xx}) \otimes \boldsymbol{\ell}'_{n_y n_z}\right) \odot \left(\boldsymbol{\ell}_{n_x^2} \otimes \operatorname{vec}'(\boldsymbol{\Sigma}_{yz})\right) \\
&\quad \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\
&\quad + 4\left(\boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}'_{n_y}\right) \odot \left(\boldsymbol{\ell}'_{n_z} \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_x}\right) \\
&\quad \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\
\operatorname{cov}(\mathbf{m}_3(\boldsymbol{x},\boldsymbol{x}), \mathbf{m}_4(\boldsymbol{y},\boldsymbol{z})) &= 0_{(n_x^2, n_y n_z)}
\end{aligned}$$

iv) Covariance with the fourth power:

$$
\begin{aligned}
cov\left(\mathbf{m}_4(\boldsymbol{x},\boldsymbol{x}),\mathbf{m}_4(\boldsymbol{y},\boldsymbol{z})\right) = {} & 4cov\left(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}),\mathbf{m}_2(\boldsymbol{y},\boldsymbol{z})\right)\odot cov\left(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}),\mathbf{m}_2(\boldsymbol{y},\boldsymbol{z})\right)\\
& + 4\left(vec(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot cov\left(\mathbf{m}_2(\boldsymbol{x},\boldsymbol{x}),\mathbf{m}_4(\boldsymbol{y},\boldsymbol{z})\right)\\
& + 2\left(\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes\boldsymbol{\ell}_{n_z}\otimes vecd'(\boldsymbol{\Sigma}_{yy})\right)\\
& \odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)\\
& + 2\left(\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz}\otimes\boldsymbol{\ell}_{n_y})\right)\\
& \odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\\
& + 2\left(vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes\boldsymbol{\ell}_{n_z}\otimes vecd'(\boldsymbol{\Sigma}_{yy})\right)\\
& \odot\left(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y}\right)\odot\left(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y}\right)\\
& + 2\left(vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz}\otimes\boldsymbol{\ell}_{n_y})\right)\\
& \odot\left(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy}\right)\odot\left(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy}\right)\\
& + 8\left(\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right)\\
& \odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)\\
& + 8\left(vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}\right)\odot\left(\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right)\\
& \odot\left(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy}\right)\odot\left(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y}\right)\\
& + 8\left(\boldsymbol{\Sigma}_{xy}\otimes\mathbf{1}_{n_x\times n_z}\right)\odot\left(\mathbf{1}_{n_x\times n_y}\otimes\boldsymbol{\Sigma}_{xz}\right)\\
& \odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y}\right)
\end{aligned}
$$

Proof. Again, the proof is straightforward –and tedious– algebra. ■

A.1.2 Multivariate Wiener-Kolmogorov filter

Consider a time-invariant linear state-space model, (1)-(2). Finally, assume that $\det(I_M - \mathbf{F}z) = 0$ implies $|z| > 0$, that is, all eigenvalues of $\mathbf{F}$ are inside the unit circle. We pose the following problem. We observe the double-infinite sequence $Y_\infty = \{\boldsymbol{y}_t\}_{t=-\infty}^{\infty}$ and seek to find the linear projection,

$$
\begin{bmatrix}\hat{\boldsymbol{\xi}}_{t-1|\infty}\\ \hat{\boldsymbol{u}}_{t|\infty}\end{bmatrix} = \mathcal{P}\left(\begin{bmatrix}\boldsymbol{\xi}_{t-1}\\ \boldsymbol{u}_t\end{bmatrix}\middle| Y_\infty\right) = \begin{bmatrix}\boldsymbol{\Xi}(L)\\ \boldsymbol{\Upsilon}(L)\end{bmatrix}\boldsymbol{y}_t
$$

where $\boldsymbol{\Xi}$ and $\boldsymbol{\Upsilon}$ are two-sided filters in the lag operator L. Let us denote

$$
\mathbf{F}^{-1}(L) = (I_M - \mathbf{F}L)^{-1} = \sum_{j=0}^{\infty}\mathbf{F}^j L^j
$$

and

$$\boldsymbol{\Psi}(L) = \mathbf{H}\mathbf{F}^{-1}(L)\mathbf{G} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j L^j$$

where $\boldsymbol{\Psi}_j = \mathbf{H}\mathbf{F}^j\mathbf{G}$ for all j. Also, notice that

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{u}_t$$

Lemma 3 *The filtering problem posed above has solution*

$$\begin{bmatrix} \boldsymbol{\Xi}(z) \\ \boldsymbol{\Upsilon}(z) \end{bmatrix} = \begin{bmatrix} z\mathbf{F}^{-1}(z)\mathbf{G} \\ I_K \end{bmatrix} \boldsymbol{\Psi}'^{-1}) \left[\boldsymbol{\Psi}(z)\boldsymbol{\Psi}'^{-1})\right]^{-1}$$

Proof. This is a well-known technique in the theory of signal-extraction. Consider the joint auto-covariance generating function for $\begin{pmatrix} \boldsymbol{y}_t' & \boldsymbol{u}_t' \end{pmatrix}'$ which is easily seen to be

$$g(z) = \begin{bmatrix} g_{yy}(z) & g_{yu}(z) \\ g_{uy}(z) & g_{uu}(z) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}(z)\boldsymbol{\Psi}'^{-1}) & \boldsymbol{\Psi}(z) \\ \boldsymbol{\Psi}'^{-1}) & I_K \end{bmatrix}$$

for any $z \in \mathbb{C}$. The Wiener-Kolmogorov filter for $\boldsymbol{u}_t$ is given by

$$\begin{aligned} \hat{\boldsymbol{u}}_{t|\infty} &= g_{uy}(L)g_{yy}^{-1}(L)\boldsymbol{y}_t \\ &= \boldsymbol{\Psi}'^{-1}) \left[\boldsymbol{\Psi}(L)\boldsymbol{\Psi}'^{-1})\right]^{-1} \boldsymbol{y}_t \end{aligned}$$

Moreover, because

$$\boldsymbol{\xi}_{t-1} = L\mathbf{F}^{-1}(L)\mathbf{G}\mathbf{u_t}$$

the filter for $\boldsymbol{\xi}_{t-1}$ follows. ■

A remark about this result is that the same formula can still be applied even if $\mathbf{F}$ has eigenvalues on or outside the unit circle provided some assumptions on initial conditions hold (see Bell (1984) for a complete discussion).

An obvious consequence of the preceding result is that we can write

$$\begin{bmatrix} \hat{\boldsymbol{\xi}}_{t-1|\infty} \\ \hat{\boldsymbol{u}}_{t|\infty} \end{bmatrix} = \begin{bmatrix} L\mathbf{F}^{-1}(L)\mathbf{G} \\ I_K \end{bmatrix} \boldsymbol{\Psi}'^{-1}) \left[\boldsymbol{\Psi}(L)\boldsymbol{\Psi}'^{-1})\right]^{-1} \boldsymbol{\Psi}(L)\mathbf{u}_t$$

A less obvious consequence is that the Wiener-Kolmogorov filter for this setting always has a finite-order VARMA representation with scalar autoregressive part.

Lemma 4 *There are positive integers p and q, a set of scalars $\phi_1 \dots \phi_p \in \mathbb{R}$ and a set of matrices $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1 \dots \boldsymbol{\Theta}_q \in \mathbb{R}^{(M+K)\times K}$ such that*

$$(1 - \phi_1 L - \cdots - \phi_p L^p) \begin{bmatrix} \hat{\boldsymbol{\xi}}_{t-1|\infty} \\ \hat{\boldsymbol{u}}_{t|\infty} \end{bmatrix} = (\boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 L + \cdots + \boldsymbol{\Theta}_q L^q)\boldsymbol{u}_t$$

Proof. It follows from the fact that the auto-covariance generating functions for this model are rational polynomials. ∎

This is a useful result to the extent that for the large class of models considered in our paper the coefficients $\phi_1 \dots \phi_p$ and matrices $\mathbf{\Theta}_0, \mathbf{\Theta}_1 \dots \mathbf{\Theta}_q$ can be obtained in terms of the parametrisation of $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{G}$ and the algorithm proposed in the next sub-section can be employed to compute the auto-covariances of $\begin{pmatrix} \hat{\boldsymbol{\xi}}'_{t-1|\infty} & \hat{\boldsymbol{u}}'_{t|\infty} \end{pmatrix}'$

A.1.3 Algorithm for computing auto-covariances of AR-VMA processes

Consider an AR-VMA process for a K_x-dimensional process $\boldsymbol{x}_t$,

$$\phi(L)\boldsymbol{x}_t = \mathbf{\Theta}(L)\boldsymbol{u}_t$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\mathbf{\Theta}(z) = \mathbf{\Theta_0} + \mathbf{\Theta_1} z + \dots \mathbf{\Theta_q} z^q$. The error process $\boldsymbol{u}_t$ is K-dimensional and it is assumed to be white noise,

$$E(\boldsymbol{u}_t) = 0_{(K,1)}$$
$$E(\boldsymbol{u}_t \boldsymbol{u}'_{t-j}) = \begin{cases} \mathbf{\Sigma} & j = 0 \\ 0_{(K,K)} & j \neq 0 \end{cases}$$

Finally, all along this sub-section we assume $|z| = 1$ implies $\phi(z) \neq 0$. Of course, it is also the case that $\phi(0) \neq 0$, yet roots of ϕ might lie inside or outside the unit circle. Under these conditions there is a unique stable process satisfying the difference equation above. A deeper result is the following:

Lemma 5 *The auto-covariance function of the Wiener-Kolmogorov filter derived in the previous sub-section is the auto-covariance function of the stable solution to the difference equation embodied in its AR-VMA representation.*

Proof. It is an immediate consequence of uniqueness. ∎

Next, we present a method to compute the auto-covariance function of the stable solution to the AR-VMA difference equation.

Let us define some auxiliary matrices,

$$\bar{\mathbf{\Phi}} = \begin{bmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ & I_{p-1} & & 0_{(p-1,1)} \end{bmatrix} \quad \bar{\mathbf{\Theta}} = \begin{bmatrix} \mathbf{\Theta}_1 & \cdots & \mathbf{\Theta}_q \end{bmatrix}$$
$$\mathbf{J}_q = \begin{bmatrix} 0_{(1,q-1)} & 0 \\ I_{q-1} & 0_{(q-1,1)} \end{bmatrix} \quad \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0_{(p-1,1)} \end{bmatrix}$$

We can write the AR-VMA as a Hyper-VAR(1),

$$\boldsymbol{X}_t = \mathbf{A}\boldsymbol{X}_{t-1} + \mathbf{Q}\boldsymbol{u}_t$$

by setting

$$\boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \vdots \\ \boldsymbol{x}_{t-p+1} \\ \boldsymbol{u}_t \\ \vdots \\ \boldsymbol{u}_{t-q+1} \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} \bar{\boldsymbol{\Phi}} \otimes I_{K_x} & \mathbf{e}_1 \otimes \bar{\boldsymbol{\Theta}} \\ 0_{(Kq,K_xp)} & \mathbf{J}_q \otimes I_K \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} \boldsymbol{\Theta}_0 \\ 0_{(K_x(p-1),K)} \\ I_K \\ 0_{(K(q-1),K)} \end{bmatrix}$$

Suppose we can find an invertible matrix $\mathbf{C}$ and a block diagonal matrix $\boldsymbol{\Lambda}$ (with Jordan blocks) such that

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-1}$$

Then, we can transform the original system by defining $\boldsymbol{Z}_t = \mathbf{C}^{-1}\boldsymbol{X}_t$, a possibly complex-valued stochastic process that satisfies

$$\boldsymbol{Z}_t = \boldsymbol{\Lambda}\boldsymbol{Z}_{t-1} + \boldsymbol{\eta}_t$$

being $\boldsymbol{\eta}_t = \mathbf{C}^{-1}\mathbf{Q}\boldsymbol{u}_t$ white-noise (and possibly complex-valued).

The following lemma supplies a computationally convenient decomposition of $\mathbf{A}$.

Lemma 6 *Let $\bar{\boldsymbol{\Phi}} = \bar{\mathbf{C}}\bar{\boldsymbol{\Lambda}}\bar{\mathbf{C}}^{-1}$ be the Jordan decomposition of $\bar{\boldsymbol{\Phi}}$. Furthermore, let*

$$\boldsymbol{\Theta}^* = \sum_{h=1}^{q} (\bar{\boldsymbol{\Phi}}^{q-h}\mathbf{e}_1 \otimes \bar{\boldsymbol{\Theta}})(\mathbf{J}_q^{h-1} \otimes I_K)$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \bar{\boldsymbol{\Lambda}} \otimes I_{K_x} & 0_{(K_xp,Kq)} \\ 0_{(Kq,K_xp)} & \mathbf{J}_q \otimes I_K \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \bar{\mathbf{C}} \otimes I_{K_x} & -(\bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x})\boldsymbol{\Theta}^* \\ 0_{(Kq,K_xp)} & I_{Kq} \end{bmatrix}$$

$$\mathbf{C}^{-1} = \begin{bmatrix} \bar{\mathbf{C}}^{-1} \otimes I_{K_x} & (\bar{\mathbf{C}}^{-1}\bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x})\boldsymbol{\Theta}^* \\ 0_{(Kq,K_xp)} & I_{Kq} \end{bmatrix}$$

Then,

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-1}$$

Proof. It is a straightforward calculation. For the North-West block we get

$$\begin{aligned} \bar{\boldsymbol{\Phi}} \otimes I_{K_x} &= \bar{\mathbf{C}}\bar{\boldsymbol{\Lambda}}\bar{\mathbf{C}}^{-1} \otimes I_{K_x} \\ &= \begin{bmatrix} \bar{\mathbf{C}} \otimes I_{K_x} & -(\bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x})\boldsymbol{\Theta}^* \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\Lambda}} \otimes I_{K_x} \\ 0_{(Kq,K_xp)} \end{bmatrix} \bar{\mathbf{C}}^{-1} \otimes I_{K_x} \end{aligned}$$

For the South-West block we have

$$0_{(Kq,K_xp)} = \begin{bmatrix} 0_{(Kq,K_xp)} & \mathbf{J}_q \otimes I_K \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{C}}^{-1} \otimes I_{K_x} \\ 0_{(Kq,K_xp)} \end{bmatrix}$$

For the South-East block we get

$$\mathbf{J}_q \otimes I_K = \begin{bmatrix} 0_{(Kq,K_xp)} & \mathbf{J}_q \otimes I_K \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{C}}^{-1} \bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x} \boldsymbol{\Theta}^* \\ I_{Kq} \end{bmatrix}$$

Finally, for the North-East block we have

$$\mathbf{e}_1 \otimes \bar{\boldsymbol{\Theta}} = \begin{bmatrix} \bar{\mathbf{C}} \bar{\boldsymbol{\Lambda}} \otimes I_{K_x} & -(\bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x}) \boldsymbol{\Theta}^* (\mathbf{J}_q \otimes I_K) \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{C}}^{-1} \bar{\boldsymbol{\Phi}}^{-q} \otimes I_{K_x} \boldsymbol{\Theta}^* \\ I_{Kq} \end{bmatrix}$$

And this concludes the proof. ■

Now, assume we can compute the auto-covariance function of $\boldsymbol{Z}_t$,

$$\Gamma_{\boldsymbol{Z}}(j) = E\left[\boldsymbol{Z}_t \overline{\boldsymbol{Z}}'_{t-j}\right]$$

where $\overline{\mathbf{z}}$ denotes the conjugate of $\mathbf{z}$ for any $\mathbf{z} \in \mathbb{C}^n$ (n a positive integer). Then, we can recover the auto-covariance function of $\boldsymbol{X}_t$ from

$$\Gamma_{\boldsymbol{X}}(j) = E\left[\boldsymbol{X}_t \boldsymbol{X}'_{t-j}\right] = E\left[(\mathbf{C}\boldsymbol{Z}_t)(\overline{\mathbf{C}\boldsymbol{Z}}_{t-j})'\right] = \mathbf{C}\Gamma_{\boldsymbol{Z}}(j)\overline{\mathbf{C}}'$$

and, of course, the auto-covariance function of $\boldsymbol{x}_t$ is the first block of $\Gamma_{\boldsymbol{X}}$.

We decompose $\mathbf{A}$ following lemma A.6 and assume eigenvalues are in decreasing order of absolute value. Let K_{U} be the number of roots outside the unit circle and $K_{\mathrm{S}} = K_x p + Kq - K_{\mathrm{U}}$ the number of roots inside the unit circle. Remember we assumed no unit roots.

Let us denote by $\mathbf{R} = \mathbf{C}\mathbf{Q}\overline{\mathbf{Q}}'\overline{\mathbf{C}}'$ the variance-covariance matrix of $\boldsymbol{\eta}_t$. We partition the system into unstable and stable parts as follows:

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{Z}_{\mathrm{U}t} \\ \boldsymbol{Z}_{\mathrm{S}t} \end{bmatrix} \quad \boldsymbol{\eta}_t = \begin{bmatrix} \boldsymbol{\eta}_{\mathrm{U}t} \\ \boldsymbol{\eta}_{\mathrm{S}t} \end{bmatrix} \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathrm{UU}} & 0_{K_{\mathrm{U}} \times K_{\mathrm{S}}} \\ 0_{K_{\mathrm{S}} \times K_{\mathrm{U}}} & \boldsymbol{\Lambda}_{\mathrm{SS}} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}_{\mathrm{UU}} & \mathbf{R}_{\mathrm{US}} \\ \mathbf{R}_{\mathrm{SU}} & \mathbf{R}_{\mathrm{SS}} \end{bmatrix}$$

Next we write

$$\boldsymbol{Z}_{\mathrm{U}t} = \boldsymbol{\Lambda}_{\mathrm{UU}}^{-1}(\boldsymbol{Z}_{\mathrm{U}t+1} - \boldsymbol{\eta}_{\mathrm{U}t+1})$$

$$\boldsymbol{Z}_{\mathrm{S}t} = \boldsymbol{\Lambda}_{\mathrm{SS}} \boldsymbol{Z}_{\mathrm{S}t-1} + \boldsymbol{\eta}_{\mathrm{S}t}$$

We also partition

$$\Gamma_{\boldsymbol{Z}}(j) = \begin{bmatrix} \Gamma_{\mathrm{UU}}(j) & \Gamma_{\mathrm{US}}(j) \\ \Gamma_{\mathrm{SU}}(j) & \Gamma_{\mathrm{SS}}(j) \end{bmatrix} = \begin{bmatrix} E\left(\boldsymbol{Z}_{\mathrm{U}t}\overline{\boldsymbol{Z}}'_{\mathrm{U}t-j}\right) & E\left(\boldsymbol{Z}_{\mathrm{U}t}\overline{\boldsymbol{Z}}'_{\mathrm{S}t-j}\right) \\ E\left(\boldsymbol{Z}_{\mathrm{S}t}\overline{\boldsymbol{Z}}'_{\mathrm{U}t-j}\right) & E\left(\boldsymbol{Z}_{\mathrm{S}t}\overline{\boldsymbol{Z}}'_{\mathrm{S}t-j}\right) \end{bmatrix}$$

We summarize the main result in a lemma.

Lemma 7 *The auto-covariance function of $\mathbf{Z}_t$ can be computed from*

$$\begin{aligned}
\operatorname{vec}\left(\Gamma_{\mathrm{UU}}(0)\right) &= \left(I_{K_{\mathrm{U}}^2} - (\mathbf{\Lambda}_{\mathrm{UU}}^{-1} \otimes \mathbf{\Lambda}_{\mathrm{UU}}^{-1})\right)^{-1} \operatorname{vec}\left(\mathbf{\Lambda}_{\mathrm{UU}}^{-1}\mathbf{R}_{\mathrm{UU}}(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-1})'\right) \\
\Gamma_{\mathrm{UU}}(j) &= \Gamma_{\mathrm{UU}}(0)(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-j})' \\
\Gamma_{\mathrm{UU}}(j) &= \overline{\Gamma}'_{\mathrm{UU}}(-j) \\
\operatorname{vec}\left(\Gamma_{\mathrm{SS}}(0)\right) &= \left(I_{K_{\mathrm{S}}^2} - (\mathbf{\Lambda}_{\mathrm{SS}} \otimes \mathbf{\Lambda}_{\mathrm{SS}})\right)^{-1} \operatorname{vec}\left(\mathbf{R}_{\mathrm{SS}}\right) \\
\Gamma_{\mathrm{SS}}(j) &= \mathbf{\Lambda}_{\mathrm{SS}}^{j}\Gamma_{\mathrm{SS}}(0) \\
\Gamma_{\mathrm{SS}}(j) &= \overline{\Gamma}'_{\mathrm{SS}}(-j) \\
\Gamma_{\mathrm{SU}}(j) &= -\sum_{h=1}^{j}(\overline{\mathbf{\Lambda}}_{\mathrm{SS}}^{j-h})'\mathbf{R}_{\mathrm{SU}}(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-h})' \\
\Gamma_{\mathrm{SU}}(j) &= 0 \\
\Gamma_{\mathrm{US}}(j) &= \overline{\Gamma}'_{\mathrm{SU}}(-j)
\end{aligned}$$

Proof. It follows from simple algebra. ■

A.2 Score

In this section we establish proposition 1. As anticipated, the argument is based on a direct application of the EM principle and, with the purpose to keep things clear, we divide it into three steps. All notation is as developed in the paper.

A.2.1 Score of the latent model

The state-space model considered in the paper under the alternative hypothesis is

$$\begin{aligned}
\boldsymbol{y}_t &= \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\xi}_t \\
\boldsymbol{\xi}_t &= \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} + \mathbf{G}(\boldsymbol{\theta})\boldsymbol{u}_t
\end{aligned}$$

together with the distributional assumptions,

$$\begin{aligned}
\boldsymbol{u}_t^{\mathrm{GH}}|\mathcal{I}_{t-1}, \boldsymbol{\phi} &\sim \mathrm{GH}(0_{(R,1)}, I_R, \boldsymbol{\beta}, \eta, \psi) \\
\boldsymbol{u}_t^{\mathrm{GS}}|\mathcal{I}_{t-1}, \boldsymbol{\phi} &\sim \mathrm{N}(0_{(K-R,1)}, I_{K-R}) \\
\boldsymbol{u}_t^{\mathrm{GH}} &\perp \boldsymbol{u}_t^{\mathrm{GS}}
\end{aligned}$$

where $\boldsymbol{\phi} = \begin{pmatrix}\boldsymbol{\theta}' & \boldsymbol{\beta}' & \eta & \psi\end{pmatrix}'$ is the full set of parameters including $\boldsymbol{\beta} \in \mathbb{R}^R$, $\eta \in \mathbb{R}$ and $\psi \in \mathbb{R}_+$ which parametrise the Generalised Hyperbolic distribution.

Let us denote $Y_T = (\boldsymbol{y}_1 \dots \boldsymbol{y}_T)$ and $\Xi_T = (\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_T)$. Let $\boldsymbol{\phi} \mapsto f(Y_T, \Xi_T|\boldsymbol{\phi})$ be the likelihood of (Y_T, Ξ_T). This is the likelihood function of the latent model. Mencía and Sentana (2012) derived the scores with respect to mean-variance and shape parameters for a general class of dynamic regression models that includes the one we study here. We state the result below.

Lemma 8 *Let* $\varsigma_t(\boldsymbol{\theta}) = (\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}))'\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})$ *with* $\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})$ *as defined in the text and let*

$$s_{\mathrm{k},t}(\boldsymbol{\theta}) = c_0 + c_1\varsigma_t(\boldsymbol{\theta}) + c_2\varsigma_t^2(\boldsymbol{\theta})$$
$$\mathbf{s}_{\mathrm{s},t}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\,[c_3 + \varsigma_t(\boldsymbol{\theta})]$$
$$s_{\mathrm{GH},t}(\boldsymbol{\theta}) = s_{\mathrm{k},t}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s},t}(\boldsymbol{\theta})$$

where $c_0 = R(R+2)/4$, $c_1 = -(R+2)/2$, $c_2 = 1/4$ *and* $c_3 = -(R+2)$. *Then,*

i) For any $\boldsymbol{\beta} \in \mathbb{R}^R$ *and* $\psi > 0$,

$$\lim_{\eta\to 0^+}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \eta}\right) = \frac{1}{T}\sum_{t=1}^{T} s_{\mathrm{GH},t}(\boldsymbol{\theta})$$
$$-\lim_{\eta\to 0^-}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \eta}\right) = \frac{1}{T}\sum_{t=1}^{T} s_{\mathrm{GH},t}(\boldsymbol{\theta})$$
$$\lim_{\eta\to 0^\pm}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \psi}\right) = 0$$

ii) For any $\boldsymbol{\beta} \in \mathbb{R}^R$ *and* $\psi > 0$,

$$\lim_{\psi\to 0^+}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \eta}\right) = 0$$
$$\lim_{\psi\to 0^+}\left(\frac{2}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \psi}\right) = \frac{1}{T}\sum_{t=1}^{T} s_{\mathrm{GH},t}(\boldsymbol{\theta})$$

iii) In any case,

$$\lim_{\eta\cdot\psi\to 0}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \boldsymbol{\beta}}\right) = 0_{(R,1)}$$
$$\lim_{\eta\cdot\psi\to 0}\left(\frac{1}{T}\frac{\partial \log f(Y_T, \Xi_T|\boldsymbol{\phi})}{\partial \boldsymbol{\theta}}\right) = \text{Gaussian : score}$$

Proof. See Mencía and Sentana (2012). ■

The EM principle instructs us to smooth $s_{\mathrm{GH},t}(\boldsymbol{\theta})$, which can be decomposed into a kurtosis component $s_{\mathrm{k},t}(\boldsymbol{\theta})$ and a skewness component $\mathbf{s}_{\mathrm{s},t}(\boldsymbol{\theta})$. We do this in turn. Proposition 1 will be an immediate consequence of lemmas B.2 and B.3 below.

A.2.2 Smoothing the kurtosis score

Lemma 9 *Under the null we can write the kurtosis component as*

$$
\begin{aligned}
s_{\mathrm{k},t|T}(\boldsymbol{\theta}) &= \mathbb{E}\left(s_{\mathrm{k},t}(\boldsymbol{\theta}) : | : Y_T\right) \\
&= \mathbf{b}'_{\mathrm{k},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{k},t|T}(\boldsymbol{\theta}) = \begin{pmatrix} b_{0,t|T}(\boldsymbol{\theta}) & \mathbf{b}'_{2,t|T}(\boldsymbol{\theta}) & \mathbf{b}'_{4,t|T}(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{m}_{2,t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{4,t|T}(\boldsymbol{\theta}) \end{pmatrix}
\end{aligned}
$$

for any $\boldsymbol{\theta}$, where we have defined

$$
\begin{aligned}
b_{0,t|T}(\boldsymbol{\theta}) &= c_0 + (c_1 + \mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))c_2)\mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})) + 2c_2\mathrm{tr}((\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))^2) \\
\mathbf{b}_{2,t|T}(\boldsymbol{\theta}) &= (c_1 + 2\mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))c_2)\left[\mathbf{E}'_{RK} \otimes \mathbf{E}'_{RK}\right]\mathrm{vec}(I_R) + 4c_2\left[\mathbf{E}'_{RK} \otimes \mathbf{E}'_{RK}\right]\mathrm{vec}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})) \\
\mathbf{b}_{4,t|T}(\boldsymbol{\theta}) &= c_2\left[\mathbf{E}'_{RK} \otimes \mathbf{E}'_{RK}\right]\boldsymbol{\ell}_{R^2}
\end{aligned}
$$

with $\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) = \mathbf{E}_{RK}\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})\mathbf{E}'_{RK}$ and

$$
\begin{aligned}
\mathbf{m}_{2,t|T}(\boldsymbol{\theta}) &= \mathrm{vec}\left[\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})'\right] \\
\mathbf{m}_{4,t|T}(\boldsymbol{\theta}) &= \mathrm{vec}\left[(\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}))(\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}))'\right]
\end{aligned}
$$

Proof. Under the null of Gaussianity and if $\boldsymbol{\theta}$ was the true value of the mean-variance parameter vector $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})|Y_T, \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta}))$ (for ease of notation we omit $\boldsymbol{\theta}$ at the conditioning from now on). Consider

$$
s_{\mathrm{k},t|T}(\boldsymbol{\theta}) = c_0 + c_1 E\left[\varsigma_t(\boldsymbol{\theta})|Y_T\right] + c_2 E\left[\varsigma^2_t(\boldsymbol{\theta})|Y_T\right]
$$

We apply lemma A.1.i) to compute the first expectation,

$$
\begin{aligned}
E\left[\varsigma_t(\boldsymbol{\theta})|Y_T\right] &= E\left[(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}))'\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})|Y_T\right] \\
&= \mathrm{tr}\left(E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}))'|Y_T\right]\right) \\
&= \mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})) + \mathrm{vec}(I_R)'\mathrm{vec}\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))'\right]
\end{aligned}
$$

We apply lemma A.1.iii) to compute the second expectation,

$$
\begin{aligned}
E\left[\varsigma_t^2(\boldsymbol{\theta}) : | : Y_T\right] &= E\left[(\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}))'\mathbf{1}_{R\times R}(\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})) : | : Y_T\right] \\
&= \mathrm{tr}\left(\mathbf{1}_{R\times R}E\left[(\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}))(\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta}))' : | : Y_T\right]\right) \\
&= 2\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\theta}) \odot \boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\right] \\
&+ \boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))'\right] \\
&+ 4\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right] \\
&+ \boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))\mathrm{vecd}(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))')'\right] \\
&+ \boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\mathrm{vecd}(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))')\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))'\right] \\
&+ \boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right]
\end{aligned}
$$

In re-arranging the formulas the following identities are useful

$$
\begin{aligned}
\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\right] &= \mathrm{vec}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))'\mathrm{vec}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})) \\
&= \mathrm{tr}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})) = \mathrm{tr}((\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))^2) \\
\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))'\right] &= \mathrm{tr}^2(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})) \\
\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right] &= \mathrm{vec}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))'\mathrm{vec}\left[\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right] \\
\boldsymbol{\ell}'_{R^2}\mathrm{vec}\left[\mathrm{vecd}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))\mathrm{vecd}(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))')'\right] &= \mathrm{tr}(\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}))\mathrm{vec}(I_R)' \\
&\times \mathrm{vec}\left[\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right]
\end{aligned}
$$

It is also useful to note

$$
\begin{aligned}
\mathrm{vec}\left[\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})'\right] &= \left[\mathbf{E}_{RK} \otimes \mathbf{E}_{RK}\right]\mathbf{m}_{2,t|T}(\boldsymbol{\theta}) \\
\mathrm{vec}\left[(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))(\boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta}))'\right] &= \left[\mathbf{E}_{RK} \otimes \mathbf{E}_{RK}\right]\mathbf{m}_{4,t|T}(\boldsymbol{\theta})
\end{aligned}
$$

This delivers the result. ■

A.2.3 Smoothing the skewness score

Lemma 10 *Under the null we can write the skewness component as*

$$
\begin{aligned}
\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) &= E\left(\mathbf{s}_{\mathrm{s},t}(\boldsymbol{\theta})|Y^T\right) \\
&= \mathbf{b}'_{\mathrm{s},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{s},t|T}(\boldsymbol{\theta}) = \begin{pmatrix}\mathbf{b}'_{1,t|T}(\boldsymbol{\theta}) & \mathbf{b}'_{3,t|T}(\boldsymbol{\theta})\end{pmatrix}\begin{pmatrix}\mathbf{m}_{1,t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{3,t|T}(\boldsymbol{\theta})\end{pmatrix}
\end{aligned}
$$

for any $\boldsymbol{\theta}$, where we have defined

$$\mathbf{b}_{1,t|T}(\boldsymbol{\theta}) = (c_3 + \mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})))\mathbf{E}'_{RK} + 2\mathbf{E}'_{RK}\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})$$
$$\mathbf{b}_{3,t|T}(\boldsymbol{\theta}) = \left[\mathbf{E}'_{RK}\boldsymbol{\ell}_R \otimes \mathbf{E}'_{RK}\right]$$

with $\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) = \mathbf{E}_{RK}\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})\mathbf{E}'_{RK}$ and

$$\mathbf{m}_{1,t|T}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})$$
$$\mathbf{m}_{3,t|T}(\boldsymbol{\theta}) = \mathrm{vec}\left[\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}))'\right]$$

Proof. Under the null of Gaussianity and if $\boldsymbol{\theta}$ was the true value of the mean-variance parameter vector $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})|Y^T, \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta}))$ (for ease of notation we omit $\boldsymbol{\theta}$ at the conditioning from now on). Consider

$$\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) = c_3 E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})|Y_T\right] + E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})\varsigma_t(\boldsymbol{\theta})|Y_T\right]$$

The first expectation is trivially $E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})|Y_T\right] = \boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta})$.

We apply lemma A.1.ii) to compute the second expectation,

$$\begin{aligned} E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})\varsigma_t(\boldsymbol{\theta}) : | : Y_T\right] &= E\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}))' : | : Y_T\right]\boldsymbol{\ell}_R \\ &= 2\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}) + \mathrm{tr}(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta}) \\ &+ \boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))'\boldsymbol{\ell}_R \end{aligned}$$

It is also useful to note

$$\mathrm{vec}\left[\boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta})(\boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{*\mathrm{GH}}_{t|T}(\boldsymbol{\theta}))'\right] = \left[\mathbf{E}_{RK} \otimes \mathbf{E}_{RK}\right]\mathbf{m}_{3,t|T}(\boldsymbol{\theta})$$

Re-arranging things we get the result. ■

A.3 Asymptotic Variance

In this section we provide a formula for the asymptotic variance of the average score $\bar{s}_{\mathrm{GH}|T}$ and its components, $\bar{s}_{\mathrm{k}|T}$ and $\bar{\mathbf{s}}_{\mathrm{s}|T}$. We also establish propositions 2 and 3 in the paper. Importantly, we supply a numerically reliable algorithm to compute the asymptotic variance which can be applied to any state-space model.

As commented in the text there are two difficulties that we need to overcome to obtain the asymptotic variance: (i) the smoothed score functions, $\bar{s}_{\mathrm{k}|T}$ and $\bar{\mathbf{s}}_{\mathrm{s}|T}$, display serial correlation in dynamic models and (ii) the serial correlation structure of the individual contributions to the smoothed scores, $s_{\text{k},t|T}$ and $\mathbf{s}_{\text{s},t|T}$, change with both t and T. Our solution is to approximate the large sample behaviour of $\sqrt{T}\bar{s}_{\text{k}|T}$ and $\sqrt{T}\bar{\mathbf{s}}_{\text{k}|T}$ with that of $\sqrt{T}\bar{s}_{\text{k}|\infty}$ and $\sqrt{T}\bar{\mathbf{s}}_{\text{k}|\infty}$, the corresponding smoothed scores based on the Wiener-Kolmogorov filter.

Under the null of Gaussianity and if $\boldsymbol{\theta}$ was the true value of the mean-variance parameter vector

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})|Y_\infty, \boldsymbol{\theta} \sim \text{N}(\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}), \boldsymbol{\Omega}_\infty(\boldsymbol{\theta}))$$

being $\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta})$ the Wiener-Kolmogorov filter of $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ based on $Y_\infty = \{\boldsymbol{y}_t\}_{t=-\infty}^{\infty}$ and $\boldsymbol{\Omega}_\infty(\boldsymbol{\theta})$ its mean-square error.

Let $s_{\text{k},t|\infty}(\boldsymbol{\theta}) = E[s_{\text{k},t}(\boldsymbol{\theta})|Y_\infty]$ and $\mathbf{s}_{\text{s},t|\infty}(\boldsymbol{\theta}) = E[\mathbf{s}_{\text{s},t}(\boldsymbol{\theta})|Y_\infty]$. Of course, $s_{\text{GH},t|\infty}(\boldsymbol{\theta})$ is defined analogously. The argument developed in the preceding section can be reproduced with virtually no change to prove:

Lemma 11 *Under the null of Gaussianity,*

$$s_{\text{k},t|\infty}(\boldsymbol{\theta}) = \mathbf{b}_\text{k}'(\boldsymbol{\theta})\mathbf{m}_{\text{k},t|\infty}(\boldsymbol{\theta}) = \begin{pmatrix} b_0(\boldsymbol{\theta}) & \mathbf{b}_2'(\boldsymbol{\theta}) & \mathbf{b}_4'(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}) \\ \mathbf{m}_{4,t|\infty}(\boldsymbol{\theta}) \end{pmatrix}$$
$$\mathbf{s}_{\text{s},t|\infty}(\boldsymbol{\theta}) = \mathbf{b}_\text{s}'(\boldsymbol{\theta})\mathbf{m}_{\text{s},t|\infty}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{b}_1'(\boldsymbol{\theta}) & \mathbf{b}_3'(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}) \\ \mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}) \end{pmatrix}$$

for any $\boldsymbol{\theta}$, *where we have defined*

$$b_0(\boldsymbol{\theta}) = c_0 + (c_1 + \text{tr}(\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}))c_2)\text{tr}(\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})) + 2c_2\text{tr}((\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}))^2)$$
$$\mathbf{b}_1(\boldsymbol{\theta}) = (c_3 + \text{tr}(\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})))\mathbf{E}_{RK}' + 2\mathbf{E}_{RK}'\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})$$
$$\mathbf{b}_2(\boldsymbol{\theta}) = (c_1 + 2\text{tr}(\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}))c_2)\left[\mathbf{E}_{RK}' \otimes \mathbf{E}_{RK}'\right]\text{vec}(I_R) + 4c_2\left[\mathbf{E}_{RK}' \otimes \mathbf{E}_{RK}'\right]\text{vec}(\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}))$$
$$\mathbf{b}_3(\boldsymbol{\theta}) = \left[\mathbf{E}_{RK}'\boldsymbol{\ell}_R \otimes \mathbf{E}_{RK}'\right]$$
$$\mathbf{b}_4(\boldsymbol{\theta}) = c_2\left[\mathbf{E}_{RK}' \otimes \mathbf{E}_{RK}'\right]\boldsymbol{\ell}_{R^2}$$

with $\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}) = \mathbf{E}_{RK}\boldsymbol{\Omega}_\infty(\boldsymbol{\theta})\mathbf{E}_{RK}'$ *and*

$$\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta})$$
$$\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}) = \text{vec}\left[\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta})'\right]$$
$$\mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}) = \text{vec}\left[\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta})(\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}))'\right]$$
$$\mathbf{m}_{4,t|\infty}(\boldsymbol{\theta}) = \text{vec}\left[(\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}))(\boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^*(\boldsymbol{\theta}))'\right]$$

Proof. Same as lemmas B.2 and B.3. ■

More compactly we can write

$$\begin{bmatrix} s_{\mathrm{k},t|\infty}(\boldsymbol{\theta}) - b_0(\boldsymbol{\theta}) \\ \mathbf{s}_{\mathrm{s},t|\infty}(\boldsymbol{\theta}) \end{bmatrix} = \mathbf{b}'(\boldsymbol{\theta})\mathbf{m}_{t|\infty}(\boldsymbol{\theta})$$

where

$$\mathbf{b}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{B}_{\mathrm{k}}(\boldsymbol{\theta}) & \mathbf{B}_{\mathrm{s}}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} 0_{(K,1)} & \mathbf{b}_1(\boldsymbol{\theta}) \\ \mathbf{b}_2(\boldsymbol{\theta}) & 0_{(K^2,R)} \\ 0_{(K^2,1)} & \mathbf{b}_3(\boldsymbol{\theta}) \\ \mathbf{b}_4(\boldsymbol{\theta}) & 0_{(K^2,R)} \end{bmatrix} \qquad \mathbf{m}_{t|\infty}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{m}_{1,t|\infty} \\ \mathbf{m}_{2,t|\infty} \\ \mathbf{m}_{3,t|\infty} \\ \mathbf{m}_{4,t|\infty} \end{bmatrix}$$

A.3.1 Auto-covariance generating function of $\mathbf{m}_{t|\infty}$

In this section everything is evaluated under the null of Gaussianity and assuming the true mean-variance parameter vector is $\boldsymbol{\theta}$. Furthermore, to simplify the notation we omit $\boldsymbol{\theta}$ as an argument in the functions defined above.

Lemma 12 *Let $\Gamma_0 = E[\boldsymbol{\varepsilon}^*_{t|\infty}(\boldsymbol{\varepsilon}^*_{t|\infty})']$ and $\Gamma_j = E[\boldsymbol{\varepsilon}^*_{t|\infty}(\boldsymbol{\varepsilon}^*_{t-j|\infty})']$ be the 0-th and j-th order auto-covariance matrix of the Wiener-Kolmogorov filter for $\boldsymbol{\varepsilon}^*_t$ based on Y_∞ for any integer j. Then,*
i) Covariance with the first power:

$$\begin{aligned}
\mathrm{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right) &= 0_{(K,K^2)} \\
\mathrm{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{3,t-j|\infty}\right) &= 2\left(\boldsymbol{\ell}_K \otimes \mathrm{vec}'(\Gamma_0)\right) \odot \left(\Gamma_j \otimes \boldsymbol{\ell}'_K\right) \\
&\quad + \left(\mathrm{vecd}'(\Gamma_0) \otimes \mathbf{1}_{K\times K}\right) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j\right) \\
\mathrm{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right) &= 0_{(K,K^2)}
\end{aligned}$$

ii) Covariance with the second power:

$$\begin{aligned}
\mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right) &= (\mathbf{1}_{K\times K} \otimes \Gamma_j) \odot (\Gamma_j \otimes \mathbf{1}_{K\times K}) \\
&\quad + \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \\
\mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{3,t-j|\infty}\right) &= 0_{(K^2,K^2)} \\
\mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right) &= 4\left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vec}'(\Gamma_0)\right) \odot \mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right) \\
&\quad + 2\left(\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}'_K \otimes \mathrm{vecd}'(\Gamma_0)\right) \\
&\quad \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \odot (\Gamma_j \otimes \mathbf{1}_{K\times K}) \\
&\quad + 2\left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vecd}'(\Gamma_0) \otimes \boldsymbol{\ell}'_K\right) \\
&\quad \odot (\mathbf{1}_{K\times K} \otimes \Gamma_j) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right)
\end{aligned}$$

iii)Covariance with the third power:

$$\begin{aligned}
\mathrm{cov}\left(\mathbf{m}_{3,t|\infty}, \mathbf{m}_{3,t-j|\infty}\right) &= \left(\mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}_K \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_K \otimes \mathrm{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{3,t-j|\infty}\right)\right) \\
&+ 2\left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \odot \left(\left(\Gamma_j \odot \Gamma_j\right) \otimes \mathbf{1}_{K\times K}\right) \\
&+ 2\left(\mathrm{vec}(\Gamma_0) \otimes \mathbf{1}_{1\times KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vecd}'(\Gamma_0) \otimes \boldsymbol{\ell}'_K\right) \\
&\odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \\
&+ 4\left(\mathrm{vec}(\Gamma_0) \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vec}'(\Gamma_0)\right) \\
&\odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \\
&+ 4\left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \\
&\odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \\
\mathrm{cov}\left(\mathbf{m}_{3,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right) &= 0_{(K^2,K^2)}
\end{aligned}$$

iv) Covariance with the fourth power:

$$\begin{aligned}
\mathrm{cov}\left(\mathbf{m}_{4,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right) &= 4\mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right) \odot \mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right) \\
&{::::}\ +4\left(\mathrm{vec}(\Gamma_0) \otimes \boldsymbol{\ell}'_{KK}\right) \odot \mathrm{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right) \\
&{::::}\ +2\left(\boldsymbol{\ell}_K \otimes \mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}_K \otimes \mathrm{vecd}'(\Gamma_0)\right) \\
&{::::}\ \odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \\
&{::::}\ +2\left(\boldsymbol{\ell}_K \otimes \mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vecd}'(\Gamma_0 \otimes \boldsymbol{\ell}_K)\right) \\
&{::::}\ \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \\
&{::::}\ +2\left(\mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}_K \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}_K \otimes \mathrm{vecd}'(\Gamma_0)\right) \\
&{::::}\ \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \\
&{::::}\ +2\left(\mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}_K \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vecd}'(\Gamma_0 \otimes \boldsymbol{\ell}_K)\right) \\
&{::::}\ \odot \left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \odot \left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \\
&{::::}\ +8\left(\boldsymbol{\ell}_K \otimes \mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vec}'(\Gamma_0)\right) \\
&{::::}\ \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \\
&{::::}\ +8\left(\mathrm{vecd}(\Gamma_0) \otimes \boldsymbol{\ell}_K \otimes \boldsymbol{\ell}'_{KK}\right) \odot \left(\boldsymbol{\ell}_{K^2} \otimes \mathrm{vec}'(\Gamma_0)\right) \\
&{::::}\ \odot \left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \\
&{::::}\ +8\left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right) \odot \left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \\
&{::::}\ \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right) \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right)
\end{aligned}$$

Proof. Follows directly from lemma A.2. ■

A.3.2 Proof of Proposition 2

Proposition 2 has two parts: (i) the asymptotic independence of the kurtosis and skewness component and (ii) the cancellation of cross-terms within the variance formulas.

The asymptotic independence follows from this (stronger) result,

Lemma 13 *For any integer j,*

$$\text{cov}(\mathbf{s}_{\text{s},t|\infty}, s_{\text{k},t-j|\infty}) = \mathbf{0}.$$

Proof. From lemma C.1 we can write

$$\begin{aligned}\text{cov}(\mathbf{s}_{\text{s},t|\infty}, s_{\text{k},t-j|\infty}) &= \mathbf{b}_1'\text{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\mathbf{b}_2 + \mathbf{b}_1'\text{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right)\mathbf{b}_4 \\ &:::: +\mathbf{b}_3'\text{cov}\left(\mathbf{m}_{3,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\mathbf{b}_2 + \mathbf{b}_3'\text{cov}\left(\mathbf{m}_{3,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right)\mathbf{b}_4\end{aligned}$$

and from lemma C.2 we know that the auto-covariance matrices involved are zero.■

On the other hand, the simplification of the formulas follows from the next result.

Lemma 14 *For any integer j,*

$$\text{cov}(\mathbf{m}_{1,t|\infty}, \mathbf{s}_{\text{s},t-j|\infty}) = 0_{(K,R)}$$

$$\text{cov}(\mathbf{m}_{2,t|\infty}, s_{\text{k},t-j|\infty}) = 0_{(K^2,1)}$$

Proof. We show how to prove the first result for the case when $R = K$. The proof for $R < K$ is similar. Notice that, because $\mathbf{\Omega}_\infty = I_K - \Gamma_0$,

$$\text{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{1,t-j|\infty}\right)\mathbf{b}_1 = -\left(2\Gamma_j\Gamma_0 + \text{tr}(\Gamma_0)\Gamma_j\right)$$

Now, it is easy to see that[10]

$$\Gamma_j\Gamma_0 = \left[\left(\boldsymbol{\ell}_K \otimes \text{vec}'(\Gamma_0)\right) \odot \left(\Gamma_j \otimes \boldsymbol{\ell}_K'\right)\right]\left(\boldsymbol{\ell}_K \otimes I_K\right)$$

$$\text{tr}(\Gamma_0)\Gamma_j = \left[\left(\text{vecd}'(\Gamma_0) \otimes \mathbf{1}_{K\times K}\right) \odot \left(\boldsymbol{\ell}_K' \otimes \Gamma_j\right)\right]\left(\boldsymbol{\ell}_K \otimes I_K\right)$$

[10] A useful property of Hadamard product is $\mathbf{x}_1'(\mathbf{A}_1 \odot \mathbf{A}_2)\mathbf{x}_2 = \text{tr}((\mathbf{A}_1'\text{diag}(\mathbf{x}_1)\mathbf{A}_2\text{diag}(\mathbf{x}_2))$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors of appropriate size and $\mathbf{A}_1$ and $\mathbf{A}_2$ are equally sized matrices. To check that the left-hand side and right-hand side of the following identities coincide we pre-multiply by $\mathbf{e}_i'$ (a vector with 1 in the i-th entry and 0s elsewhere) and post-multiply by $\mathbf{e}_j$ (a vector with 1 in the j-th entry and 0s elsewhere) and apply this property and those of the trace.

and, consequently,

$$\text{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{1,t-j|\infty}\right)\mathbf{b}_1 + \text{cov}\left(\mathbf{m}_{1,t|\infty}, \mathbf{m}_{3,t-j|\infty}\right)\mathbf{b}_3 = 0_{(K,K)}$$

For the second result (and again for $R = K$), notice that, because $\mathbf{\Omega}_\infty = I_K - \Gamma_0$,

$$\mathbf{b}_2 = -\frac{1}{2}\text{tr}(\Gamma_0)\text{vec}(I_K) - \text{vec}(\Gamma_0)$$

Now, it is easy to see that[11]

$$\begin{aligned}
\text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\text{vec}(\Gamma_0) &= \left[\left(\boldsymbol{\ell}_{K^2} \otimes \text{vec}'(\Gamma_0)\right) \odot \text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\right]\boldsymbol{\ell}_{K^2} \\
\text{tr}(\Gamma_0)\text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\text{vec}(I_K) &= 2\Big[\left(\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}'_K \otimes \text{vecd}'(\Gamma_0)\right) \\
&\quad \odot \left(\boldsymbol{\ell}_K \otimes \Gamma_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\Gamma_j \otimes \mathbf{1}_{K\times K}\right)\Big]\boldsymbol{\ell}_{K^2} \\
\text{tr}(\Gamma_0)\text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\text{vec}(I_K) &= 2\Big[\left(\boldsymbol{\ell}_{K^2} \otimes \text{vecd}'(\Gamma_0) \otimes \boldsymbol{\ell}'_K\right) \\
&\quad \odot \left(\mathbf{1}_{K\times K} \otimes \Gamma_j\right) \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j \otimes \boldsymbol{\ell}_K\right)\Big]\boldsymbol{\ell}_{K^2}
\end{aligned}$$

consequently,

$$\text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{2,t-j|\infty}\right)\mathbf{b}_2 + \text{cov}\left(\mathbf{m}_{2,t|\infty}, \mathbf{m}_{4,t-j|\infty}\right)\mathbf{b}_4 = 0_{(K^2,1)}$$

and this finishes the proof. ■

This allows us to write

$$\lim_{T\to\infty} \text{Var}\left(\begin{bmatrix}\sqrt{T}\bar{s}_{\text{k}|T}(\boldsymbol{\theta}_0) \\ \sqrt{T}\bar{s}_{\text{k}|T}(\boldsymbol{\theta}_0)\end{bmatrix}\right) = \begin{bmatrix}\mathcal{C}_{\text{k}|\infty}(\boldsymbol{\theta}_0) & 0_{(1,R)} \\ 0_{(R,1)} & \mathcal{C}_{\text{s}|\infty}(\boldsymbol{\theta}_0)\end{bmatrix}$$

where

$$\begin{aligned}
\mathcal{C}_{\text{k}|\infty}(\boldsymbol{\theta}_0) &= \mathbf{b}'_4(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_4(\boldsymbol{\theta}_0)\mathbf{b}_4(\boldsymbol{\theta}_0) - \mathbf{b}'_2(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_2(\boldsymbol{\theta}_0)\mathbf{b}_2(\boldsymbol{\theta}_0) \\
\mathcal{C}_{\text{s}|\infty}(\boldsymbol{\theta}_0) &= \mathbf{b}'_3(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_3(\boldsymbol{\theta}_0)\mathbf{b}_3(\boldsymbol{\theta}_0) - \mathbf{b}'_1(\boldsymbol{\theta}_0)\boldsymbol{\kappa}_1(\boldsymbol{\theta}_0)\mathbf{b}_1(\boldsymbol{\theta}_0)
\end{aligned}$$

with

$$\boldsymbol{\kappa}_h(\boldsymbol{\theta}_0) = \sum_{j=-\infty}^{\infty} \text{cov}(\mathbf{m}_{h,t|\infty}, \mathbf{m}_{h,t-j|\infty})$$

for $h = 1 \ldots 4$, as cross-terms cancel out.

A.3.3 Proof of Proposition 3

TBD

[11] Same remarks as in previous footnote.

B Algorithm for computing the asymptotic variance

Finally, we comment on a numerically reliable and efficient algorithm to obtain the asymptotic variance of the test statistics. As a preliminary step we suppose the researcher has (i) specified the model and (ii) computed the Maximum Likelihood estimates corresponding to Gaussianity, which we denote $\hat{\boldsymbol{\theta}}_T$ (if the model needs to be estimated).

Then, the researcher should proceed as follows:

1. Obtain the AR-VMA representation of the Wiener-Kolmogorov filter for the innovations. This can be done symbolic in Mathematica in terms of matrices $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{G}$ and evaluated at the Maximum Likelihood estimates afterwards. We refer the reader to lemma A.3.

2. Compute the auto-covariance function implied by the Wiener-Kolmogorov filter of the innovations using the techniques developed in lemmas A.6 and A.7. The decomposition outlined in lemma A.6 is convenient to handle large systems as it reduces the size of the matrices on which the Jordan decomposition needs to be performed.

3. From lemma C.2 obtain the auto-covariance function of $\mathbf{m}_{h,t|\infty}$ for $h = 1 \dots 4$.

4. Add auto-covariance matrices of $\mathbf{m}_{h,t|\infty}$ for $h = 1 \dots 4$ until some convergence criterion is satisfied. This gives a numerical approximation to $\boldsymbol{\kappa}_h(\hat{\boldsymbol{\theta}}_T)$ for $h = 1 \dots 4$.

5. Compute $\mathbf{b}_h(\hat{\boldsymbol{\theta}}_T)$ from lemma C.1, which only requires knowledge of the 0-th order auto-covariance matrix of the Wiener-Kolmogorov filter since $\boldsymbol{\Omega}_\infty = I_K - \Gamma_0$.

Codes for all the steps are available upon request.

C Tables and Figures

Table 1: Rejection rates (in %) under the null at 1%, 5%, and 10% significance levels

		Static FM			AR(1) + noise			RW + noise			Cointegrated		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
H_J	Kt	1.0	4.5	10.0	1.5	4.1	7.8	1.4	3.7	7.6	1.2	4.2	8.7
	Sk				1.2	5.2	9.9	1.2	5.0	9.8	1.0	4.5	8.7
	GH				1.9	5.2	9.0	1.9	4.9	8.6	1.4	4.6	8.4
$H_{S_\mathbf{f}}$	Kt	1.4	4.1	8.1	1.6	3.8	7.7	1.5	4.0	7.9	1.4	3.9	7.5
	Sk	0.8	4.6	9.7	1.2	4.7	9.3	1.1	4.9	10.0	1.1	4.4	9.0
	GH	1.8	4.5	7.9	2.0	4.9	8.1	1.9	4.7	8.0	1.7	4.5	8.0
$H_{S_\mathbf{v}}$	Kt	1.1	4.1	8.7	1.2	3.8	7.6	1.2	3.3	7.2	1.4	3.7	7.8
	Sk	1.2	5.0	9.8	1.1	5.2	10.0	1.1	4.7	9.6	1.0	4.7	9.2
	GH	1.4	5.0	9.2	1.7	4.5	8.1	1.6	4.1	7.6	1.7	4.5	8.4
H_I	Kt	2.5	5.4	9.0	1.9	4.4	7.5	1.9	4.1	7.1	2.0	4.8	7.6
	Sk	1.1	4.8	9.2	1.7	5.5	9.9	1.2	4.9	9.8	1.1	4.7	9.2
	GH	2.9	6.3	9.8	2.8	5.7	9.0	2.1	5.1	8.4	2.3	5.6	9.1
H_R	Kt	1.0	4.0	8.7	1.5	4.0	7.9	1.5	3.9	7.9	5.9	8.0	9.8
HK_f	Kt	1.5	4.0	7.9	1.5	3.7	7.5	1.5	3.9	7.8	1.5	4.0	7.5
	Sk	0.8	4.5	9.8	1.1	4.7	9.2	1.0	4.8	9.8	1.1	4.6	9.3
HK_{v_1}	Kt	1.3	3.5	7.1	1.2	3.7	7.4	1.2	3.3	7.1	1.2	3.4	7.0
	Sk	1.0	5.3	10.1	1.0	5.1	9.8	1.1	4.6	9.4	1.0	5.0	9.6
HK_{v_2}	Kt	1.4	4.2	8.0							1.4	3.5	7.4
	Sk	1.2	4.8	9.4							0.9	4.8	9.6
HK_{v_3}	Kt												
	Sk												

Notes: $\boldsymbol{\varepsilon}^* \sim N_K(\mathbf{0};\mathbf{I})$. Static FM (DGP 1): Trivariate static factor model with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = (1,1,1)'$ and $\boldsymbol{\gamma} = q^{-1}(1,1,1)'$ in (CITE); AR(1) + noise (DGP 2):AR(1) signal plus noise stationary UCARIMA model with $\alpha_x = .75$ and the signal-to-noise ratio, $\sigma_x^2/\sigma_\epsilon^2 = \sigma_f^2/[(1-\alpha_x^2)\sigma_v^2] = 2$; RW + noise (DGP 3): Random walk plus noise UCARIMA model with signal-to-noise ratio σ_f^2/σ_v^2 is equal to 2; and Cointegrated (DGP 4): Bivariate cointegrated, dynamic single factor model with $\alpha_x = .5$, $\alpha_{\epsilon_1} = .2$, $\alpha_{\epsilon_2} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ such that $q_1 = 2$ and $q_2 = .5$ in $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2 = [\sigma_f^2(1-\alpha_{\epsilon_i}^2)]/[(1-\alpha_x^2)\sigma_v^2]$. In the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta,\psi,\boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}} : \mathbf{v}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; $H_{S_\mathbf{f}} : \mathbf{f}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; and $H_I : \varepsilon_{it}^* \sim GH(\eta_i,\psi_i,\beta_i)$, for $i = 1,...,K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 2: Monte Carlo rejection rates (in %) under alternative hypotheses at the 5% significance level for the trivariate static factor model

		Student t				asymmetric Student t			
		H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I	H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I
H_J	Kt	98.5	21.3	83.3	76.0	99.5	54.9	86.8	98.9
	Sk								
	GH								
$H_{S_\mathbf{f}}$	Kt	66.9	51.6	5.1	53.0	91.8	81.2	13.7	81.8
	Sk	29.8	23.5	5.3	23.9	99.1	93.0	13.8	95.4
	GH	65.3	50.4	5.7	51.6	98.7	92.3	15.8	94.5
$H_{S_\mathbf{v}}$	Kt	93.6	4.0	91.7	65.2	94.8	4.5	92.7	97.1
	Sk	54.1	5.3	47.1	29.9	98.0	12.0	69.2	95.5
	GH	91.2	5.0	88.3	61.5	99.3	10.7	92.5	98.9
H_I	Kt	95.3	41.8	85.5	82.3	98.5	75.3	87.4	99.4
	Sk	57.6	17.2	43.1	42.2	99.2	86.9	52.0	99.9
	GH	94.8	40.7	84.5	81.2	99.7	87.3	86.9	99.9
H_R	Kt	98.8	23.1	84.9	78.1	99.6	56.6	88.3	99.1
HK_f	Kt	66.3	51.2	5.0	52.6	93.2	82.6	13.3	83.4
	Sk	29.9	23.7	5.3	24.0	99.2	93.1	13.9	95.4
HK_{v_1}	Kt	64.1	3.4	58.9	39.1	66.0	4.2	59.9	75.6
	Sk	27.9	4.6	26.6	19.2	40.6	4.8	28.4	75.3
HK_{v_2}	Kt	63.3	3.8	58.7	38.8	66.0	4.1	60.1	75.4
	Sk	28.2	4.6	26.6	18.9	41.1	5.3	28.5	75.5
HK_{v_3}	Kt								
	Sk								

Notes: Trivariate static factor model (DGP 1) with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c}$ =$(1,1,1)'$ and $\boldsymbol{\gamma} = q^{-1}(1,1,1)'$ in (CITE). Innovations for Student t and asymmetric Student t are $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8)$ and $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8,\mathbf{-1})$, respectively, with $\boldsymbol{\varepsilon}_t^{*\text{N}} \sim N_{K-R}(\mathbf{0};\mathbf{I})$ according to the definitions of H_J, $H_{S_\mathbf{v}}$, $H_{S_\mathbf{f}}$, and H_I. In the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta,\psi,\boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}}: \mathbf{v}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; $H_{S_\mathbf{f}}: \mathbf{f}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; and $H_I: \varepsilon_{it}^* \sim GH(\eta_i,\psi_i,\beta_i)$, for $i = 1,...,K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 3: Monte Carlo rejection rates (in %) under alternative hypotheses at the 5% significance level for the AR(1) signal plus noise stationary UCARIMA model

		Student t				asymmetric Student t			
		H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I	H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I
H_J	Kt	53.5	11.9	18.4	30.8	90.6	32.6	45.8	70.5
	Sk	25.3	9.9	12.8	17.0	97.2	43.8	56.0	85.5
	GH	51.6	13.6	19.5	30.7	98.1	52.2	64.8	89.5
$H_{S_\mathbf{f}}$	Kt	41.5	14.9	8.9	22.8	81.4	38.6	30.0	57.7
	Sk	20.0	10.4	7.9	12.9	92.9	48.6	22.3	73.9
	GH	40.9	16.2	9.9	23.3	92.3	49.6	26.8	73.3
$H_{S_\mathbf{v}}$	Kt	44.7	6.8	21.6	27.6	85.6	17.5	51.7	64.7
	Sk	21.0	6.4	13.3	14.1	95.0	23.1	62.0	86.2
	GH	44.2	7.4	22.3	27.7	95.1	19.9	63.8	80.8
H_I	Kt	48.0	22.4	13.2	33.0	87.3	51.6	35.0	73.2
	Sk	22.8	12.9	8.6	16.9	94.3	62.2	38.5	80.7
	GH	48.0	23.4	14.2	33.6	94.8	63.4	42.6	83.5
H_R	Kt	49.2	20.9	8.9	28.8	87.2	50.5	23.3	68.5
HK_f	Kt	41.3	14.8	8.8	22.6	81.4	38.6	22.9	57.7
	Sk	19.8	10.3	7.8	12.7	93.0	48.6	22.1	73.4
HK_v	Kt	44.5	6.7	21.5	27.6	85.6	17.3	51.6	64.7
	Sk	20.9	6.4	13.1	14.0	95.0	15.5	61.8	80.7

Notes: AR(1) signal plus noise stationary UCARIMA model (DGP 2) with $\alpha_x = .75$ and the signal-to-noise ratio, $\sigma_x^2/\sigma_\epsilon^2 = 2$. Innovations for Student t and asymmetric Student t are $\boldsymbol{\varepsilon}_t^{*\mathrm{GH}} \sim t(\mathbf{0};\mathbf{I};1/8)$ and $\boldsymbol{\varepsilon}_t^{*\mathrm{GH}} \sim t(\mathbf{0};\mathbf{I};1/8,\mathbf{-1})$, respectively, with $\boldsymbol{\varepsilon}_t^{*\mathrm{N}} \sim N_{K-R}(\mathbf{0};\mathbf{I})$ according to the definitions of H_J, $H_{S_\mathbf{v}}$, $H_{S_\mathbf{f}}$, and H_I. In the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta,\psi,\boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}}$: $\mathbf{v}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; $H_{S_\mathbf{f}}$: $\mathbf{f}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; and H_I : $\varepsilon_{it}^* \sim GH(\eta_i,\psi_i,\beta_i)$, for $i = 1,...,K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 4: Monte Carlo rejection rates (in %) under alternative hypotheses at the 5% significance level for the Random walk plus noise UCARIMA model

		Student t				asymmetric Student t			
		H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I	H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I
H_J	Kt	49.7	19.7	10.7	30.1	88.1	48.2	28.6	70.7
	Sk	24.0	13.1	8.6	16.6	95.0	63.0	36.2	78.9
	GH	47.9	21.2	11.8	30.4	95.5	63.5	38.6	82.4
$H_{S_\mathbf{f}}$	Kt	41.1	24.5	5.7	28.0	80.5	54.6	13.5	64.8
	Sk	19.9	13.8	5.8	14.9	88.7	68.4	8.7	68.9
	GH	40.9	24.8	6.6	28.2	89.8	68.1	14.4	74.9
$H_{S_\mathbf{v}}$	Kt	34.3	8.1	14.2	21.1	78.5	22.4	37.2	57.0
	Sk	14.2	6.2	9.2	10.7	72.5	7.8	42.8	44.8
	GH	33.9	8.9	14.6	21.1	83.9	21.7	46.0	62.2
H_I	Kt	48.0	22.4	13.2	33.0	87.3	51.5	35.0	73.2
	Sk	22.8	12.9	8.6	16.9	94.3	62.2	38.5	80.7
	GH	48.0	23.4	14.2	33.6	94.8	63.4	42.6	83.5
H_R	Kt	49.2	20.9	8.9	28.8	87.2	50.5	23.3	68.5
HK_f	Kt	41.0	24.4	5.6	28.0	80.4	54.5	13.4	64.7
	Sk	19.8	13.7	5.7	14.8	88.7	68.3	8.7	68.7
HK_v	Kt	34.2	8.0	14.2	20.9	78.4	22.3	37.2	56.8
	Sk	14.1	6.1	9.1	10.1	72.3	7.8	42.6	44.7

Notes: Random walk plus noise UCARIMA model (DGP 3) with signal-to-noise ratio σ_f^2/σ_v^2 is equal to 2. Innovations for Student t and asymmetric Student t are $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8)$ and $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8,\mathbf{-1})$, respectively, with $\boldsymbol{\varepsilon}_t^{*\text{N}} \sim N_{K-R}(\mathbf{0};\mathbf{I})$ according to the definitions of H_J, $H_{S_\mathbf{v}}$, $H_{S_\mathbf{f}}$, and H_I. In the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta,\psi,\boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}} : \mathbf{v}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; $H_{S_\mathbf{f}} : \mathbf{f}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; and $H_I : \varepsilon_{it}^* \sim GH(\eta_i,\psi_i,\beta_i)$, for $i = 1,...,K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 5: Monte Carlo rejection rates (in %) under alternative hypotheses at the 5% significance level for the bivariate cointegrated, dynamic single factor model

		Student t				asymmetric Student t			
		H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I	H_J	$H_{S_\mathbf{f}}$	$H_{S_\mathbf{v}}$	H_I
H_J	Kt	85.8	14.2	56.2	56.4	96.9	37.8	72.7	92.3
	Sk	41.8	10.8	26.7	25.8	98.8	54.4	71.7	97.1
	GH	81.6	15.4	52.4	52.6	99.4	54.7	80.9	98.0
$H_{S_\mathbf{f}}$	Kt	44.4	29.3	5.8	32.0	79.5	59.8	15.8	63.7
	Sk	20.6	16.0	5.3	16.0	87.5	74.9	9.3	74.4
	GH	43.6	29.3	6.2	31.6	89.4	74.1	16.0	16.1
$H_{S_\mathbf{v}}$	Kt	72.5	4.4	65.5	48.9	87.9	6.0	76.7	87.2
	Sk	33.7	5.0	30.5	23.4	92.3	6.6	77.5	92.4
	GH	69.5	5.5	62.8	46.8	95.7	7.2	85.8	95.3
H_I	Kt	76.9	24.4	56.9	62.2	94.0	54.4	72.1	54.4
	Sk	37.6	12.1	25.7	29.3	97.2	65.3	66.4	98.7
	GH	75.8	24.6	55.8	61.8	98.1	66.5	78.3	99.0
H_R	Kt	86.6	24.5	56.7	59.3	97.1	48.1	70.5	92.7
HK_f	Kt	45.1	29.8	5.9	32.1	81.6	61.3	16.3	65.6
	Sk	21.2	16.6	5.6	16.6	87.8	75.3	9.7	74.8
HK_{v_1}	Kt	40.2	4.0	35.0	27.3	55.9	5.4	45.0	62.3
	Sk	18.9	5.0	17.8	14.3	44.2	5.5	35.0	60.2
HK_{v_2}	Kt	57.8	4.4	49.7	42.0	75.6	4.9	61.5	77.0
	Sk	26.1	5.3	23.1	20.8	78.8	5.6	59.7	86.6

Notes: Bivariate cointegrated, dynamic single factor model (DGP 4) with $\alpha_x = .5$, $\alpha_{\epsilon_1} = .2$, $\alpha_{\epsilon_2} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ such that $q_1 = 2$ and $q_2 = .5$ with $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$. Innovations for Student t and asymmetric Student t are $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8)$ and $\boldsymbol{\varepsilon}_t^{*\text{GH}} \sim t(\mathbf{0};\mathbf{I};1/8,\mathbf{-1})$, respectively, with $\boldsymbol{\varepsilon}_t^{*\text{N}} \sim N_{K-R}(\mathbf{0};\mathbf{I})$ according to the definitions of H_J, $H_{S_\mathbf{v}}$, $H_{S_\mathbf{f}}$, and H_I. In the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta,\psi,\boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}} : \mathbf{v}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; $H_{S_\mathbf{f}} : \mathbf{f}_t^* \sim GH_R(\eta,\psi,\boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0};\mathbf{I})$; and $H_I : \varepsilon_{it}^* \sim GH(\eta_i,\psi_i,\beta_i)$, for $i = 1,...,K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 6: Parameter estimates and normality tests: Great Moderation without the Great Recession

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.357	0.054
δ	0.021	0.094
α_x	0.534	0.045
α_{ϵ_1}	-0.677	0.036
$\alpha_{\epsilon 2}$	0.941	0.040
σ_x^2	0.134	0.026
$\sigma_{\epsilon_1}^2$	0.010	0.004
$\sigma_{\epsilon_2}^2$	0.152	0.020

Panel B: Normality tests

		statistic	p-value			statistic	p-value
	Kt	2.683	0.101		Kt	1.827	0.177
H_J	Sk	3.684	0.298	H_R	Sk	1.516	0.469
	GH	6.367	0.173		GH	3.342	0.342
	Kt	0.663	0.416		Kt	0.007	0.934
$H_{S_{\mathbf{f}}}$	Sk	1.510	0.219	HK$_f$	Sk	0.025	0.874
	GH	2.173	0.337		GH	0.032	0.984
	Kt	6.342	0.012		Kt	12.138	0.000
$H_{S_{\mathbf{v}}}$	Sk	8.318	0.016	HK$_{v_1}$	Sk	0.000	0.991
	GH	14.660	0.002		GH	12.138	0.002
	Kt	34.300	0.000		Kt	0.006	0.938
H_I	Sk	2.099	0.552	HK$_{v_2}$	Sk	0.003	0.958
	GH	36.399	0.000		GH	0.008	0.996

Notes: Data: Quarterly GDP and GDI from 1984Q2 to 2007Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. In Panel B, the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta, \psi, \boldsymbol{\beta})$; similarly, $H_{S_{\mathbf{v}}} : \mathbf{v}_t^* \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; $H_{S_{\mathbf{f}}} : \mathbf{f}_t^* \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; and $H_I : \varepsilon_{it}^* \sim GH(\eta_i, \psi_i, \beta_i)$, for $i = 1, ..., K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 7: Parameter estimates and normality tests: Great moderation with the Great Recession

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.230	0.071
δ	0.005	0.068
α_x	0.643	0.086
α_{ϵ_1}	-0.383	0.267
$\alpha_{\epsilon 2}$	0.939	0.032
σ^2_x	0.168	0.033
$\sigma^2_{\epsilon_1}$	0.022	0.012
$\sigma^2_{\epsilon_2}$	0.149	0.024

Panel B: Normality tests

		statistic	p-value			statistic	p-value
H_J	Kt	20.117	0.000	H_R	Kt	21.529	0.000
	Sk	18.799	0.000		Sk	7.950	0.010
	GH	38.916	0.000		GH	29.474	0.000
$H_{S_\mathbf{f}}$	Kt	66.339	0.000	HK_f	Kt	37.149	0.000
	Sk	22.840	0.000		Sk	4.339	0.037
	GH	89.179	0.000		GH	41.489	0.000
$H_{S_\mathbf{v}}$	Kt	8.000	0.004	HK_{v_1}	Kt	0.850	0.356
	Sk	4.513	0.105		Sk	0.000	0.999
	GH	13.013	0.005		GH	0.850	0.654
H_I	Kt	79.941	0.000	HK_{v_2}	Kt	0.104	0.746
	Sk	22.920	0.000		Sk	0.000	0.997
	GH	102.861	0.000		GH	0.104	0.949

Notes: Data: Quarterly GDP and GDI from 1984Q2 to 2015Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. In Panel B, the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}^*_t \sim GH_K(\eta, \psi, \boldsymbol{\beta})$; similarly, $H_{S_\mathbf{v}} : \mathbf{v}^*_t \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{f}^*_t \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; $H_{S_\mathbf{f}} : \mathbf{f}^*_t \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{v}^*_t \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; and $H_I : \varepsilon^*_{it} \sim GH(\eta_i, \psi_i, \beta_i)$, for $i = 1, ..., K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Table 8: Parameter estimates and normality tests: A longer sample

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.376	0.063
δ	0.053	0.045
α_x	0.506	0.053
α_{ϵ_1}	0.910	0.033
$\alpha_{\epsilon 2}$	0.069	0.394
σ_x^2	0.538	0.059
$\sigma_{\epsilon_1}^2$	0.140	0.026
$\sigma_{\epsilon_2}^2$	0.034	0.019

Panel B: Normality tests

		statistic	p-value			statistic	p-value
	Kt	17.329	0.000		Kt	13.517	0.000
H_J	Sk	7.503	0.058	H_R	Sk	1.284	0.526
	GH	24.833	0.000		GH	14.801	0.002
	Kt	9.416	0.002		Kt	1.475	0.225
$H_{S_{\mathbf{f}}}$	Sk	1.296	0.255	HK$_f$	Sk	0.006	0.940
	GH	10.713	0.005		GH	1.481	0.477
	Kt	8.108	0.004		Kt	0.016	0.900
$H_{S_{\mathbf{v}}}$	Sk	7.724	0.021	HK$_{v_1}$	Sk	0.015	0.904
	GH	15.832	0.001		GH	0.031	0.985
	Kt	24.194	0.000		Kt	0.041	0.840
H_I	Sk	3.436	0.329	HK$_{v_1}$	Sk	0.000	0.984
	GH	27.630	0.000		GH	0.041	0.980

Notes: Data: Quarterly GDP and GDI from 1952Q1 to 2015Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. In Panel B, the first column H_J refers to normality score tests of Propositions 4 and 5 against $\boldsymbol{\varepsilon}_t^* \sim GH_K(\eta, \psi, \boldsymbol{\beta})$; similarly, $H_{S_{\mathbf{v}}} : \mathbf{v}_t^* \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{f}_t^* \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; $H_{S_{\mathbf{f}}} : \mathbf{f}_t^* \sim GH_R(\eta, \psi, \boldsymbol{\beta})$, with $\mathbf{v}_t^* \sim N_{K-R}(\mathbf{0}; \mathbf{I})$; and $H_I : \varepsilon_{it}^* \sim GH(\eta_i, \psi_i, \beta_i)$, for $i = 1, ..., K$. H_R denotes the tests of section 3.5.2. HK refers to the Harvey and Koopman (1992)'s tests of section 3.5.1. Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, respectively, while GH corresponds to the sum of the two.

Figure 1: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q2 to 2007Q2.

Figure 1a: Smoothed innovations for the underlying factor

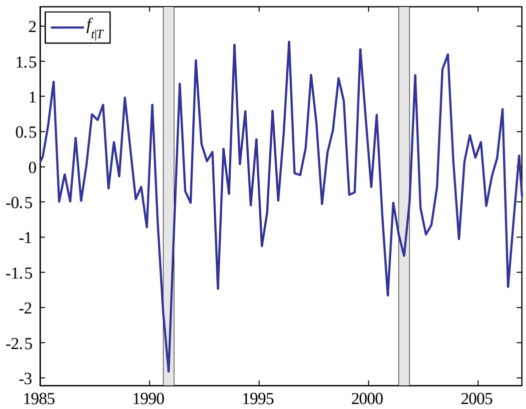


Figure 1b: Smoothed innovations for the measurement errors

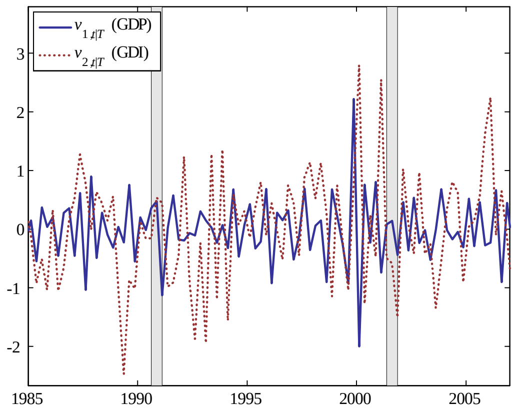


Figure 1c: Influence functions for for the underlying factor (kurtosis)

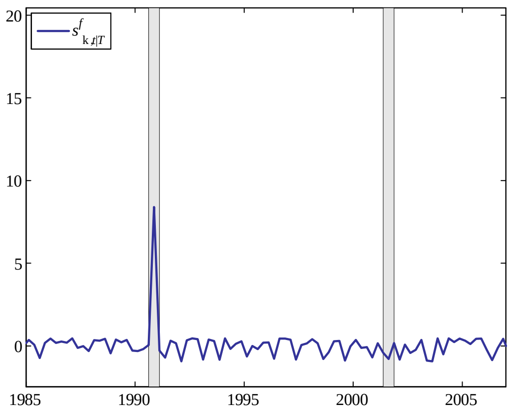


Figure 1d: Influence functions for for the measurement errors (kurtosis)

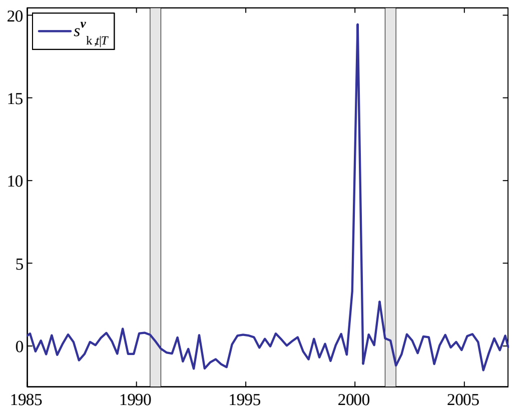


Figure 1e: Influence functions for for the underlying factor (skewness)

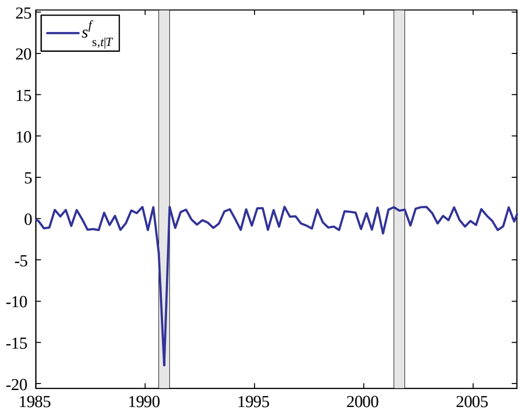
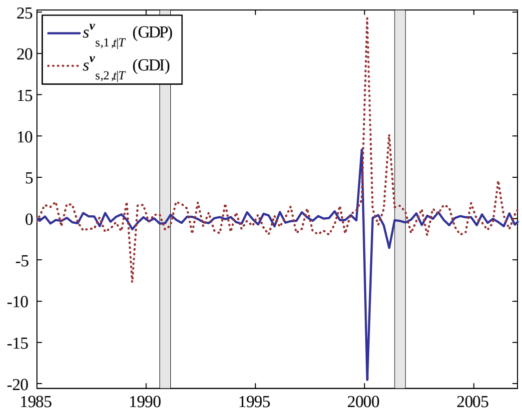


Figure 1f: Influence functions for for the measurement errors (skewness)



Notes: Data: Quarterly GDP and GDI from 1984Q2 to 2007Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. Shaded areas represent NBER recessions.

Figure 2: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q2 to 2015Q2.

Figure 2a: Smoothed innovations for the underlying factor

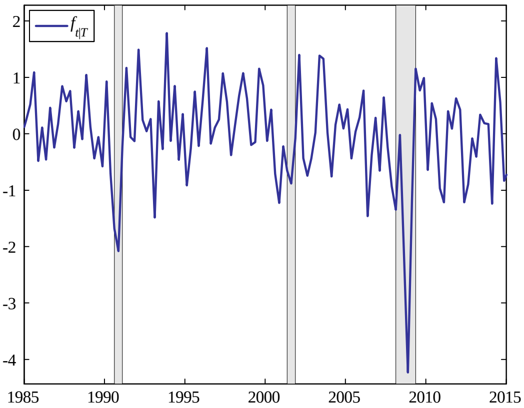


Figure 2b: Smoothed innovations for the measurement errors

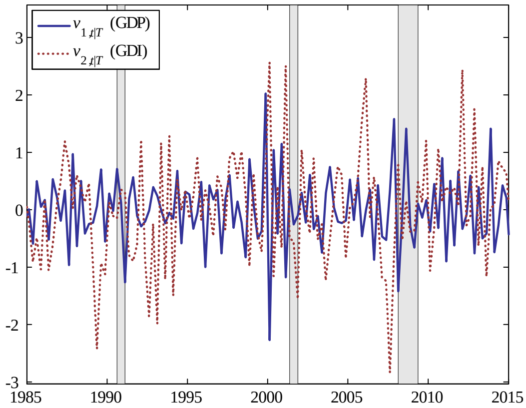


Figure 2c: Influence functions for for the underlying factor (kurtosis)

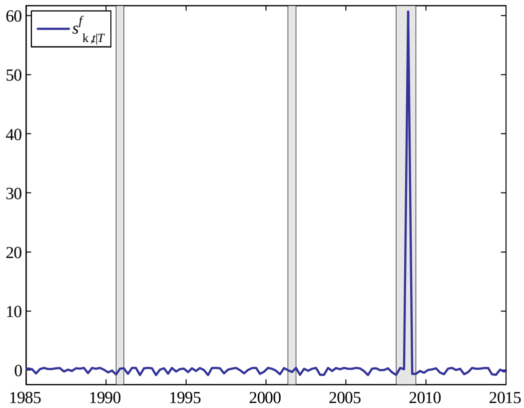


Figure 2d: Influence functions for for the measurement errors (kurtosis)

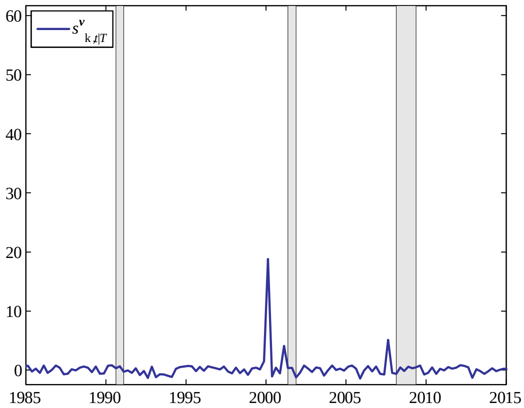


Figure 2e: Influence functions for for the underlying factor (skewness)

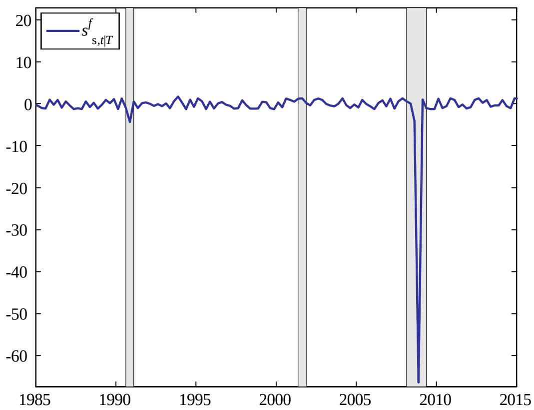
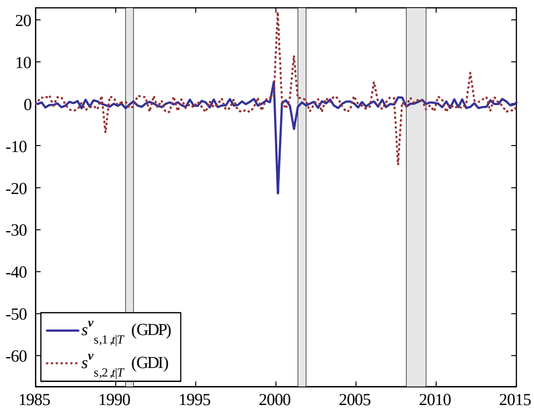


Figure 2f: Influence functions for for the measurement errors (skewness)



Notes: Data: Quarterly GDP and GDI from 1984Q2 to 2015Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. Shaded areas represent NBER recessions.

Figure 3: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1952Q1 to 2015Q2.

Figure 3a: Smoothed innovations for the underlying factor

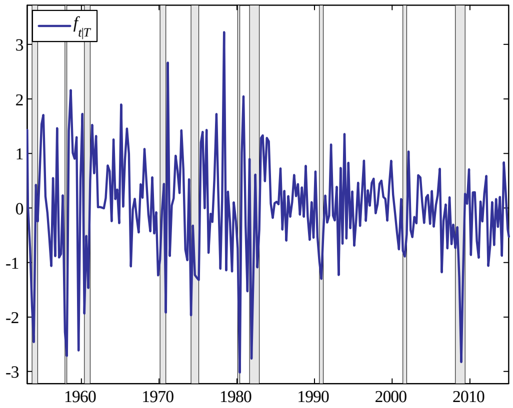


Figure 3b: Smoothed innovations for the measurement errors

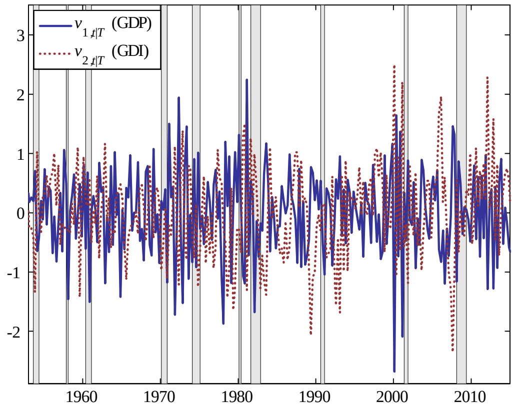


Figure 3c: Influence functions for for the underlying factor (kurtosis)

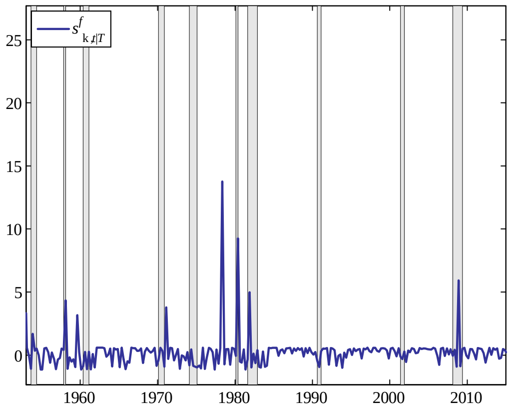


Figure 3d: Influence functions for for the measurement errors (kurtosis)

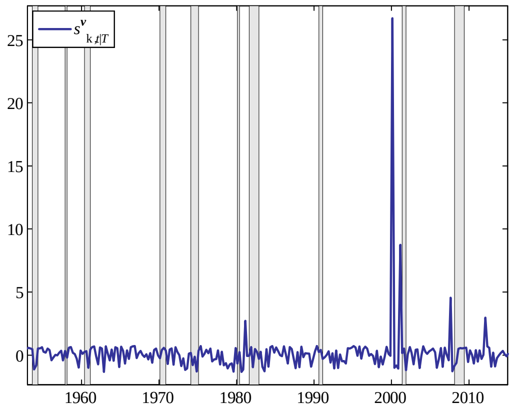


Figure 3e: Influence functions for for the underlying factor (skewness)

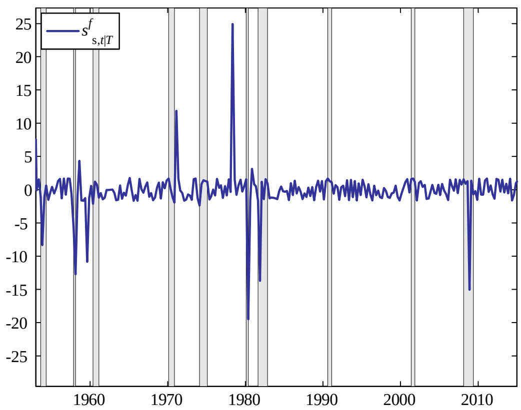
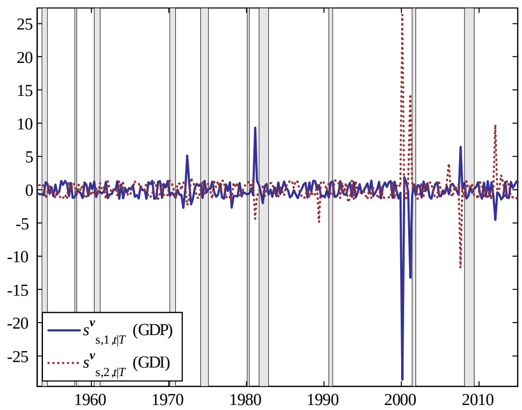


Figure 3f: Influence functions for for the measurement errors (skewness)



Notes: Data: Quarterly GDP and GDI from 1952Q1 to 2015Q2. Bivariate cointegrated, dynamic single factor model; see section 2.3.3 for parameter definitions. Shaded areas represent NBER recessions.