cemfi

working paper 1708

Normality Tests for Latent Variables

Tincho Almuzara
Dante Amengual
Enrique Sentana

February 2017

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 1708
February 2017

Normality Tests for Latent Variables

Abstract

We exploit the rationale behind the Expectation Maximization algorithm to derive simple to implement and interpret score tests of normality in the innovations to the latent variables in state space models against generalized hyperbolic alternatives, including symmetric and asymmetric Student ts. We decompose our tests into third and fourth moment components, and obtain one-sided likelihood ratio analogues, whose asymptotic distribution we provide. When we apply them to a cointegrated dynamic factor model which combines the expenditure and income versions of US aggregate real output to improve its measurement, we reject normality if the sample period extends beyond the Great Moderation.

JEL Codes: C32, C52, E01.

Keywords: Gross domestic product, gross domestic income, kurtosis, Kuhn-Tucker test, skewness, supremum test, Wiener-Kolmogorov-Kalman smoother.

Tincho Almuzara
CEMFI
tincho.almuzara@gmail.com

Dante Amengual
CEMFI
amengual@cemfi.es

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

We are grateful to Gabriele Fiorentini and Javier Mencía, as well as to seminar audiences at Boston University, the 2016 Barcelona GSE Summer Forum, the 69th ESEM (Geneva), the 2016 NBER-NSF Time Series Conference (New York), the 41st SAEe (Bilbao), the 2016 Encuentro SEU (Montevideo) and the 7th ICEEE (Messina) for helpful comments, discussions and suggestions. Of course, the usual caveat applies. 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. The list of empirical studies that make use of those models is vast. In this paper, we consider a particularly relevant application whereby we infer aggregate (real) production from its expenditure (GDP) and income (GDI) measures. In theory, both measures of output 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 Department of Commerce. However, the Great Recession 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 et al (2016)). 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 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 emphasized by Smith, Weale and Satchell (1998), though, dynamic considerations matter. There are at least two important reasons: (i) the expenditure and income measures should be cointegrated with the true GDP, and therefore between themselves, with cointegrating vector (1,-1); and (ii) the associated measurement errors should be stationary but they may well be serially correlated. For those reasons, a single factor model with unit loadings on an I(1) common factor and covariance stationary specific factors provides a rather natural way of capturing the dynamics of the observed series.

We will use this model in section 5, but in developing it, one particularly relevant decision we must make is the normality of the underlying variables, which implies the normality of the observed variables and justifies the use of the Kalman filter for inferring the true underlying output from its two measures. In contrast, if the innovations are not Gaussian, the Kalman filter only provides the best linear filter for the latent variable, which can be noticeably different from its conditional expectation. To illustrate this point, consider the simplest possible example in which a negatively skewed signal x is observed cloaked in some additive symmetric noise ϵ.

As can be seen in Figure 1, the linear projection can display important biases relative to the conditional expectation of x given the observed series $y = x + \epsilon$. Intuitively, the conditional expectation takes into account that the asymmetry in x implies that large negative/positive realizations of y are more/less likely to result from the signal, while the linear projection assigns a constant fraction of y to x regardless.

The remarkable increase in computing power has made possible the implementation of simulation-based estimation and filtering techniques for non-Gaussian dynamic latent variable models (see e.g. Johannes and Polson (2009)). However, the majority of practitioners continue to rely on the Kalman filter, which is far simpler to implement and explain. Undoubtedly, those practitioners would benefit from the existence of diagnostics that could tell them the extent to which normality of the latent variables is at odds with the data. Although there are many readily 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 computational reasons, the advantage of LM tests is that rejections provide a clear indication of the specific directions along which modelling efforts should focus.

Nowadays, the computational advantages of LM tests might seem irrelevant, but in our case they are of first-order importance because 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. We overcome this stumbling block by using the Expectation Maximization (EM) principle to obtain the scores of the parameters that characterize 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 neglected serial dependence in non-Gaussian static 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 implied by the model in those estimates. Unlike us, though, none of those authors justify their procedures by appealing to the likelihood principle or consider multivariate models.

For most practical purposes, departures from normality can be attributed to two different sources: excess kurtosis and skewness. Although our EM-based LM approach can be applied far more generally, we follow Mencía and Sentana (2012) in considering Generalized 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 they lead to easy to interpret moment tests that focus on third and fourth moments. In particular, they coincide with the moments underlying the Jarque and Bera (1980) test in the univariate case. At the same time, the number of moments that are effectively tested in multivariate contexts is proportional to the number of series involved, unlike tests against Hermite expansions of the multivariate normal density, which suffer from the curse of dimensionality (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, so we can treat them as if they were known.

The rest of the paper is organized as follows. Section 2 describes the econometric model, as well as the GH alternatives. We derive the normality tests in section 3, and discuss the results of our Monte Carlo experiments in section 4. Section 5 explores in detail the information about aggregate output in the GDP and GDI measures. Finally, we present our conclusions in section 6. 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

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

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

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

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 a K-dimensional vector of standardized 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

[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.

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 exogenous regressors or those in which some of the system matrices are deterministic functions of time or observable predertermined 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 (see e.g. section 2.3 of Fiorentini, Galesi and Sentana (2016) and the references therein).

2.2 Null and alternative hypotheses

In the next section we derive computationally simple tests of the null hypothesis that the structural innovations are Gaussian against the alternative that they follow a member of the *GH* 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. As we mentioned in the introduction, the main advantages of these *GH* alternatives is that they lead to easy to interpret moment tests that focus on third and fourth moments, but in such a way that the number of conditions which are effectively tested is proportional to the number of series involved.

Mencía and Sentana (2012) derive a standardized version of the *GH* distribution with zero mean and identity covariance matrix, which therefore depends exclusively on three shape parameters that we can set to zero under normality: $\boldsymbol{\beta}$, which introduces asymmetries, and η and ψ,

[2] Minor changes to the testing procedures we propose will render them applicable to those situations.

whose product $\tau = \eta\psi$ effectively controls excess kurtosis in the vicinity of the Gaussian null.

In many applications, 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. In our empirical application, for example, it matters whether the potential non-normality is a feature of the true GDP or its measurement errors. 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 $\boldsymbol{\varepsilon}_t^{*\mathrm{N}}$.[3]

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 $\mathbf{Y}_T$, 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

$$\begin{aligned}
&\mathbf{m}_{1,t|T}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta}),\\
&\mathbf{m}_{2,t|T}(\boldsymbol{\theta}) = vec[\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})'],\\
&\mathbf{m}_{3,t|T}(\boldsymbol{\theta}) = vec\{\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})]'\},\ \text{and}\\
&\mathbf{m}_{4,t|T}(\boldsymbol{\theta}) = vec\{[\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})][\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta})]'\}.
\end{aligned} \tag{5}$$

3.1 The score under Gaussianity

LM tests are usually obtained from the score associated to the (marginal) likelihood function of the observed variables, $f_T(\mathbf{Y}_T;\boldsymbol{\phi})$ say, evaluated under the Gaussian null. Unfortunately, the functional form of $f_T(\mathbf{Y}_T;\boldsymbol{\phi})$ is generally unknown under the alternative, and consequently the same is true of its score vector evaluated 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 $f_T(\mathbf{Y}_T, \boldsymbol{\Xi}_T;\boldsymbol{\phi})$ denote the (joint) likelihood function for both observed $\{\mathbf{y}_t\}$ and state $\{\boldsymbol{\xi}_t\}$ variables of model (1)–(2) for a sample of size T and $f_T(\boldsymbol{\Xi}_T|\mathbf{Y}_T;\boldsymbol{\phi})$ the conditional likelihood function of the latent variables given the observed ones. Since the Kullback inequality implies

[3] We might also envisage an alternative situation in which the elements of $\boldsymbol{\varepsilon}_t^*$ are cross-sectionally independent but non-Gaussian, but for the sake of conciseness, we relegate its analysis to appendix D.

that $E\left[\partial \ln f_T(\mathbf{\Xi}_T|\mathbf{Y}_T;\boldsymbol{\phi})/\partial\boldsymbol{\phi}|\mathbf{Y}_T;\boldsymbol{\phi}\right] = \mathbf{0}$, it follows that we can obtain $\partial \ln f_T(\mathbf{Y}_T;\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ as the expected value (given $\mathbf{Y}_T$ and $\boldsymbol{\phi}$) of the unobservable score corresponding to $f_T(\mathbf{\Xi}_T|\mathbf{Y}_T;\boldsymbol{\phi})$. Specifically,

$$\frac{\partial \ln f_T(\mathbf{Y}_T;\boldsymbol{\phi})}{\partial\boldsymbol{\phi}} = E\left[\left.\frac{\partial \ln f_T(\mathbf{Y}_T,\mathbf{\Xi}_T;\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right|\mathbf{Y}_T;\boldsymbol{\phi}\right]. \tag{6}$$

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

But we still face two additional difficulties. First, there are three different paths along which a symmetric GH distribution converges to a Gaussian distribution. Fortunately, Mencía and Sentana (2012) showed that the score of the relevant kurtosis parameter evaluated under the null of normality is proportional along those three paths to the score with respect to $\tau = \eta\psi$ evaluated at $\tau = 0$. Second, $\boldsymbol{\beta}$ vanishes from the log-likelihood function as $\tau \to 0$ (see again 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. To apply this idea to the LM test, we need:

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*

$$\begin{aligned}
\bar{s}_{\mathrm{GH},T}(\boldsymbol{\theta},\boldsymbol{\beta}) &= \frac{1}{T}\sum_{t=1}^{T}\left[s_{\mathrm{k},t|T}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})\right], \qquad (7)\\
s_{\mathrm{k},t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{k},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{k},t|T}(\boldsymbol{\theta}),\\
\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{s},t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{s},t|T}(\boldsymbol{\theta}),
\end{aligned}$$

where

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

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

$$\begin{aligned}
&b_{0,t|T}(\boldsymbol{\theta}) = c_0 + \{c_1 + c_2 tr[\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\}tr[\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})] + 2c_2 tr\{[\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]^2\},\\
&\mathbf{b}_{1,t|T}(\boldsymbol{\theta}) = [c_3 + tr(\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\mathbf{E}'_{RK} + 2\mathbf{E}'_{RK}\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}),\\
&\mathbf{b}_{2,t|T}(\boldsymbol{\theta}) = \{c_1 + 2c_2 tr[\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\}\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right)vec(\mathbf{I}_R) + 4c_2\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right)vec[\mathbf{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})],\\
&\mathbf{b}_{3,t|T}(\boldsymbol{\theta}) = \mathbf{E}'_{RK}\boldsymbol{\ell}_R\otimes\mathbf{E}'_{RK},\\
&\mathbf{b}_{4,t|T}(\boldsymbol{\theta}) = c_2\left(\mathbf{E}'_{RK}\otimes\mathbf{E}'_{RK}\right)\boldsymbol{\ell}_{R^2},
\end{aligned}$$

with $c_0 = R(R+2)/4$, $c_1 = -(R+2)/2$, $c_2 = 1/4$, $c_3 = -(R+2)$ and $\boldsymbol{\ell}_H$ 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 R skewness components, $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$.

We could then derive the associated test statistic, say $LM_T(\boldsymbol{\theta}, \boldsymbol{\beta})$, after finding the asymptotic variance for (7). But since it is often unclear what value of $\boldsymbol{\beta}$ to choose, we prefer a second approach, which consists in computing the LM test for all possible values of $\boldsymbol{\beta}$ and then taking the supremum over those parameter values. Either way, we first require the asymptotic covariance matrix 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 $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$ and $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta})$ 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 $\boldsymbol{\varepsilon}_t^*$ being *iid*. Consequently, the same will be true of $s_{\mathrm{k},t|T}(\boldsymbol{\theta}_0)$ and $\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}_0)$. 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 $\sqrt{T}\bar{s}_{\mathrm{k},t|T}(\boldsymbol{\theta}_0)$ and $\sqrt{T}\bar{\mathbf{s}}_{\mathrm{s},t|T}(\boldsymbol{\theta}_0)$ 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 $\boldsymbol{\varepsilon}_t^*$ 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 (5) 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 $\boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})$.[4]

We can then prove the following result:

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

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

and

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

respectively, where

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

denotes 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.

[4] 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.

3.2.1 Parameter uncertainty

In practice, we do not generally know $\boldsymbol{\theta}_0$. Therefore, we need to obtain the asymptotic covariance matrix of $\sqrt{T}\bar{s}_{\mathrm{k},t|T}(\hat{\boldsymbol{\theta}}_T)$ and $\sqrt{T}\bar{\mathbf{s}}_{\mathrm{s},t|T}(\hat{\boldsymbol{\theta}}_T)$, where $\hat{\boldsymbol{\theta}}_T$ is the Gaussian Maximum Likelihood estimator of $\boldsymbol{\theta}$, which is the efficient estimator under the null. Importantly, the following proposition shows that the sampling variability of the Gaussian ML 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,*

$$\lim_{T\to\infty} Cov[\sqrt{T}\bar{\mathbf{s}}_{\mathrm{MV},T}(\boldsymbol{\theta}_0), \sqrt{T}\bar{s}_{\mathrm{k},t|T}(\boldsymbol{\theta}_0)|\boldsymbol{\theta}_0] = \mathbf{0}$$

and

$$\lim_{T\to\infty} Cov[\sqrt{T}\bar{\mathbf{s}}_{\mathrm{MV},T}(\boldsymbol{\theta}_0), \sqrt{T}\bar{\mathbf{s}}_{\mathrm{s},t|T}(\boldsymbol{\theta}_0)|\boldsymbol{\theta}_0] = \mathbf{0}.$$

This result generalizes Proposition 3 in Fiorentini and Sentana (2007), which contains an analogous result for dynamic location-scale multivariate models with non-Gaussian innovations. 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 normal 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 Gaussian ML estimates $\hat{\boldsymbol{\phi}}_T = (\hat{\boldsymbol{\theta}}'_T, \mathbf{0}')'$.

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

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

which is asymptotically distributed as a χ_1^2 under the null.

The fact that $\eta = 0$ lies at the boundary of the admissible parameter space invalidates the usual distribution of the LR and Wald tests, which under the null will be a 50:50 mixture of χ_0^2 (=0 with probability 1) and χ_1^2. Although the distribution of the LM test statistic remains valid, 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 follow Fiorentini, Sentana and Calzolari (2003) in using the Kühn-Tucker (KT) multiplier test introduced by Gouriéroux, Holly and Monfort (1980) instead, which is equivalent in large samples to the LR and Wald tests. Thus, 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 strictly positive *and* the LM statistic exceeds the $100(1-2\varkappa)$ percentile of a χ_1^2 distribution. In this respect, it is important to mention that when there is a single restriction, as in our case, those one-sided tests would be asymptotically locally most powerful.

3.3.2 Multivariate normal 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

$$LM_T^{GH}(\hat{\boldsymbol{\theta}}_T, \boldsymbol{\beta}) = \frac{T}{\mathcal{C}_{\mathrm{k}|\infty}(\boldsymbol{\theta}_0) + \boldsymbol{\beta}'\mathcal{C}_{\mathrm{s}|\infty}(\boldsymbol{\theta}_0)\boldsymbol{\beta}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \left[s_{\mathrm{k},t|T}(\hat{\boldsymbol{\theta}}_T) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s},t|T}(\hat{\boldsymbol{\theta}}_T) \right] \right\}^2,$$

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

But we can maximize $LM_T^{GH}(\boldsymbol{\theta}, \boldsymbol{\beta})$ 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 (7) is equal to*

$$\sup_{\boldsymbol{\beta}} LM_T^{GH}(\hat{\boldsymbol{\theta}}_T, \boldsymbol{\beta}) = LM_T^{Student}(\hat{\boldsymbol{\theta}}_T) + T\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\mathrm{s},t|T}(\hat{\boldsymbol{\theta}}_T)\right]' \mathcal{C}_{\mathrm{s}|\infty}^{-1}(\boldsymbol{\theta}_0)\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\mathrm{s},t|T}(\hat{\boldsymbol{\theta}}_T)\right]$$

which is asymptotically distributed as a χ_{R+1}^2 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_R^2 and χ_{R+1}^2, whose p-values are the equally weighted average of those χ^2 p-values.

3.4 Two illustrative examples

In section 5 we will use our methods for improving GDP measurement. But since they apply far more generally, in this section we illustrate them with two popular textbook examples: a static factor model and the so-called local-level dynamic model.

3.4.1 Static factor models

We start by considering a single factor version of a traditional (i.e. static, conditionally homoskedastic and exact) factor model. Specifically,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\pi} + \mathbf{c}f_t + \mathbf{v}_t, \\ \left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array} \right) \Big| \mathcal{I}_{t-1}; \boldsymbol{\phi} &\sim iid\ D\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array} \right), \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array} \right), \boldsymbol{\eta} \right], \end{aligned} \tag{8}$$

where $\mathbf{y}_t$ is an $N \times 1$ vector of observable variables with constant conditional mean $\boldsymbol{\pi}$, f_t is an unobserved common factor, whose constant variance we have normalized 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 matrix of constant idiosyncratic variances, and $\boldsymbol{\theta} = (\mathbf{c}', \boldsymbol{\gamma}')'$, with $\boldsymbol{\gamma} = vecd(\boldsymbol{\Gamma})$.

We can easily express model (8) 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 $\boldsymbol{\varepsilon}_t^* = (f_t, \mathbf{v}_t^{*\prime})'$, where $\mathbf{v}_t^* = \boldsymbol{\Gamma}^{-1/2}\mathbf{v}_t$. Note that this specification trivially implies that

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

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.

Letting $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})$, we can show that

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

so 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 $N+1$. Hence, we will have that under the null,

$$\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta}) | \mathbf{Y}_T; \boldsymbol{\theta} \sim N[\boldsymbol{\varepsilon}^*_{t|t}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})],$$

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 look 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:

$$\left.\begin{array}{r} H_3(f_t) = f_t^3 - 3f_t, \\ H_4(f_t) = f_t^4 - 6f_t^2 + 3, \end{array}\right\} \tag{9}$$

which coincide with the third and fourth Hermite polynomials for f_t underlying the usual Jarque and Bera (1980) test.

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 (9) 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 $\mathbf{v}_t^*$.

3.4.2 The local-level model

Consider now the random walk plus noise model studied in Harvey and Koopman (1992),

$$\begin{aligned} y_t &= \pi + x_t + v_t \\ x_t &= x_{t-1} + f_t \\ \left.\begin{pmatrix} f_t \\ v_t \end{pmatrix}\right| \mathcal{I}_{t-1};\boldsymbol{\phi} &\sim iid\ D\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}, \boldsymbol{\eta}\right], \end{aligned}$$

where x_t is the "signal" component, v_t the orthogonal "non-signal" component, and $\boldsymbol{\theta}$ refers to the model parameters that characterize the autocovariance structure of the observed series.

Since there are only two shocks, we could 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.

For the sake of brevity, let us focus on 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 evaluated under the null will be given by

$$E\left[s_{\eta t}^{S_v}(\boldsymbol{\theta},\mathbf{0})|\mathbf{Y}_T\right]=\frac{1}{2}\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta})\right]^2-\sqrt{\frac{3}{2}}\left[1-\omega_{\mathrm{v},t|T}(\boldsymbol{\theta})\right]v_{t|T}^{*2}(\boldsymbol{\theta})+\frac{1}{2}\sqrt{\frac{1}{6}}v_{t|T}^{*4}(\boldsymbol{\theta}). \quad (10)$$

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 (10) is proportional to the fourth order Hermite polynomial of the standardized 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 yields 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 (10) are serially correlated.

3.5 Comparison with alternative approaches

3.5.1 Univariate tests applied to the smoothed innovations

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.

Their asymmetry test is based on the skewness coefficient

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

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 $sk_{\varepsilon_i^*}$ will be given by $\zeta_{\varepsilon_i^*}(\boldsymbol{\theta}_0, 3)$, where

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

provides the sum of powers of the autocorrelations, which are the autocorrelations of the powers of the original Gaussian series (see Lomnicki (1961)).

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

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

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

It is interesting to compare these tests to our LM tests based on Propositions 4 and 5. The procedures proposed by Harvey and Koopman (1992) can be regarded as moment tests of

$$\begin{aligned} E[f^{**3}_{t|T}(\boldsymbol{\theta})] &= 0, \qquad E[f^{**4}_{t|T}(\boldsymbol{\theta}) - 3] = 0, \\ E[v^{**3}_{t|T}(\boldsymbol{\theta})] &= 0, \qquad E[v^{**4}_{t|T}(\boldsymbol{\theta}) - 3] = 0, \end{aligned}$$

where $f^{**}_{t|T}(\boldsymbol{\theta})$ and $v^{**}_{t|T}(\boldsymbol{\theta})$ are standardized smoothed innovations. Thus, the main difference is that they look at third and fourth moments, while we use the log-likelihood scores, which are proportional to the third and fourth Hermite polynomials. The main advantage of the latter is that they are not affected by the sampling variability in $\hat{\boldsymbol{\theta}}_T$, as we have shown in Proposition 3. Nevertheless, Harvey and Koopman (1992) indicate that their tests are also asymptotically insensitive to parameter uncertainty when the standardization of $f^{**}_{t|T}(\boldsymbol{\theta})$ and $v^{**}_{t|T}(\boldsymbol{\theta})$ relies on sample moments (see also Bontemps and Meddahi (2005)).[5] In fact, we can show that their tests and ours are asymptotically equivalent under the null hypothesis in the local level model in section 3.4.2.

3.5.2 Reduced form tests

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 $M \geq N$.

[5] 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)).

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,

$$\mathbf{w}_t = \left[\mathbf{I}_N + \sum_{j=1}^{q_y} \mathbf{B}_j(\boldsymbol{\theta})L^j\right]^{-1} \left[\mathbf{I}_N - \sum_{j=1}^{q_y} \mathbf{A}_j(\boldsymbol{\theta})L^j\right] (\mathbf{y}_t - \boldsymbol{\pi}). \quad (11)$$

This relationship is the basis for the comparison of 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. 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.[6] 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 $\mathbf{Y}_{t-1}$ 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 standardized reduced form innovations

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

[6] Although we would expect it to be closer to a normal than $\boldsymbol{\varepsilon}_t^*$ because of the averaging implicit in (11).

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 expressions for the asymptotic covariance matrix of the sample average of those influence functions in terms of $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = V(\mathbf{w}_t)$, which typically coincides with the steady state value of $\boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\theta})$.

We can show that in the static factor models in section 3.4.1, 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 entire vector of 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. Nevertheless, our tests have the advantage over the reduced form ones that we can focus our attention on the common factors or the specific components separately.[7]

4 Monte Carlo simulations

In this section, we study the finite sample size and power properties of the testing procedures discussed above by means of several extensive Monte Carlo exercises. We do so in the context of three different models: the cointegrated single factor model we use in our empirical application in section 5, as well as the illustrative examples we discussed in sections 3.4.1 and 3.4.2.

4.1 Simulation and estimation details

We assess the power properties of our tests by generating non-Gaussian data in three alternative designs:

1. All structural innovations are jointly GH: $\boldsymbol{\varepsilon}^{*}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$ (alternative J);

[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. Straightforward calculations show that $LM^{Student}_T(\boldsymbol{\theta})$ is numerically identical irrespective of whether we are testing normality against Student t in both or just one of the innovations. The same applies to $\sup LM^{GH}_T(\boldsymbol{\theta})$.

2. The distribution of the innovations to the signal component is GH while the idiosyncratic shocks are Gaussian: $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ (alternative S_f);

3. The joint distribution of the innovations to the idiosyncratic variables is GH while the common component is Gaussian: $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ (alternative $S_{\mathbf{v}}$).

We consider two examples of GH distributions: a symmetric Student t with 8 degrees of freedom and an asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\ell_{K\times 1}$. Thus, we end up with a total of seven different specifications for $\boldsymbol{\varepsilon}_t^*$, including the Gaussian null. For each distributional assumption, we generate 10,000 samples of size T observations each exploiting the location-scale mixture of normal representation of the GH distribution (see Mencía and Sentana (2012) for details).

We use standard Matlab routines for estimation. In the case of the local-level model, we rely on its Ima$(1, 1)$ reduced form representation to improve the computational efficiency of the algorithm. Finally, we compute the asymptotic variances of the test statistics by truncating the infinite sums in Proposition 2 when the additional terms lead to increments lower than 10^{-5}.

Given that in all three models we observe a "pile-up" problem whereby the fraction of negative values of the average kurtosis scores exceeds 50% under the null, we employ a parametric bootstrap procedure based on 10,000 simulated samples. In this way, we can automatically compute size-adjusted rejection rates, as forcefully argued by Horowitz and Savin (2000). Importantly, our bootstrap procedure does not exploit the asymptotic orthogonality of the scores between mean and variance parameters on the one hand and shape parameters on the other in Proposition 3. On the contrary, it explicitly takes into account the sensitivity of the critical values to the estimated values of $\boldsymbol{\theta}$ in order not to rule out higher order refinements (see Appendix D.1 in Amengual and Sentana (2015) for details).

4.2 Small sample properties

4.2.1 Cointegrated dynamic factor model

We simulate data from the model (12) that we use in our empirical application, with $\alpha_x = .5$, $\alpha_{\epsilon_E} = .2$, $\alpha_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2 = [\sigma_f^2(1 - \alpha_{\epsilon_i}^2)]/[(1 - \alpha_x^2)\sigma_{v_i}^2]$ represents the signal-to-noise ratio for y_{it} for $i = E, I$.

Panels A of Tables 1 and 2 report rejection rates under the null at the 1%, 5% and 10% levels for $T = 100$ and $T = 250$, respectively, which roughly correspond to the sample sizes in our empirical application in section 5. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 5 corresponding to the J, S_f and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in section 3.5.2. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two. The results make clear that the parametric bootstrap works remarkably well for both sample sizes.[8]

Panels B of the same tables report the rejection rates at the 5% level of the tests under each of different alternative hypotheses that we consider. As expected, the most powerful test for any given alternative is typically the score test we have designed against that particular alternative. In that regard, we find that while the reduced form tests have non-trivial power, especially under alternative J, they are clearly dominated by the tests aimed at the structural innovations.

4.2.2 Static factor model

Table 3 shows the analogous results for a trivariate version of the static factor model (8) for $T = 250$ with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = (1, 1, 1)'$ and $\boldsymbol{\gamma} = q^{-1}(1, 1, 1)'$, where q reflects the signal-to-noise ratio, which we set to 2. We omit the rows corresponding to the Red test because we argued in section 3.5.2 that tests of the null hypothesis that the entire vector of reduced form innovations $\mathbf{w}_t^*$ is Gaussian against the alternative hypothesis that it follows a GH distribution are numerically identical to the analogous tests for the latent variables $\boldsymbol{\varepsilon}_t^*$.

Once again, the parametric bootstrap rejection rates are reliable. Similarly, the kurtosis component of the test designed for a specific alternative is the most powerful when the alternative is symmetric Student t, while the corresponding supremum tests present the highest rejection rates when the alternative is asymmetric.[9] Interestingly, tests designed against alternatives S_f and $S_{\mathbf{v}}$ have close to trivial power when the true DGP is a Student t for $S_{\mathbf{v}}$ and S_f, respectively.

4.2.3 Local level model

Table 4 contains the results for samples of size $T = 250$ of the local-level model in section 3.4.2 in which the signal-to-noise ratio $q = \sigma_f^2/\sigma_v^2$ is set to 2, as in Harvey and Koopman (1992). For comparison purposes, we also include the original Harvey and Koopman (1992) tests. Our results confirm the asymptotic equivalence between their tests and the less powerful two-sided

[8] Given the number of Monte Carlo replications, the 95% asymptotic confidence intervals for the rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels.

[9] The only exception is $H_{S_{\mathbf{v}}}$ when the DGP is such that the idiosyncratic shocks follow an asymmetric Student t, in which case the kurtosis component has more power than the corresponding supremum test.

versions of ours (not reported). More generally, we by and large reach the same conclusions for size and power as in the other two examples.

5 Inferring real output from GDP and GDI

Figure 2a contains the temporal evolution of the US quarterly (log) GDP and GDI series between 1984Q3 and 2015Q2, with shaded areas indicating NBER recessions. Although the two series differ, their (1,-1) cointegration relationship is evident. In turn, Figure 2b shows that their first differences are also highly correlated, but with a rich dynamic bivariate structure. Finally, Figure 2c makes clear that the statistical discrepancy is a persistent but stationary series whose movements are unrelated to the business cycle. For those reasons, the cointegrated single factor model with covariance stationary specific factors that we mentioned in the introduction is a natural candidate for capturing the dynamics of the two output series. Specifically, if y_{Et} and y_{It} denote (log) GDP and GDI, respectively, the model that we consider is

$$\begin{aligned}
\begin{pmatrix} y_{Et} \\ y_{It} \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} x_t + \begin{pmatrix} \epsilon_{Et} \\ \epsilon_{It} \end{pmatrix} \qquad (12) \\
(1-\alpha_x L)(\Delta x_t - \mu) &= f_t \\
(1-\alpha_{\epsilon_E} L)(\epsilon_{Et} - \delta/2) &= v_{Et} \\
(1-\alpha_{\epsilon_I} L)(\epsilon_{It} + \delta/2) &= v_{It} \\
\left.\begin{pmatrix} f_t \\ v_{Et} \\ v_{It} \end{pmatrix}\right| \mathcal{I}_{t-1}; \boldsymbol{\phi} &\sim iid\ D\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & 0 & 0 \\ 0 & \sigma_{v_E}^2 & 0 \\ 0 & 0 & \sigma_{v_I}^2 \end{pmatrix}, \boldsymbol{\eta}\right],
\end{aligned}$$

where x_t is the "true GDP" common factor, whose rate of growth follows an AR(1) process with mean μ, autoregressive coefficient α_x and innovation variance σ_f^2, while ϵ_{Et} and ϵ_{It} are the measurement errors in the (log) expenditure and income measures, respectively, which follow covariance stationary AR(1) processes with unconditional means $\pm\delta/2$, autoregressive coefficients α_{ϵ_E} and α_{ϵ_I}, and innovation variances $\sigma_{v_E}^2$ and $\sigma_{v_I}^2$.[10] Importantly, our 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.[11]

We initially estimate the model using data from 1984Q3 to 2007Q2. We chose the final date to exclude the Great Recession from the sample. As for the start date, it marks the

[10] Our specification of the serial correlation structure of the latent series follows from the empirical analysis in earlier versions of Fiorentini and Sentana (2016), who found evidence in favour of AR(1) processes for both the first difference of the common factor and the levels of the measurement errors.

[11] For identification purposes, though, we assume without loss of generality that the magnitude of those biases is the same for the two output series.

beginning of the so-called Great Moderation, as in Nalewaik (2010). We estimate the model in the time domain on the basis of the bivariate Gaussian likelihood of the stationary transformation $\Delta y_{Et} + \Delta y_{It}$ and $y_{Et} - y_{It}$, systematically exploring its surface to make sure that we have found the global maximum. Panel A of Table 5 presents the estimates of the model parameters and their corresponding standard errors obtained from the asymptotic information matrix, which we compute using its frequency domain closed-form expression. As expected, we find that the growth rate of the "true" aggregate real output series is reasonably persistent. Our estimates also suggest that GDP provides a better measure of output than GDI, in the sense that GDP measurement errors have both a smaller autoregressive coefficient –in absolute value– and a smaller variance parameter. Indeed, the negative serial correlation coefficient for the GDP measurement error implies a tendency to compensate prior measurement errors, while the highly persistent GDI measurement error indicates that the difference between the growth rates of GDI and the true output measure are close to white noise.

In turn, the normality tests reported in Panel B of Table 5 suggest that the soothing effects of the so–called Great Moderation propagated beyond second moments because the normality of the innovations to the underlying GDP growth rates is not rejected at conventional levels. In contrast, we clearly reject the null of Gaussian innovations in the measurement errors. In fact, we reject not only when we use the joint test but also when we look at the skewness and kurtosis components separately. In contrast, the bivariate normality test of the reduced form innovations fails to reject its null hypothesis, which confirms the power advantages of looking at the structural innovations documented in section 4.

To gain some further insight, in Figure 3 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). Panels 3d and 3f indicate that an unusual measurement issue in both series around the first quarter of 2000 leads to the rejection of the Gaussian null for the measurement errors.

In Table 6 we present analogous results for a slightly larger sample that includes the Great Recession (1984Q3-2015Q2). As can be seen from Panel A, there are no dramatic changes in the parameter estimates, except perhaps for a higher persistence in the common factor, whose innovations have an unsurprisingly larger variance too. Nevertheless, the smoothed series are almost identical over the overlapping period. Figure 4 presents the evolution of the two output measures and our smoothed estimate in the period surrounding the Great Recession. As can be seen, GDP kept increasing over the entire 2007 while GDI began to show early warning signs of stagnation one year before. In the fourth quarter of 2008, though, both series experimented a dramatic drop, with GDI recovering slightly earlier than GDP. Our estimate tends to closely follow the GDP series, but taking into account the differing behavior of GDI around the turning points.

The large fall in output experienced in 2008Q4 implies that we also reject the normality of the common factor over this extended period. In that regard, we would like to emphasize 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 5, which is entirely analogous to Figure 3 but including the Great Recession, confirms that 2008Q4 has a huge impact on the skewness and kurtosis scores of the common factor, resulting in a strong rejection of the null.

Nevertheless, if we take a longer historical perspective, and start our sample soon after the Treasury - Federal Reserve Accord whereby the Fed stopped its wartime pegging of interest rates, the Great Recession is no longer an isolated outlier. There are several other periods, including the turbulences in the late 70's, early 80's, in which the normality of the "true GDP" innovations is clearly rejected (see appendix E for details).

6 Conclusions

We exploit the rationale behind the Expectation Maximization 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 Generalized Hyperbolic alternatives, which include the symmetric and asymmetric Student t together with many other popular distributions. 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, explicitly comparing them to alternative tests. For the three models that we consider, our results indicate that the fraction of negative values of the average kurtosis scores exceeds 50% under the null. For that reason, we employ a parametric bootstrap procedure, which improves their reliability under the null. In terms of power, we find that the most powerful test for any given alternative is very often the score test we have designed against that particular alternative. We also find that while the tests that are based on the reduced form innovations have non-trivial power, they are clearly dominated by our proposed tests, which aim at the structural innovations.

When we apply our tests to a cointegrated dynamic factor model which combines the expenditure and income versions of US aggregate real output to improve its measurement, we reject normality of the innovations to the true GDP if the sample span extends beyond the Great Moderation (1984Q3-2007Q2). In contrast, the GDP/GDI measurement errors seem to be non-normal regardless of the period. Therefore, our results suggest that it is probably worth considering simulation-based non-linear procedures to improve over the Kalman filter.

From a methodological point of view, our EM-based approach can be successfully used in cross-sectional contexts too. In particular, it is straightforward to employ it for proving that many of the diagnostics suggested by Pagan and Vella (1989) for Tobit models do indeed coincide with the LM tests against specific alternatives in Chesher and Irish (1987) and Gouriéroux et al (1987). Analyzing other non-Gaussian contexts constitutes another interesting avenue for future research.

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.

Aruoba, S. B., F.X. Diebold, J. Nalewaik, F. Schorfheide and D. Song (2016): "Improving GDP measurement: a measurement-error perspective", *Journal of Econometrics* 191, 384–397.

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

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

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

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

Chesher, A. and M. Irish (1987): "Residual analysis in the grouped data and censored normal linear model", *Journal of Econometrics* 34, 33–62.

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, 1–38.

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

Fiorentini G., A. Galesi, A. and E. Sentana (2016): "A spectral EM algorithm for dynamic factor models", Bank of Spain Working Paper 1619, forthcoming in the *Journal of Econometrics*.

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, 532–546.

Fiorentini G., E. Sentana and G. Calzolari (2004): "On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models", *Economics Letters* 83, 307-312.

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

Fiorentini G. and E. Sentana (2015): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved components and time series econometrics*, 118-189, Oxford University Press.

Fiorentini G. and E. Sentana (2016): "Dynamic specification test for dynamic factor models", mimeo, CEMFI.

Greenaway-McGrevy, R. (2011): "Is GDP or GDI a better measure of output? A statistical approach", Bureau of Economic Analysis WP 2011–08.

Gouriéroux 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.

Gouriéroux, C., A. Monfort, E. Renault and A. Trognon (1987): "Generalized residuals", *Journal of Econometrics* 34, 5–32.

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, 377–389.

Horowitz, J. and N.E. Savin (2000): "Empirically relevant critical values for hypothesis tests: a bootstrap approach", *Journal of Econometrics* 95, 375–389.

Jarque, C. M. and A. Bera (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economics Letters* 6, 255–259.

Johannes, M. and Polson, N. (2009): "Particle filtering " in T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds.) *Handbook of financial time series,* 1015-1029, Springer-Verlag.

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, 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, 193–216.

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

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

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

Mencía J. (2012): "Testing nonlinear dependence in the hedge fund industry", *Journal of Financial Econometrics* 10, 545–587.

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

Nalewaik, Jeremy J. (2010): "The income- and expenditure-side measures of output growth", *Brookings Papers on Economic Activity* 1, 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, 385–402.

Pagan A. and F. Vella (1989): "Diagnostic tests for models based on individual data: a survey", *Journal of Applied Econometrics* 4, S29–S59.

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

Ruud, P. (1991): "Extensions of estimation methods using the EM algorithm", *Journal of Econometrics* 49, 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, 109–134.

Stone, R., D.G. Champernowne and J.E. Meade (1942): "The precision of national income estimates", *Review of Economic Studies* 9, 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, 385–400.

Weale, M. (1992): "Estimation of data measured with error and subject to linear restrictions", *Journal of Applied Econometrics* 7, 167–174.

Table 1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the bivariate cointegrated, dynamic single factor model ($T = 100$)

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	1.15	4.72	9.43	55.73	6.72	44.09	71.44	12.64	55.04
	Sk	1.00	4.92	10.30	31.77	6.79	25.05	68.09	17.31	50.62
	GH	1.02	4.67	9.79	48.13	6.88	37.11	74.04	16.26	57.02
$H_{S_{\mathbf{f}}}$	Kt	0.94	4.71	9.60	19.54	13.83	6.70	39.00	26.72	13.38
	Sk	0.91	4.69	9.79	13.03	10.07	6.11	33.83	29.56	10.26
	GH	0.95	4.69	9.65	18.22	12.90	6.40	39.76	30.13	13.08
$H_{S_{\mathbf{v}}}$	Kt	1.08	4.75	9.70	48.35	4.84	46.40	58.30	5.02	55.61
	Sk	1.09	4.87	9.92	27.60	5.29	27.15	51.41	6.30	54.84
	GH	1.04	4.83	9.94	42.96	5.14	41.71	61.15	5.74	60.98
Red	Kt	1.04	4.76	9.58	53.15	7.71	37.89	70.70	14.98	48.17
	Sk	0.88	4.61	8.91	24.33	5.02	21.65	31.23	5.30	23.59
	GH	0.99	4.45	9.02	47.45	6.79	34.36	65.45	12.37	44.22

Notes: Results based on 10,000 samples of size $T = 100$ from model (12) with $\alpha_x = .5$, $\alpha_{\epsilon_E} = .2$, $\alpha_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$ represents the signal-to-noise ratio for y_{it} for $i = E, I$. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t^* \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$, respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 5 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in section 3.5.2. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_{R\times 1}$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table 2: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the bivariate cointegrated, dynamic single factor model ($T = 250$)

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%)					
					Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	0.83	4.67	9.72	88.54	9.89	76.00	96.80	23.30	86.98
	Sk	1.02	5.33	10.19	42.42	8.77	33.85	95.50	36.18	82.65
	GH	0.98	4.99	9.85	80.82	9.73	66.07	98.55	34.51	90.56
$H_{S_{\mathbf{f}}}$	Kt	1.07	4.81	9.79	34.44	22.74	8.27	64.40	48.53	22.74
	Sk	1.11	5.25	10.04	17.07	12.33	6.45	55.84	58.49	16.27
	GH	1.09	5.08	10.09	31.41	20.69	7.86	67.19	59.01	22.76
$H_{S_{\mathbf{v}}}$	Kt	0.86	4.78	9.78	81.86	5.60	79.33	91.86	6.87	88.03
	Sk	1.15	5.21	10.15	35.49	6.07	35.22	83.47	8.32	86.65
	GH	1.03	4.89	9.83	74.06	5.83	72.00	93.88	7.99	92.91
Red	Kt	0.93	4.68	9.61	85.85	11.43	66.66	96.22	27.25	80.75
	Sk	1.22	5.15	10.72	31.06	5.41	27.85	41.49	6.24	31.54
	GH	0.98	4.71	9.96	80.97	9.57	60.67	94.33	23.22	76.20

Notes: Results based on 10,000 samples of size $T = 250$ from model (12) with $\alpha_x = .5$, $\alpha_{\epsilon_E} = .2$, $\alpha_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$ represents the signal-to-noise ratio for y_{it} for $i = E, I$. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t^* \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$, respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 5 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in section 3.5.2. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_{R\times 1}$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table 3: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the trivariate static factor model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%)					
					Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	1.02	4.83	9.68	99.33	30.74	89.94	99.76	63.13	92.36
	Sk	1.20	4.53	9.31	59.68	14.34	41.52	99.86	77.97	60.28
	GH	1.15	4.62	9.18	97.83	24.31	81.83	99.97	78.73	87.86
$H_{S_{\mathbf{f}}}$	Kt	1.01	5.04	9.54	71.78	57.02	6.75	93.12	83.90	15.60
	Sk	0.74	4.35	10.02	29.12	23.08	4.94	99.09	92.77	13.25
	GH	1.10	4.40	9.49	65.51	50.65	5.72	98.70	92.80	15.73
$H_{S_{\mathbf{v}}}$	Kt	1.01	4.60	9.87	96.00	4.79	94.63	96.87	5.83	95.68
	Sk	1.07	4.91	10.03	54.03	5.28	47.02	97.97	11.97	69.17
	GH	0.95	5.31	9.99	91.63	5.10	88.99	99.37	10.90	92.89

Notes: Results based on 10,000 samples of size $T = 250$ from a trivariate version of the static factor model (8) with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = (1,1,1)'$ and $\boldsymbol{\gamma} = q^{-1}(1,1,1)'$, where q reflects the signal-to-noise ratio, which we set to 2. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t^* \sim GH(\eta,\psi,\boldsymbol{\beta})$, $f_t \sim GH(\eta,\psi,\beta)$, $\mathbf{v}_t \sim N(\mathbf{0},\mathbf{I}_N)$ and $\mathbf{v}_t \sim GH(\eta,\psi,\boldsymbol{\beta})$, $f_t \sim N(0,1)$, respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 5 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_{R\times 1}$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table 4: Monte Carlo rejection rates (in %) under the null and alternative hypotheses for the local-level model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	S_v	J	S_f	S_v
H_J	Kt	1.15	5.15	10.07	56.63	25.12	13.82	90.53	53.15	33.40
	Sk	1.06	5.20	10.26	24.27	13.33	8.87	95.14	63.39	36.51
	GH	1.19	5.02	10.33	48.81	21.71	12.22	95.64	64.17	39.28
H_{S_f}	Kt	1.14	5.23	10.69	47.35	29.64	7.80	83.55	59.20	16.38
	Sk	1.03	4.82	10.22	19.81	13.77	5.81	88.63	68.30	8.68
	GH	1.22	5.17	9.94	42.65	26.21	7.24	90.45	69.28	15.45
H_{S_v}	Kt	1.03	4.72	9.93	40.70	11.13	18.34	82.43	26.60	41.91
	Sk	1.05	4.89	9.92	14.67	6.47	9.49	72.92	8.18	43.37
	GH	1.04	4.70	9.84	35.77	9.94	15.97	84.85	22.82	47.82
Red	Kt	1.08	5.37	10.30	55.48	25.49	11.25	89.72	54.63	27.29
	Sk	1.17	4.99	10.04	22.31	13.11	6.83	94.90	63.05	16.58
	GH	1.20	5.22	10.09	49.66	22.93	10.34	95.58	64.10	26.14
HK$_f$	Kt	1.14	5.49	10.68	43.99	26.97	7.33	82.00	56.92	15.26
	Sk	1.04	4.83	10.19	19.82	13.75	5.79	88.67	68.30	8.73
	GH	1.22	5.23	9.96	41.95	25.67	7.06	90.29	69.15	15.06
HK$_v$	Kt	1.03	4.41	10.33	36.81	9.64	16.18	80.21	24.26	39.17
	Sk	1.05	4.89	9.99	14.66	6.51	9.49	72.91	8.19	43.38
	GH	1.05	4.81	9.98	35.25	9.70	15.51	84.54	22.41	47.29

Notes: Results based on 10,000 samples of size $T = 250$ from the local-level model discussed in section 3.4.2 in which the signal-to-noise ratio $q = \sigma_f^2/\sigma_v^2$ is set to 2. The column labels J, S_f, S_{v} refer to the alternative $\boldsymbol{\varepsilon}_t^* \sim GH(\eta,\psi,\boldsymbol{\beta})$, $f_t \sim GH(\eta,\psi,\beta)$, $v_t \sim N(0,1)$ and $v_t \sim GH(\eta,\psi,\beta)$, $f_t \sim N(0,1)$, respectively. The row labels H_J, H_{S_f}, and H_{S_v} refer to the score tests in Propositions 4 and 5 corresponding to the J, S_f, and S_v alternative hypotheses, Red denotes the reduced form tests discussed in section 3.5.2, while HK denotes the original Harvey and Koopman (1992) tests discussed in section 3.5.1. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_{R\times 1}$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table 5: Parameter estimates and normality tests during Great Moderation

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.765	0.330
δ	0.181	0.040
α_x	0.536	0.105
α_{ϵ_E}	-0.672	0.152
α_{ϵ_I}	0.940	0.036
σ_f^2	0.135	0.027
$\sigma_{v_E}^2$	0.010	0.005
$\sigma_{v_I}^2$	0.153	0.025

Panel B: Normality tests

		statistic	p-value
$H_{S_\mathbf{f}}$	Kt	0.646	0.211
	Sk	1.540	0.215
	GH	2.186	0.237
$H_{S_\mathbf{v}}$	Kt	5.901	0.008
	Sk	7.914	0.019
	GH	13.815	0.002
H_R	Kt	1.585	0.104
	Sk	1.478	0.478
	GH	3.063	0.299

Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2007Q2. Model: Bivariate cointegrated, dynamic single factor model (12); see section 5 for parameter definitions. In Panel A, estimates are Gaussian ML of the bivariate Gaussian likelihood of the stationary transformation $\Delta y_{Et} + \Delta y_{It}$ and $y_{Et} - y_{It}$ in the time domain. Standard errors are obtained from the asymptotic information matrix, which is computed using its frequency domain closed-form expression. In Panel B, the row labels H_{S_f} and $H_{S_\mathbf{v}}$ refer to the score tests in Propositions 4 and 5 corresponding to the S_f and $S_\mathbf{v}$ alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in section 3.5.2. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table 6: Parameter estimates and normality tests during the Great moderation and the Great Recession

Panel A: ML estimates

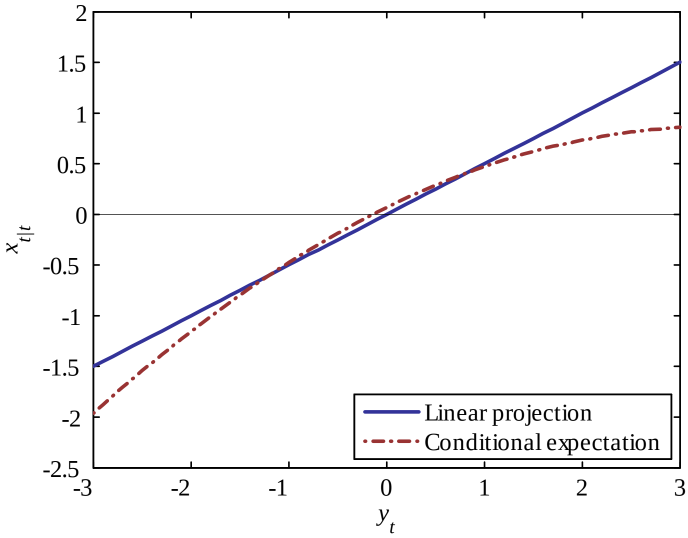
Param.	estimate	std. err.
μ	0.642	0.196
δ	0.033	0.036
α_x	0.643	0.080
α_{ϵ_E}	-0.384	0.204
α_{ϵ_I}	0.938	0.032
σ_f^2	0.169	0.031
$\sigma_{v_E}^2$	0.022	0.010
$\sigma_{v_I}^2$	0.150	0.023

Panel B: Normality tests

		statistic	p-value
H_{S_f}	Kt	64.691	0.000
	Sk	22.542	0.000
	GH	87.233	0.000
H_{S_v}	Kt	8.210	0.002
	Sk	4.398	0.111
	GH	12.607	0.004
H_R	Kt	20.828	0.000
	Sk	7.818	0.020
	GH	28.645	0.000

Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2015Q2. Model: Bivariate cointegrated, dynamic single factor model (12); see section 5 for parameter definitions. In Panel A, estimates are Gaussian ML of the bivariate Gaussian likelihood of the stationary transformation $\Delta y_{Et} + \Delta y_{It}$ and $y_{Et} - y_{It}$ in the time domain. Standard errors are obtained from the asymptotic information matrix, which is computed using its frequency domain closed-form expression. In Panel B, the row labels H_{S_f} and H_{S_v} refer to the score tests in Propositions 4 and 5 corresponding to the S_f and S_v alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in section 3.5.2. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Figure 1: Linear projection versus conditional expectation in a non-Gaussian univariate static factor model



Notes: The observed variable is $y_t = \frac{1}{2}x_t + \frac{\sqrt{3}}{2}\epsilon_t$. We assume that the joint distribution of x_t and ϵ_t is asymmetric Student t with zero mean, identity covariance matrix, 8 degrees of freedom and skewness vector parameter $\mathbf{b} = (-1, 0)'$. Given that the joint distribution of y_t and x_t will also be an asymmetric Student t, we can use the expressions in Mencía (2012) to compute the conditional expectation of x_t given y_t.

Figure 2: Expenditure (GDP) and income (GDI) measures of real output

Figure 2a: Quarterly real (log) GDP and (log) GDI

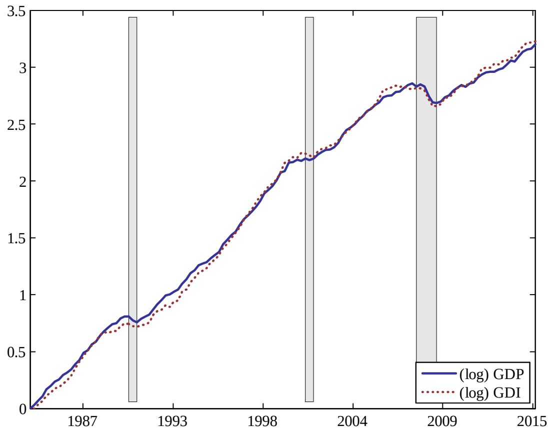


Figure 2b: Quarterly real GDP and GDI growth

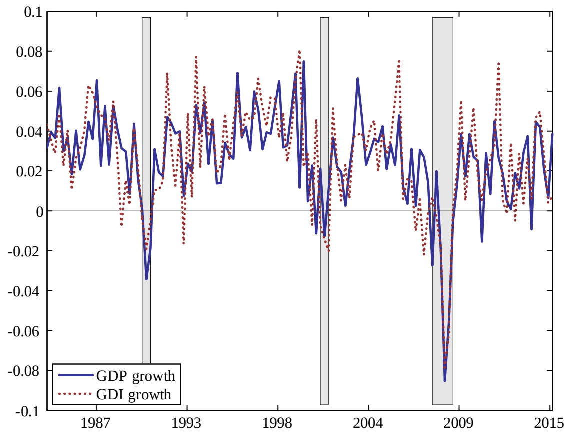
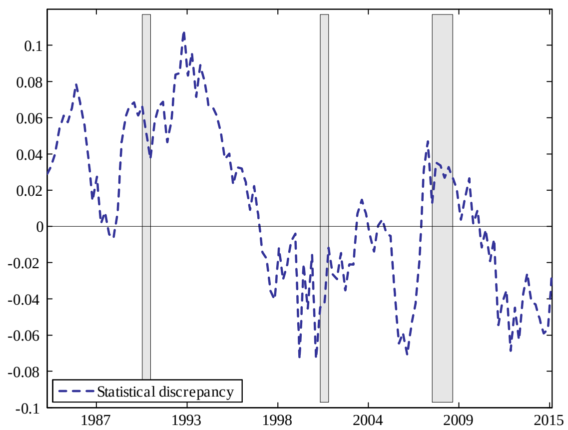


Figure 2c: Statistical discrepancy



Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2015Q2. Statistical discrepancy is defined as $\log(GDP) - \log(GDI)$. Shaded areas represent NBER recessions.

Figure 3: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q3 to 2007Q2.

Figure 3a: Smoothed innovations
for the underlying factor

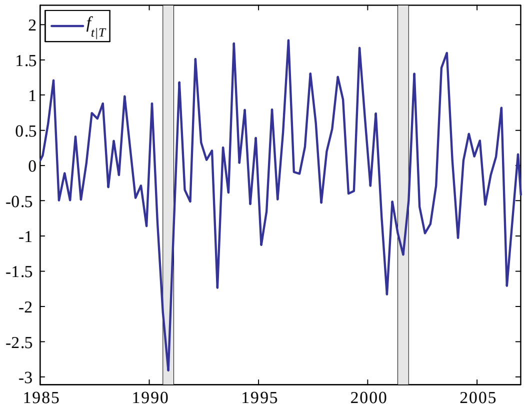


Figure 3b: Smoothed innovations
for the measurement errors

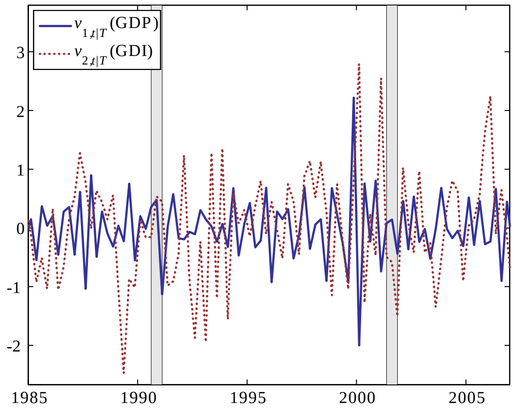


Figure 3c: Influence functions
for the underlying factor (kurtosis)

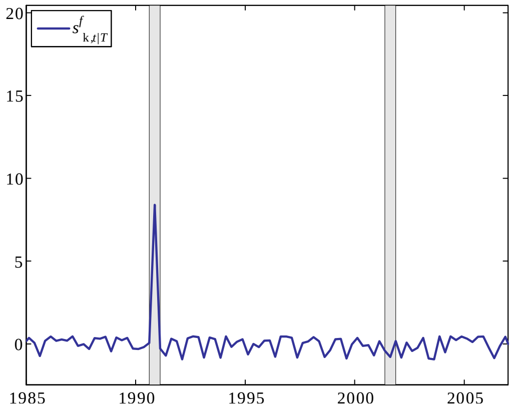


Figure 3d: Influence functions
for the measurement errors (kurtosis)

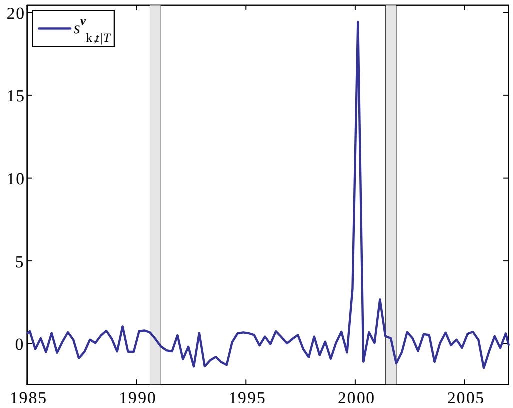


Figure 3e: Influence functions
for the underlying factor (skewness)

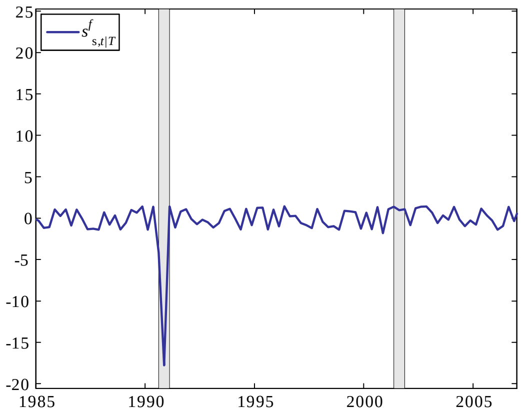
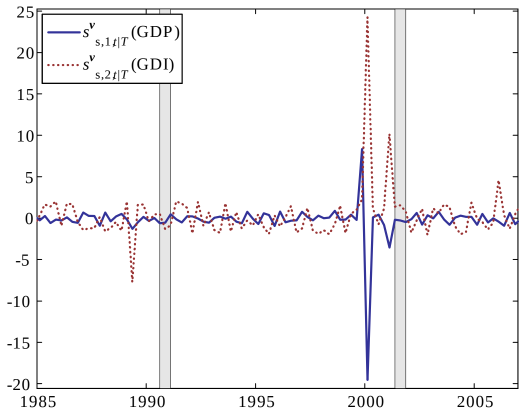
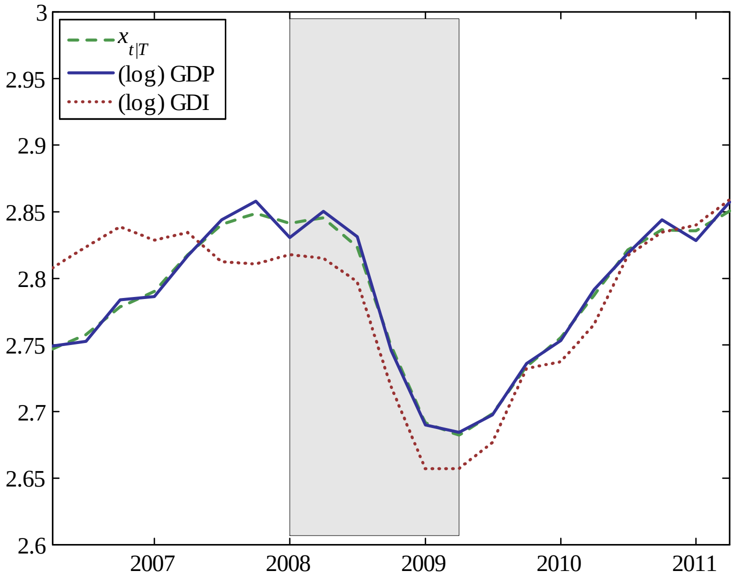


Figure 3f: Influence functions
for the measurement errors (skewness)



Notes: Smoothed innovations and influence functions were obtained from fitting the bivariate cointegrated, dynamic single factor model (12) to the quarterly real GDP and GDI from 1984Q3 to 2007Q2; see Table 4 for parameter estimates. Shaded areas represent NBER recessions.

Figure 4: GDP, GDI and smoothed estimate of real output around the Great Recession



Notes: The smoothed estimate $x_{t|T}$ was obtained from fitting the bivariate cointegrated, dynamic single factor model (12) to the quarterly real GDP and GDI from 1984Q3 to 2015Q2; see Table 5 for parameter estimates. The shaded area represents the NBER recession.

Figure 5: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q3 to 2015Q2.

Figure 5a: Smoothed innovations for the underlying factor

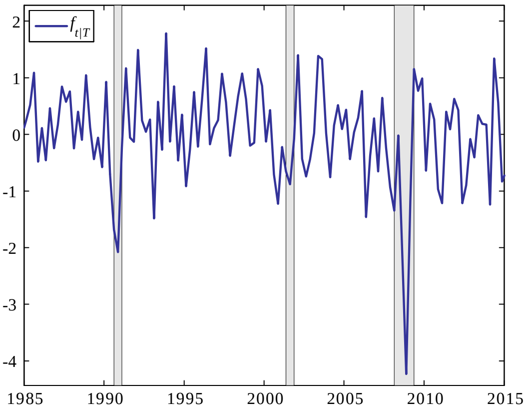


Figure 5b: Smoothed innovations for the measurement errors

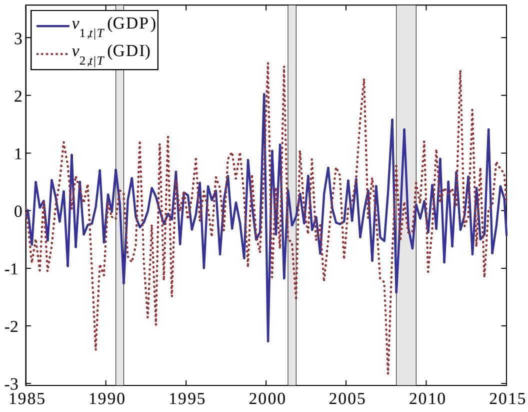


Figure 5c: Influence functions for the underlying factor (kurtosis)

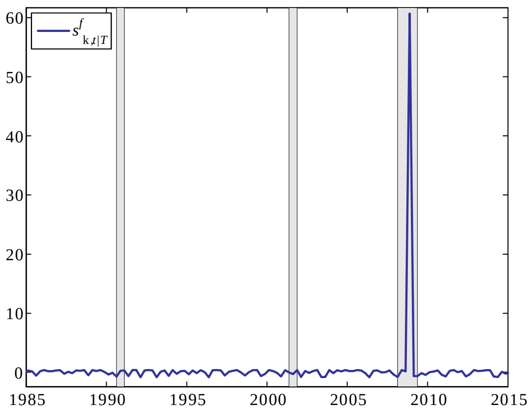


Figure 5d: Influence functions for the measurement errors (kurtosis)

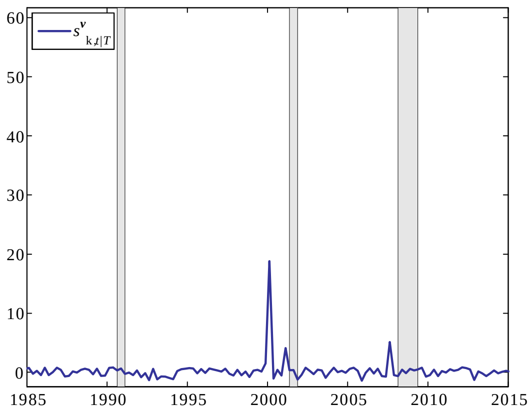


Figure 5e: Influence functions for the underlying factor (skewness)

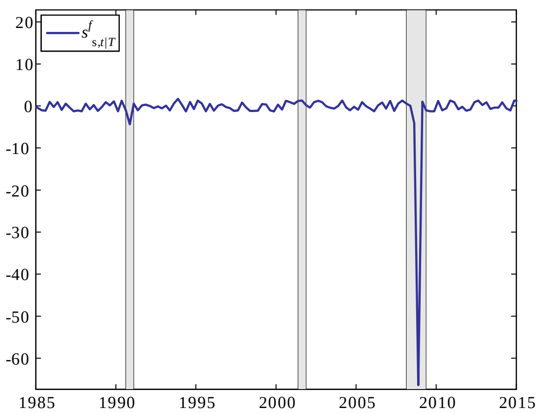
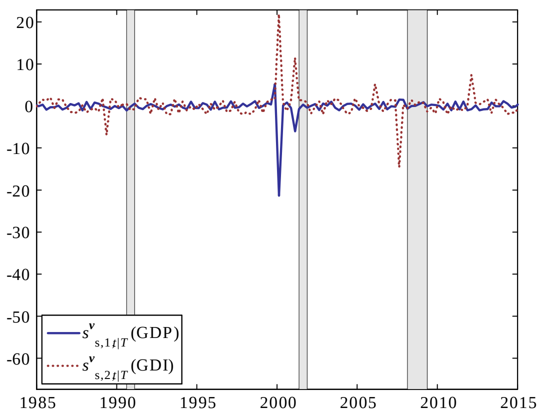


Figure 5f: Influence functions for the measurement errors (skewness)



Notes: Smoothed innovations and influence functions were obtained from fitting the bivariate cointegrated, dynamic single factor model (12) to the quarterly real GDP and GDI from 1984Q3 to 2015Q2; see Table 5 for parameter estimates. Shaded areas represent NBER recessions.

Appendix

A Proofs and auxiliary results

Lemmata

Lemma 1 *Let* $\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *be an* n_z*-dimensional real Gaussian random vector. Then,*
i) Expectation of second powers:

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

ii) Expectation of third powers:

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

iii) Expectation of fourth powers:

$$\begin{aligned} E\left[(\mathbf{z} \odot \mathbf{z})(\mathbf{z} \odot \mathbf{z})'\right] &= (\boldsymbol{\mu} \odot \boldsymbol{\mu})(\boldsymbol{\mu} \odot \boldsymbol{\mu})' + 2(\boldsymbol{\Sigma} \odot \boldsymbol{\Sigma}) + vecd(\boldsymbol{\Sigma})vecd'(\boldsymbol{\Sigma}) \\ &\quad + 4(\boldsymbol{\Sigma} \odot \boldsymbol{\mu}\boldsymbol{\mu}') + vecd(\boldsymbol{\mu}\boldsymbol{\mu}')vecd'(\boldsymbol{\Sigma}) + vecd(\boldsymbol{\Sigma})vecd'(\boldsymbol{\mu}\boldsymbol{\mu}'), \end{aligned}$$

where $\odot$ *denotes the Hadamard (or elementwise) product,* $vecd(\cdot)$ *is operator which stacks the diagonal elements of a square matrix in vector form and* $\boldsymbol{\ell}_{n_z}$ *is a vector of* n_z *ones.*

Proof. The proof is tedious but straightforward. □

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

$$\begin{aligned} \mathbf{m}_2(\mathbf{w}_1, \mathbf{w}_2) &= vec(\mathbf{w}_1\mathbf{w}_2'), \\ \mathbf{m}_3(\mathbf{w}_1, \mathbf{w}_2) &= vec[\mathbf{w}_1(\mathbf{w}_2 \odot \mathbf{w}_2)'], \\ \mathbf{m}_4(\mathbf{w}_1, \mathbf{w}_2) &= vec[(\mathbf{w}_1 \odot \mathbf{w}_1)(\mathbf{w}_2 \odot \mathbf{w}_2)'], \end{aligned}$$

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

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{xy}' & \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xz}' & \boldsymbol{\Sigma}_{yz}' & \boldsymbol{\Sigma}_{zz} \end{pmatrix}\right]$$

where $\mathbf{x}$ *is* n_x*-dimensional,* $\mathbf{y}$ *is* n_y*-dimensional, and* $\mathbf{z}$ *is* n_z*-dimensional. Then,*
i) Covariance with the first power:

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

ii) Covariance with the second power:

$$\begin{aligned} cov\left[\mathbf{m}_2(\mathbf{x}, \mathbf{x}), \mathbf{m}_2(\mathbf{y}, \mathbf{z})\right] &= (\mathbf{1}_{n_x \times n_z} \otimes \boldsymbol{\Sigma}_{xy}) \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\ &\quad + (\boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}_{n_y}') \odot \left(\boldsymbol{\ell}_{n_x}' \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_z}\right), \\ cov\left[\mathbf{m}_2(\mathbf{x}, \mathbf{x}), \mathbf{m}_3(\mathbf{y}, \mathbf{z})\right] &= \mathbf{0}, \end{aligned}$$

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

iii) Covariance with the third power:

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

$$cov\left[\mathbf{m}_3(\mathbf{x},\mathbf{x}),\mathbf{m}_4(\mathbf{y},\mathbf{z})\right]=\mathbf{0},$$

iv) Covariance with the fourth power:

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

where $\otimes$ denotes Kronecker product and $\mathbf{1}_{n_1\times n_2}$ denotes a matrix of ones of dimension $n_1\times n_2$.

Proof. Again, the proof is tedious but straightforward. □

Lemma 3 *Consider the model (1)-(2) where $\boldsymbol{\varepsilon}^*_t=(\boldsymbol{\varepsilon}^{*\mathrm{GH}\prime}_t,\boldsymbol{\varepsilon}^{*\mathrm{N}\prime}_t)'$, with $\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t\sim GH_R(\eta,\psi,\boldsymbol{\beta})$ and $\boldsymbol{\varepsilon}^{*\mathrm{N}}_t\sim N(\mathbf{0};\mathbf{I}_{K-R})$. Let $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})'\boldsymbol{\varepsilon}^{*\mathrm{GH}}_t(\boldsymbol{\theta})$ and*

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

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 $\psi>0$,

$$
\begin{aligned}
\lim_{\eta\to0^+}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\eta} &= -\lim_{\eta\to0^-}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\eta}=\frac{1}{T}\sum_{t=1}^{T}s_{\mathrm{GH},t}(\boldsymbol{\theta}),\ \text{and}\\
\lim_{\eta\to0^\pm}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\psi} &= 0,
\end{aligned}
$$

ii) For any $\boldsymbol{\beta}\in\mathbb{R}^R$ and $\eta\in\mathbb{R}$,

$$\lim_{\psi\to0^+}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\eta}=0,\ \text{and}\ \lim_{\psi\to0^+}\frac{2}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\psi}=\frac{1}{T}\sum_{t=1}^{T}s_{\mathrm{GH},t}(\boldsymbol{\theta}),$$

iii) Either way,

$$\lim_{\eta\cdot\psi\to0}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\boldsymbol{\beta}}=\mathbf{0},\ \text{and}\ \lim_{\eta\cdot\psi\to0}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\Xi_T|\boldsymbol{\phi})}{\partial\boldsymbol{\theta}}=\text{Gaussian score}.$$

Proof. See Mencía and Sentana (2012). $\square$

Lemma 4 *Consider the model (1)-(2) where* $\{\boldsymbol{\varepsilon}_t^*\}_{t=-\infty}^{\infty}$ *is white noise with identity covariance matrix. Moreover, assume that all the eigenvalues of* $\mathbf{F}$ *are inside the unit circle. If we observe the double-infinite sequence* $\mathbf{Y}_\infty = \{\mathbf{y}_t\}_{t=-\infty}^{\infty}$*, then the linear projection*

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_{t-1|\infty} \\ \hat{\boldsymbol{\varepsilon}}_{t|\infty}^* \end{pmatrix} = \mathcal{P}\left[\left. \begin{pmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\varepsilon}_t^* \end{pmatrix} \right| \mathbf{Y}_\infty \right] = \begin{bmatrix} \boldsymbol{\Psi}(L) \\ \boldsymbol{\Upsilon}(L) \end{bmatrix} \mathbf{y}_t,$$

where $\boldsymbol{\Psi}$ *and* $\boldsymbol{\Upsilon}$ *are absolutely summable two-sided filters in the lag operator* L*, is given by*

$$\begin{bmatrix} \boldsymbol{\Psi}(\mathbf{z}) \\ \boldsymbol{\Upsilon}(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \mathbf{z}\mathbf{F}^{-1}(\mathbf{z})\mathbf{M} \\ \mathbf{I}_K \end{bmatrix} \mathbf{D}'(\mathbf{z}^{-1}) \left[\mathbf{D}(\mathbf{z})\mathbf{D}'(\mathbf{z}^{-1}) \right]^{-1},$$

where

$$\mathbf{F}^{-1}(L) = (\mathbf{I}_M - \mathbf{F}L)^{-1} = \sum_{j=0}^{\infty} \mathbf{F}^j L^j \text{ and } \mathbf{D}(L) = \mathbf{H}\mathbf{F}^{-1}(L)\mathbf{M} = \sum_{j=0}^{\infty} \mathbf{D}_j L^j$$

with $\mathbf{D}_j = \mathbf{H}\mathbf{F}^j\mathbf{M}$ *for all* j*.*

Proof. Given that $\mathbf{y}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t^*$, the joint autocovariance generating function for $(\mathbf{y}_t', \boldsymbol{\varepsilon}_t^{*\prime})'$ is easily seen to be

$$\mathbf{G}(\mathbf{z}) = \begin{bmatrix} \mathbf{G}_{yy}(\mathbf{z}) & \mathbf{G}_{y\varepsilon}(\mathbf{z}) \\ \mathbf{G}_{\varepsilon y}(\mathbf{z}) & \mathbf{G}_{\varepsilon\varepsilon}(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \mathbf{D}(\mathbf{z})\mathbf{D}'(\mathbf{z}^{-1}) & \mathbf{D}(\mathbf{z}) \\ \mathbf{D}'(\mathbf{z}^{-1}) & \mathbf{I}_K \end{bmatrix}$$

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

$$\hat{\boldsymbol{\varepsilon}}_{t|\infty}^* = \mathbf{G}_{\varepsilon y}(L)\mathbf{G}_{yy}^{-1}(L)\mathbf{y}_t = \mathbf{D}'(L^{-1}) \left[\mathbf{D}(L)\mathbf{D}'(L^{-1}) \right]^{-1} \mathbf{y}_t$$

It is then easily checked that for every t, $\hat{\boldsymbol{\varepsilon}}_{t|\infty}^*$ is well-defined as a mean-square limit under the assumptions of the Lemma. Moreover, because

$$\boldsymbol{\xi}_{t-1} = L\mathbf{F}^{-1}(L)\mathbf{M}\boldsymbol{\varepsilon}_t^*,$$

the filter for $\boldsymbol{\xi}_{t-1}$ follows from the filter for $\boldsymbol{\varepsilon}_t^*$, and it is also well-defined. $\square$

Proposition 1

From Lemma 3, we can obtain the expression for the score with respect to τ for a fixed value of the skewness parameter vector $\boldsymbol{\beta}$, $s_{\text{GH},t}(\boldsymbol{\theta}) = s_{\text{k},t}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\text{s},t}(\boldsymbol{\theta})$, which corresponds to the M-step of the EM algorithm. Next, we can apply the E-step to each of the components separately.

As for $s_{\text{k},t}(\boldsymbol{\theta})$, we have that $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})|\,\mathbf{Y}_T, \boldsymbol{\theta} \sim N[\boldsymbol{\varepsilon}_{t|T}^*(\boldsymbol{\theta}), \boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})]$ under the null of normality, so that

$$s_{\text{k},t|T}(\boldsymbol{\theta}) = c_0 + c_1 E\left[\varsigma_t(\boldsymbol{\theta})|\,\mathbf{Y}_T, \boldsymbol{\theta}\right] + c_2 E\left[\varsigma_t^2(\boldsymbol{\theta})\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$$

involves the computation of $E\left[\varsigma_t(\boldsymbol{\theta})|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$ and $E\left[\varsigma_t^2(\boldsymbol{\theta})\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$. To compute the first expectation, we can write

$$\begin{aligned} E\left[\varsigma_t(\boldsymbol{\theta})|\,\mathbf{Y}_T\right] &= E\left[\boldsymbol{\varepsilon}_t^{*\text{GH}}(\boldsymbol{\theta})'\boldsymbol{\varepsilon}_t^{*\text{GH}}(\boldsymbol{\theta})\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right] \\ &= tr\left\{E\left[\boldsymbol{\varepsilon}_t^{*\text{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\text{GH}}(\boldsymbol{\theta})'\,|\mathbf{Y}_T, \boldsymbol{\theta}\right]\right\} \\ &= tr[\boldsymbol{\Omega}_{t|T}^{\text{GH}}(\boldsymbol{\theta})] + vec(\mathbf{I}_R)'vec[\boldsymbol{\varepsilon}_{t|T}^{*\text{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{*\text{GH}}(\boldsymbol{\theta})'], \end{aligned}$$

where the first equality follows from the fact that $tr(A'B) = tr(BA')$, and the second one from Lemma 1.i. As for the second expectation,

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

where the first equality is a rewriting of $\varsigma_t^2(\boldsymbol{\theta})$, the second one follows from the aforementioned property of the trace, and the third one from Lemma 1.iii. Finally, to obtain the expression for $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$ we have made use of the the following identities

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

together with

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

Similarly, in order to compute

$$
\mathbf{s}_{\mathrm{s},t|T}(\boldsymbol{\theta}) = c_3E\left[\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\,|\mathbf{Y}_T,\boldsymbol{\theta}\right] + E\left[\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\varsigma_t(\boldsymbol{\theta})\,|\mathbf{Y}_T,\boldsymbol{\theta}\right],
$$

we need the expectation of the first component, which is trivially $E\left[\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\,|\mathbf{Y}_T,\boldsymbol{\theta}\right] = \boldsymbol{\varepsilon}_{t|T}^{*\mathrm{GH}}(\boldsymbol{\theta})$. We also need

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

where we have used the fact that $\varsigma_t(\boldsymbol{\theta}) = \left[\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_t^{*\mathrm{GH}}(\boldsymbol{\theta})\right]'\boldsymbol{\ell}_R$ in the first equality, and applied Lemma 1.ii. in the last one. Finally, we obtain the desired result by exploiting the fact that

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

after re-arranging terms. □

Proposition 2

Following the same steps as in Proposition 1, but conditioning on $\mathbf{Y}_\infty$ instead of $\mathbf{Y}_T$, we can obtain $\mathbf{s}_{\text{k},t|\infty}(\boldsymbol{\theta}) = E\left[\mathbf{s}_{\text{k},t}(\boldsymbol{\theta})\,|\mathbf{Y}_\infty, \boldsymbol{\theta}\right]$ and $\mathbf{s}_{\text{s},t|\infty}(\boldsymbol{\theta}) = E\left[\mathbf{s}_{\text{s},t}(\boldsymbol{\theta})\,|\mathbf{Y}_\infty, \boldsymbol{\theta}\right]$. Specifically, we can write

$$\begin{bmatrix} s_{\text{k},t|\infty}(\boldsymbol{\theta}) - b_0(\boldsymbol{\theta}) \\ \mathbf{s}_{\text{s},t|\infty}(\boldsymbol{\theta}) \end{bmatrix} = \mathbf{B}'(\boldsymbol{\theta})\mathbf{m}_{t|\infty}(\boldsymbol{\theta}) \text{ where } \mathbf{B}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0} & \mathbf{b}_1(\boldsymbol{\theta}) \\ \mathbf{b}_2(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_3(\boldsymbol{\theta}) \\ \mathbf{b}_4(\boldsymbol{\theta}) & \mathbf{0} \end{bmatrix},$$

and $\mathbf{m}_{t|\infty}(\boldsymbol{\theta}) = [\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{4,t|\infty}(\boldsymbol{\theta})]'$, where

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

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

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

Next, we can use Lemma 4 to obtain $\boldsymbol{\Gamma}_j = E[\boldsymbol{\varepsilon}_{t|\infty}^{*\text{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t-j|\infty}^{*\text{GH}}(\boldsymbol{\theta})']$, which corresponds to the j^{th} order autocovariance matrix of the Wiener-Kolmogorov filter for $\boldsymbol{\varepsilon}_t^*$ based on $\mathbf{Y}_\infty$ for any integer j. Further, we can apply Lemma 2 to obtain:

i) Covariance matrices with the first power:

$$cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right] = \mathbf{0}, \tag{A1}$$

$$\begin{aligned}
cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t-j|\infty}(\boldsymbol{\theta})\right] &= 2\left[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right)\\
&\quad + \left[vecd'(\boldsymbol{\Gamma}_0) \otimes \mathbf{1}_{K\times K}\right] \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j\right),
\end{aligned} \tag{A2}$$

$$cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right] = \mathbf{0}, \tag{A3}$$

ii) Covariance matrices with the second power:

$$\begin{aligned}
cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right] &= (\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\\
&\quad + \left(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K\right),\\
cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t-j|\infty}(\boldsymbol{\theta})\right] &= \mathbf{0},
\end{aligned} \tag{A4}$$

$$\begin{aligned}
cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right] &= 4\left[\boldsymbol{\ell}_{K^2} \otimes vec'(\boldsymbol{\Gamma}_0)\right] \odot cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right]\\
&+2\left[\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}'_K \otimes vecd'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right) \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\\
&+2\left[\boldsymbol{\ell}_{K^2} \otimes vecd'(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}'_K\right] \odot (\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot \left(\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K\right),
\end{aligned} \tag{A5}$$

iii) Covariance matrices with the third power:

$$\begin{aligned}
cov\left[\mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t-j|\infty}(\boldsymbol{\theta})\right] &= [vecd(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}_K \otimes \boldsymbol{\ell}'_{K^2}] \odot \{\boldsymbol{\ell}_K \otimes cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t-j|\infty}(\boldsymbol{\theta})\right]\}\\
&+2\,(\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot [(\boldsymbol{\Gamma}_j \odot \boldsymbol{\Gamma}_j) \otimes \mathbf{1}_{K\times K}]\\
&+2[vec(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}'_{K^2}] \odot [\boldsymbol{\ell}_{K^2} \otimes vecd'(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}'_K] \odot (\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K)\\
&+4[vec(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}'_{K^2}] \odot [\boldsymbol{\ell}_{K^2} \otimes vec'(\boldsymbol{\Gamma}_0)] \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\\
&+4(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K) \odot (\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K) \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K}),
\end{aligned}$$

$$cov\left[\mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right]=\mathbf{0}, \tag{A6}$$

iv) Covariance matrix of the fourth power:

$$\begin{aligned}
cov\left[\mathbf{m}_{4,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right] &= 4cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right]\odot cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right] \\
&+4[vec(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right] \\
&+2[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes\boldsymbol{\ell}_K\otimes vecd'(\boldsymbol{\Gamma}_0)\right]\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K})\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K}) \\
&+2[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vecd'(\boldsymbol{\Gamma}_0\otimes\boldsymbol{\ell}_K)\right]\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right) \\
&+2[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes\boldsymbol{\ell}_K\otimes vecd'(\boldsymbol{\Gamma}_0)\right]\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right) \\
&+2[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vecd'(\boldsymbol{\Gamma}_0\otimes\boldsymbol{\ell}_K)\right]\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j) \\
&+8[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vec'(\boldsymbol{\Gamma}_0)\right]\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K}) \\
&+8[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vec'(\boldsymbol{\Gamma}_0)\right]\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right) \\
&+8\left(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K}\right)\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right).
\end{aligned}$$

Then, we can show the asymptotic independence of the kurtosis and skewness components by noticing that

$$\begin{aligned}
cov\left[\mathbf{s}_{\mathrm{s},t|\infty}(\boldsymbol{\theta}),s_{\mathrm{k},t-j|\infty}(\boldsymbol{\theta})\right] &= \mathbf{b}'_1 cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right]\mathbf{b}_2 \\
&+\mathbf{b}'_1 cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right]\mathbf{b}_4 \\
&+\mathbf{b}'_3 cov\left[\mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right]\mathbf{b}_2 \\
&+\mathbf{b}'_3 cov\left[\mathbf{m}_{3,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{4,t-j|\infty}(\boldsymbol{\theta})\right]\mathbf{b}_4 \\
&=\mathbf{0},
\end{aligned}$$

where the last equality follows from (A1), (A3), (A4) and (A6). Moreover, we can simplify even further the relevant expressions by exploiting the cancellation of cross-terms within the variance formulas,

$$cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}),\mathbf{s}_{\mathrm{s},t-j|\infty}(\boldsymbol{\theta})\right]=\mathbf{0},\ \text{ and }\ cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}),s_{\mathrm{k},t-j|\infty}(\boldsymbol{\theta})\right]=\mathbf{0}. \tag{A7}$$

For the sake of brevity, we prove the above equalities for the case when $R=K$; the proof for the case $R<K$ is similar, but more tedious.

To show the first equality in (A7) notice that for any j, we obtain

$$cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}),\mathbf{m}_{1,t-j|\infty}(\boldsymbol{\theta})\right]\mathbf{b}_1=-\left[2\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0+tr(\boldsymbol{\Gamma}_0)\boldsymbol{\Gamma}_j\right],$$

because $\boldsymbol{\Omega}_\infty=\mathbf{I}_K-\boldsymbol{\Gamma}_0$ and $\mathbf{b}_1=-tr(\boldsymbol{\Gamma}_0)\mathbf{I}_K-\boldsymbol{\Gamma}_0$. The remaining part follows from exploiting the following equalities:

$$\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0=\left\{\left[\boldsymbol{\ell}_K\otimes vec'(\boldsymbol{\Gamma}_0)\right]\odot\left(\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right)\right\}(\boldsymbol{\ell}_K\otimes\mathbf{I}_K) \tag{A8}$$

and

$$tr(\boldsymbol{\Gamma}_0)\boldsymbol{\Gamma}_j=\left\{\left[vecd'(\boldsymbol{\Gamma}_0)\otimes\mathbf{1}_{K\times K}\right]\odot\left(\boldsymbol{\ell}'_K\otimes\Gamma_j\right)\right\}(\boldsymbol{\ell}_K\otimes\mathbf{I}_K). \tag{A9}$$

For instance, to show (A8), define

$$\mathbf{E}_K=\begin{bmatrix}\mathbf{e}_1\mathbf{e}'_1 & \dots & \mathbf{e}_K\mathbf{e}'_K\end{bmatrix},$$

with $(\mathbf{e}_1|\dots|\mathbf{e}_K)=\mathbf{I}_K$, as the unique $K\times K^2$ "diagonalization" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A})=\mathbf{E}'_K vec(\mathbf{A})$ (see Magnus (1988)). Similarly, let

$$\mathbf{E}_{K^2}=\begin{bmatrix}(\mathbf{e}_1\mathbf{e}'_1\otimes\mathbf{e}_1\mathbf{e}'_1) & (\mathbf{e}_1\mathbf{e}'_2\otimes\mathbf{e}_1\mathbf{e}'_2) & \dots & (\mathbf{e}_K\mathbf{e}'_{K-1}\otimes\mathbf{e}_K\mathbf{e}'_{K-1}) & (\mathbf{e}_K\mathbf{e}'_K\otimes\mathbf{e}_K\mathbf{e}'_K)\end{bmatrix},$$

which is $K^2 \times K^4$. Some straightforward algebra delivers the following key identities:

$$\begin{aligned}
\mathbf{e}_i'\mathbf{E}_K &= (\mathbf{e}_i \otimes \mathbf{e}_i)', \\
(\mathbf{e}_i \otimes \mathbf{e}_i)'\mathbf{E}_{K^2} &= (\mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i)', \\
\mathbf{E}_{K^2}'(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_i &= (\mathbf{I}_K \otimes \mathbf{e}_i \otimes \mathbf{I}_K \otimes \mathbf{e}_i)vec(\mathbf{I}_K), \\
\mathbf{E}_{K^2}'\boldsymbol{\ell}_{K^2} &= vec(\mathbf{I}_{K^2}),
\end{aligned}$$

for all $i = 1, \ldots, K$. Moreover, $\mathbf{E}_K$ and $\mathbf{E}_{K^2}$ have the important property that

$$(\mathbf{A} \odot \mathbf{B}) = \mathbf{E}_K(\mathbf{A} \otimes \mathbf{B})\mathbf{E}_{K^2}'$$

for any pair of $K \times K^2$ matrices $\mathbf{A}$ and $\mathbf{B}$. As a consequence, we have that for any pair of indices $i_1, i_2 = 1, \ldots, K$,

$$\begin{aligned}
\mathbf{e}_{i_1}'\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \odot (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K')\}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_{i_2} &= \mathbf{e}_{i_1}'\mathbf{E}_K\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \otimes (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K')\} \\
&\quad \times \mathbf{E}_{K^2}'(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_{i_2} \\
&= (\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_1})'\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \otimes (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K')\} \times (\mathbf{I}_K \otimes \mathbf{e}_{i_2} \otimes \mathbf{I}_K \otimes \mathbf{e}_{i_2})vec(\mathbf{I}_K) \\
&= \{\mathbf{e}_{i_1}'[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)](\mathbf{I}_K \otimes \mathbf{e}_{i_2}) \otimes \mathbf{e}_{i_1}'(\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K')(\mathbf{I}_K \otimes \mathbf{e}_{i_2})\} \times vec(\mathbf{I}_K) \\
&= (\mathbf{e}_{i_2}'\boldsymbol{\Gamma}_0 \otimes \mathbf{e}_{i_1}'\boldsymbol{\Gamma}_j)vec(\mathbf{I}_K) = \mathbf{e}_{i_1}'\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0\mathbf{e}_{i_2}.
\end{aligned}$$

But since i_1, i_2 were arbitrary, we can conclude that (A8) holds. Analogous calculations allow us to show (A9). Therefore (A8) and (A9), together with the fact that $\mathbf{b}_3 = \boldsymbol{\ell}_K \otimes \mathbf{I}_K$ and (A2) imply that

$$cov\left[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{s}_{\mathrm{s},t-j|\infty}(\boldsymbol{\theta})\right] = cov[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{1,t|\infty}'(\boldsymbol{\theta})]\mathbf{b}_1' + cov[\mathbf{m}_{1,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{3,t|\infty}'(\boldsymbol{\theta})]\mathbf{b}_3' = \mathbf{0}.$$

As for the second equality in (A7), again given that $\boldsymbol{\Omega}_\infty = \mathbf{I}_K - \boldsymbol{\Gamma}_0$ and

$$\mathbf{b}_2 = -\frac{1}{2}tr(\boldsymbol{\Gamma}_0)vec(\mathbf{I}_K) - vec(\boldsymbol{\Gamma}_0),$$

we can then use the same tedious but straightforward arguments as before to show that

$$cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right] vec(\boldsymbol{\Gamma}_0) = \{[\boldsymbol{\ell}_{K^2} \otimes vec'(\boldsymbol{\Gamma}_0)] \odot cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right]\}\boldsymbol{\ell}_{K^2}$$

and

$$\begin{aligned}
tr(\boldsymbol{\Gamma}_0)cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t-j|\infty}(\boldsymbol{\theta})\right] vec(\mathbf{I}_K) &= \{\left[\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}_K' \otimes vecd'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K'\right) \\
\odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\}\boldsymbol{\ell}_{K^2} &+ \{[\boldsymbol{\ell}_{K^2} \otimes vecd'(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}_K'] \odot (\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot \boldsymbol{\ell}_K' \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K)\}\boldsymbol{\ell}_{K^2},
\end{aligned}$$

which together with the fact that $\mathbf{b}_4 = \boldsymbol{\ell}_{K^2}/4$ and (A5) imply that

$$cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), s_{\mathrm{k},t-j|\infty}(\boldsymbol{\theta})\right] = cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{2,t|\infty}'(\boldsymbol{\theta})\right]\mathbf{b}_2' + cov\left[\mathbf{m}_{2,t|\infty}(\boldsymbol{\theta}), \mathbf{m}_{4,t|\infty}'(\boldsymbol{\theta})\right]\mathbf{b}_4' = \mathbf{0},$$

as desired. This allows us to write

$$\lim_{T\to\infty} V\begin{bmatrix}\sqrt{T}\bar{s}_{\mathrm{k}|T}(\boldsymbol{\theta}_0) \\ \sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}|T}(\boldsymbol{\theta}_0)\end{bmatrix} = \begin{bmatrix}\mathcal{C}_{\mathrm{k}|\infty}(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_{\mathrm{s}|\infty}(\boldsymbol{\theta}_0)\end{bmatrix}$$

where the expressions for $\mathcal{C}_{\mathrm{k}|\infty}(\boldsymbol{\theta}_0)$ and $\mathcal{C}_{\mathrm{s}|\infty}(\boldsymbol{\theta}_0)$ can be found in the statement of the Proposition. □

Proposition 3

As stated in Lemma 3, the score with respect to the mean-variance parameter vector $\boldsymbol{\theta}$ converges to the Gaussian score as we approach the null hypothesis in any of the possible directions in which the *GH* distribution approaches Gaussianity. For the latent model we can find an explicit formula for it. Specifically, assume $\mathbf{H}(\boldsymbol{\theta})$ has full row-rank and $\mathbf{M}(\boldsymbol{\theta})$ is square and non-singular. Necessary conditions for this are $N \leq K = M$. To simplify the exposition and without loss of generality we set $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{I}_K$ and reinterpret the state vector as a re-scaled rotation of the state vector in the original setting. In particular, the re-specification is achieved by $\tilde{\boldsymbol{\xi}}_t = \mathbf{M}^{-1}(\boldsymbol{\theta})\boldsymbol{\xi}_t$, $\tilde{\mathbf{H}}(\boldsymbol{\theta}) = \mathbf{M}^{-1}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})$ and $\tilde{\mathbf{F}}(\boldsymbol{\theta}) = \mathbf{M}^{-1}(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})$. Hence, we effectively express the data generating process for $\{\mathbf{y}_t\}$ as

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

In what follows we assume $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{I}_K$ and keep the original notation for $\boldsymbol{\xi}_t$, $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{F}(\boldsymbol{\theta})$. The score with respect to the mean-variance parameter vector can then be expressed in terms of

$$\mathbf{J}(\boldsymbol{\theta}) = \left[\begin{array}{cc} \mathbf{I}_{K-N} & \mathbf{0} \\ \multicolumn{2}{c}{\mathbf{H}(\boldsymbol{\theta})} \end{array} \right],$$

which is non-singular. Applying the EM principle, let us define

$$\begin{aligned} \mathbf{s}_{\mathrm{MV},t|T}(\boldsymbol{\theta}) &= \mathbf{s}_{\mathrm{MVl},t|T}(\boldsymbol{\theta}) + \mathbf{s}_{\mathrm{MVs},t|T}(\boldsymbol{\theta}), \\ \mathbf{s}_{\mathrm{MVl},t|T}(\boldsymbol{\theta}) &= E\left\{ \mathbf{A}(\boldsymbol{\theta})vec\left[\boldsymbol{\xi}_{t-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})'\right] \middle| \mathbf{Y}_\infty \right\}, \\ \mathbf{s}_{\mathrm{MVs},t|T}(\boldsymbol{\theta}) &= E\left\{ \mathbf{B}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})' - \mathbf{I}_K\right] \middle| \mathbf{Y}_\infty \right\}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &= \frac{\partial vec'[\mathbf{F}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} + \frac{\partial vec'[\mathbf{J}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\mathbf{F}(\boldsymbol{\theta}) \otimes \mathbf{J}^{-1'}(\boldsymbol{\theta})] \\ \mathbf{B}(\boldsymbol{\theta}) &= \frac{1}{2}\frac{\partial vec'[\mathbf{J}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\mathbf{I}_K \otimes \mathbf{J}^{-1'}(\boldsymbol{\theta})] + \frac{1}{2}\frac{\partial vec'[\mathbf{J}'(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\mathbf{J}^{-1'}(\boldsymbol{\theta}) \otimes \mathbf{I}_K]. \end{aligned}$$

Then, the score with respect to the mean-variance parameter vector under the null can be written as

$$\mathbf{s}_{\mathrm{MV},t|T}(\boldsymbol{\theta}) = \mathbf{b}^{\mathrm{MV}}_{t|T}(\boldsymbol{\theta})'\mathbf{m}^{\mathrm{MV}}_{t|T}(\boldsymbol{\theta})$$

where$\boldsymbol{\Omega}^{\xi\varepsilon}_{t|T}(\boldsymbol{\theta})$ is the conditional covariance between $\boldsymbol{\xi}_{t-1}$ and $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$,

$$\mathbf{b}^{\mathrm{MV}}_{t|T}(\boldsymbol{\theta}) = [b^{\mathrm{MV}}_{0,t|T}(\boldsymbol{\theta}), \mathbf{b}^{\mathrm{MV}}_{2,t|T}(\boldsymbol{\theta})', \mathbf{b}^{\mathrm{MV}}_{\xi\varepsilon,t|T}(\boldsymbol{\theta})']',$$

with

$$\begin{aligned} b^{\mathrm{MV}}_{0,t|T}(\boldsymbol{\theta}) &= \mathbf{A}(\boldsymbol{\theta})vec[\boldsymbol{\Omega}^{\xi\varepsilon}_{t|T}(\boldsymbol{\theta})] + \mathbf{B}(\boldsymbol{\theta})vec[\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta}) - \mathbf{I}_K] \\ \mathbf{b}^{\mathrm{MV}}_{2,t|T}(\boldsymbol{\theta}) &= \mathbf{B}(\boldsymbol{\theta}) \\ \mathbf{b}^{\mathrm{MV}}_{\xi\varepsilon,t|T}(\boldsymbol{\theta}) &= \mathbf{A}(\boldsymbol{\theta}) \end{aligned}$$

and $\mathbf{m}^{\mathrm{MV}}_{t|T}(\boldsymbol{\theta}) = [1, \mathbf{m}'_{2,t|T}(\boldsymbol{\theta}), \mathbf{m}'_{\xi\varepsilon,t|T}(\boldsymbol{\theta})]'$, with $\mathbf{m}_{\xi\varepsilon,t|T}(\boldsymbol{\theta}) = vec[\boldsymbol{\xi}_{t-1|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^*_{t|T}(\boldsymbol{\theta})']$. The rest of the proof is a consequence of (A7) and

$$cov\left[\mathbf{m}_{\xi\varepsilon,t|\infty}(\boldsymbol{\theta}), s_{\mathrm{k},t-j|\infty}(\boldsymbol{\theta})\right] = \mathbf{0} \quad \text{and} \quad cov\left[\mathbf{m}_{\xi\varepsilon,t|\infty}(\boldsymbol{\theta}), s_{\mathrm{s},t-j|\infty}(\boldsymbol{\theta})\right] = \mathbf{0},$$

which can be shown in the same way as it has been done with the conditions in (A7). □

Proposition 4

It follows from Propositions 1, 2, and 3. $\square$

Proposition 5

It follows from applying the same argument as in the proof of Proposition 5 in Mencía and Sentana (2012) together with the results in Propositions 1, 2 and 3. $\square$

B Algorithm for computing the asymptotic variance

In this section we describe a numerically reliable and computationally efficient algorithm to obtain the asymptotic variance of the test statistics. As a preliminary step, we assume the researcher has (i) specified the model and (ii) computed if necessary the Gaussian maximum likelihood estimates, $\hat{\boldsymbol{\theta}}_T$.

It turns out that the Wiener-Kolmogorov filter for the setting in this paper always has a finite-order VARMA representation with scalar autoregressive part. This feature follows from the fact that the autocovariance generating functions for this model are rational polynomials. Specifically, there are positive integers p and q, a set of scalars $\phi_1, ..., \phi_p \in \mathbb{R}$ and a set of matrices $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, ..., \boldsymbol{\Theta}_q \in \mathbb{R}^{(M+K)\times K}$ such that

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

This is a useful result to the extent that for the class of models considered in this paper the coefficients $\phi_1, ..., \phi_p$ and matrices $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, ..., \boldsymbol{\Theta}_q$ can be obtained in terms of the parametrization of $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{M}$. Thus, the following algorithm can be employed to compute the auto-covariances of $(\hat{\boldsymbol{\xi}}'_{t-1|\infty}, \hat{\boldsymbol{\varepsilon}}^{*\prime}_{t|\infty})'$:

STEP 1: *Obtain the* VARMA *representation of the Wiener-Kolmogorov filter for the innovations.* This can be done using symbolic software –such as Mathematica– in terms of the matrices $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{M}$. We refer the reader to Lemma 4. Importantly, the VAR component is scalar.

STEP 2: *Compute the autocovariance function implied by the Wiener-Kolmogorov filter of the innovations.* To do so, consider a VARMA process with scalar VAR part for a K_x-dimensional process $\mathbf{x}_t$,

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

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\boldsymbol{\Theta}(z) = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 z + \ldots \boldsymbol{\Theta}_q z^q$. The error process $\mathbf{u}_t$ is K-dimensional and it is assumed to be white noise, i.e. $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}'_t) = \boldsymbol{\Sigma}$, $E(\mathbf{u}_t\mathbf{u}'_{t-j}) = \mathbf{0}$ for $j \neq 0$. Next, write the VARMA process in companion VAR(1) form as

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

where $\mathbf{X}_t = (\mathbf{x}_t, ..., \mathbf{x}_{t-p+1}, \mathbf{u}_t, ..., \mathbf{u}_{t-p+1})'$,

$$\mathbf{A} = \begin{pmatrix} \bar{\boldsymbol{\Phi}} \otimes \mathbf{I}_{K_x} & \mathbf{e}_1 \otimes \bar{\boldsymbol{\Theta}} \\ \mathbf{0} & \mathbf{J}_q \otimes \mathbf{I}_K \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \boldsymbol{\Theta}_0 \\ \mathbf{0} \\ \mathbf{I}_K \\ \mathbf{0} \end{pmatrix}$$

with $\mathbf{e}_1$ being the first vector of the canonical basis in $\mathbb{R}^p$,

$$\bar{\boldsymbol{\Phi}} = \begin{pmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix}, \quad \bar{\boldsymbol{\Theta}} = \begin{pmatrix} \boldsymbol{\Theta}_1 & \cdots & \boldsymbol{\Theta}_q \end{pmatrix}, \text{ and } \mathbf{J}_q = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{q-1} & \mathbf{0} \end{pmatrix}.$$

Suppose we can find an invertible matrix $\mathbf{C}$ and a block diagonal matrix $\mathbf{\Lambda}$ (with Jordan blocks) such that $\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}$. Then, we can transform the original system by defining $\mathbf{Z}_t = \mathbf{C}^{-1}\mathbf{X}_t$, a possibly complex-valued stochastic process that satisfies

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

with $\boldsymbol{\eta}_t = \mathbf{C}^{-1}\mathbf{Q}\mathbf{u}_t$ being white-noise (and possibly complex-valued). It can be shown that a computationally convenient decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1} \tag{B10}$$

where

$$\mathbf{C} = \begin{pmatrix} \bar{\mathbf{C}} \otimes \mathbf{I}_{K_x} & -(\bar{\mathbf{\Phi}}^{-q} \otimes \mathbf{I}_{K_x})\mathbf{\Theta}^* \\ \mathbf{0} & \mathbf{I}_{Kq} \end{pmatrix}, \; \mathbf{\Lambda} = \begin{pmatrix} \bar{\mathbf{\Lambda}} \otimes \mathbf{I}_{K_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_q \otimes \mathbf{I}_K \end{pmatrix},$$

and

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

with

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

and $\bar{\mathbf{\Phi}} = \bar{\mathbf{C}}\bar{\mathbf{\Lambda}}\bar{\mathbf{C}}^{-1}$ providing the Jordan decomposition of $\bar{\mathbf{\Phi}}$. Notice that the decomposition outlined above is convenient to handle large systems as it reduces substantially the size of the matrices on which the Jordan decomposition needs to be performed.

We can also show that the autocovariance function of the Wiener-Kolmogorov filter derived in Lemma 4 is the autocovariance function of the stable solution to the difference equation embodied in its VARMA representation. For that reason, we decompose $\mathbf{A}$ as in (B10), with the absolute values of the eigenvalues in decreasing order. But since we have assumed no unit roots, $K_{\mathrm{S}} = K_x p + Kq - K_{\mathrm{U}}$, where K_{U} is the number of roots outside the unit circle and K_{S} the number of roots inside the unit circle.

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

$$\mathbf{Z}_t = \begin{pmatrix} \mathbf{Z}_{\mathrm{U}t} \\ \mathbf{Z}_{\mathrm{S}t} \end{pmatrix}, \; \boldsymbol{\eta}_t = \begin{pmatrix} \boldsymbol{\eta}_{\mathrm{U}t} \\ \boldsymbol{\eta}_{\mathrm{S}t} \end{pmatrix}, \; \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{\mathrm{UU}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{\mathrm{SS}} \end{pmatrix}, \text{ and } \mathbf{R} = \begin{pmatrix} \mathbf{R}_{\mathrm{UU}} & \mathbf{R}_{\mathrm{US}} \\ \mathbf{R}_{\mathrm{SU}} & \mathbf{R}_{\mathrm{SS}} \end{pmatrix}.$$

Next, if we write

$$\mathbf{Z}_{\mathrm{U}t} = \mathbf{\Lambda}_{\mathrm{UU}}^{-1}(\mathbf{Z}_{\mathrm{U}t+1} - \boldsymbol{\eta}_{\mathrm{U}t+1}) \quad \text{and} \quad \mathbf{Z}_{\mathrm{S}t} = \mathbf{\Lambda}_{\mathrm{SS}}\mathbf{Z}_{\mathrm{S}t-1} + \boldsymbol{\eta}_{\mathrm{S}t};$$

and partition

$$\mathbf{\Gamma}_{\mathbf{Z}}(j) = \begin{bmatrix} \mathbf{\Gamma}_{\mathrm{UU}}(j) & \mathbf{\Gamma}_{\mathrm{US}}(j) \\ \mathbf{\Gamma}_{\mathrm{SU}}(j) & \mathbf{\Gamma}_{\mathrm{SS}}(j) \end{bmatrix} = \begin{bmatrix} E(\mathbf{Z}_{\mathrm{U}t}\bar{\mathbf{Z}}'_{\mathrm{U}t-j}) & E(\mathbf{Z}_{\mathrm{U}t}\bar{\mathbf{Z}}'_{\mathrm{S}t-j}) \\ E(\mathbf{Z}_{\mathrm{S}t}\bar{\mathbf{Z}}'_{\mathrm{U}t-j}) & E(\mathbf{Z}_{\mathrm{S}t}\bar{\mathbf{Z}}'_{\mathrm{S}t-j}) \end{bmatrix},$$

we can show that the autocovariance function of $\mathbf{Z}_t$ can be computed from

$$\begin{aligned}
vec[\mathbf{\Gamma}_{\mathrm{UU}}(0)] &= [\mathbf{I}_{K_{\mathrm{U}}^2} - (\mathbf{\Lambda}_{\mathrm{UU}}^{-1} \otimes \mathbf{\Lambda}_{\mathrm{UU}}^{-1})]^{-1} vec[\mathbf{\Lambda}_{\mathrm{UU}}^{-1} \mathbf{R}_{\mathrm{UU}} (\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-1})'], \\
\mathbf{\Gamma}_{\mathrm{UU}}(j) &= \mathbf{\Gamma}_{\mathrm{UU}}(0) (\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-j})', \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{UU}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{UU}}(-j), \quad \text{for } j < 0 \\
vec[\mathbf{\Gamma}_{\mathrm{SS}}(0)] &= [\mathbf{I}_{K_{\mathrm{S}}^2} - (\mathbf{\Lambda}_{\mathrm{SS}} \otimes \mathbf{\Lambda}_{\mathrm{SS}})]^{-1} vec\,(\mathbf{R}_{\mathrm{SS}})\,, \\
\mathbf{\Gamma}_{\mathrm{SS}}(j) &= \mathbf{\Lambda}_{\mathrm{SS}}^{j} \mathbf{\Gamma}_{\mathrm{SS}}(0), \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{SS}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{SS}}(-j), \quad \text{for } j < 0 \\
\mathbf{\Gamma}_{\mathrm{SU}}(j) &= -\sum_{h=1}^{j} (\overline{\mathbf{\Lambda}}_{\mathrm{SS}}^{j-h})' \mathbf{R}_{\mathrm{SU}} (\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-h})', \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{SU}}(j) &= \mathbf{0}, \quad \text{for } j \leq 0, \quad \text{and} \\
\mathbf{\Gamma}_{\mathrm{US}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{SU}}(-j).
\end{aligned}$$

Finally, we can recover the auto-covariance function of $\mathbf{X}_t$ from

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

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

STEP 3: *Compute the expressions that appear in Proposition 2.* To do so, one can obtain the autocovariance function of $\mathbf{m}_{h,t|\infty}(\boldsymbol{\theta})$ for $h = 1, \ldots, 4$ from the expressions in i), ii), iii) and iv) in the proof of Proposition *2*.

Next, add up the autocovariance matrices of $\mathbf{m}_{h,t|\infty}(\boldsymbol{\theta})$ for $h = 1 \ldots 4$ until some convergence criterion is satisfied. This gives a numerical approximation to $\boldsymbol{\kappa}_h(\hat{\boldsymbol{\theta}}_T)$ for $h = 1, \ldots, 4$. Finally, compute $\mathbf{b}_h(\hat{\boldsymbol{\theta}}_T)$, which only requires knowledge of the contemporaneous covariance matrix of the Wiener-Kolmogorov filter since $\mathbf{\Omega}_\infty = \mathbf{I}_K - \mathbf{\Gamma}_0$.

Codes for all the steps and detailed derivations for the expressions in STEP 2 are available upon request.

C The *GH* as a location-scale mixture of normals

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 Generalized Inverse Gaussian (*GIG*). If $\boldsymbol{\varepsilon}^*$ is a *GH* vector, then it can be expressed as

$$\boldsymbol{\varepsilon}^* = \boldsymbol{\alpha} + \mathbf{\Upsilon}\boldsymbol{\beta}\zeta^{-1} + \zeta^{-\frac{1}{2}}\mathbf{\Upsilon}^{\frac{1}{2}}\boldsymbol{\varepsilon}^\circ, \tag{C11}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^K$, $\mathbf{\Upsilon}$ 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 $\boldsymbol{\alpha}+\mathbf{\Upsilon}\boldsymbol{\beta}\zeta^{-1}$ and covariance matrix $\mathbf{\Upsilon}\zeta^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\mathbf{\Upsilon}$ 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 $\mathbf{\Upsilon}$ 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 $\psi = (1+\gamma)^{-1}$. Thus, 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.

D An alternative with independent latent variables

We might also envisage an additional alternative in which the elements of $\boldsymbol{\varepsilon}_t^*$ are cross-sectionally independent but non-Gaussian. Specifically, under such an alternative, each structural innovation would be independently distributed as a univariate GH:

$$\varepsilon_{it}^* \sim GH(\eta_i, \psi_i, \beta_i), \text{ for } i = 1, ..., K \text{ (alternative } I).$$

The main difference with the results in Propositions 4 and 5 is that now there are K different kurtosis parameters under the alternative, and therefore, K different scores. In addition, those scores could in principle be correlated. Therefore, the test statistic against symmetric Student t alternatives should be

$$LM_{IT}^{Student}(\boldsymbol{\theta}) = T\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right]' V^{-1}\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right]\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right],$$

where $\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta}) = (s_{\mathrm{k},t|T}^{1}(\boldsymbol{\theta}), ..., s_{\mathrm{k},t|T}^{K}(\boldsymbol{\theta}))'$. Under the null, the asymptotic distribution of this statistic will be χ_K^2. An analogous argument applies to the test statistic against skewness

$$LM_{IT}^{Skew}(\boldsymbol{\theta}) = T\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{p}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right]' V^{-1}\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{p}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right]\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{p}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})\right],$$

where $\mathbf{p}_{t|T}^{\mathrm{s}}(\boldsymbol{\theta}) = (s_{\mathrm{s},t|T}^{1}(\boldsymbol{\theta}), ..., s_{\mathrm{s},t|T}^{K}(\boldsymbol{\theta}))'$, which will also be asymptotically distributed as a χ_K^2 under the null. Given that the orthogonality between kurtosis and skewness components is preserved in this context too, we will have that

$$LM_{IT}^{GH}(\boldsymbol{\theta}) = LM_{IT}^{Student}(\boldsymbol{\theta}) + LM_{IT}^{Skew}(\boldsymbol{\theta})$$

will be asymptotically distributed as a χ_{2K}^2 under the null.

The asymptotic dependence between the elements of $\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})$, though, complicate the distribution of the one-sided, Kühn-Tucker version of the test. Specifically, we should now consider $\max[-T^{-1}\sum_{t=1}^{T} s_{\mathrm{k},t|T}^{i}(\hat{\boldsymbol{\theta}}_T, 0), 0]$ for each $i = 1, ..., K$. As a result, the joint test statistic will be a mixture of $K+1$ χ^2's, with degrees of freedom ranging from 0 to K, whose weights depend on the probability attached to each of the orthants based on the distribution of $\mathbf{n}_{t|T}^{\mathrm{k}}(\boldsymbol{\theta})$ under the null (see Gouriéroux, Holly and Monfort (1980)). Nevertheless, computation of the mixture weights as a function of the asymptotic variance is straightforward.

In Tables D1-D3 below we report the rejection rates of the aforementioned testing procedures for the same Monte Carlo design as in the paper.

Tables D1–D3: Monte Carlo rejection rates (in %) under the null and alternative hypotheses for the H_I tests

Table D1: Bivariate, cointegrated, dynamic single factor model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_I	Kt	0.98	4.86	9.60	58.22	5.25	55.74	71.50	6.28	66.66
	Sk	1.10	5.00	9.76	24.55	5.75	23.93	58.46	5.63	58.20
	GH	1.08	4.97	10.02	53.18	5.65	51.06	73.79	5.98	70.46

Table D2: Trivariate static factor model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_I	Kt	0.97	4.56	9.50	69.49	4.45	65.09	71.26	5.48	65.65
	Sk	0.93	5.41	10.07	28.05	4.64	26.57	40.83	4.92	28.58
	GH	0.98	4.95	9.55	63.93	4.42	59.00	69.07	5.44	60.33

Table D3: Local-level model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	S_v	J	S_f	S_v
H_I	Kt	1.57	5.75	11.46	53.21	28.57	16.75	86.13	55.56	38.31
	Sk	1.10	5.06	10.10	23.25	13.20	8.79	94.37	62.73	38.94
	GH	1.51	5.73	11.23	49.61	26.47	15.16	94.62	64.32	43.14

Notes: Results based on 10,000 samples of size $T = 250$. DGP for Table F1: Bivariate, cointegrated, dynamic single factor model of (12) with $\alpha_x = .5$, $\alpha_{\epsilon_E} = .2$, $\alpha_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$ represents the signal-to-noise ratio for y_{it} for $i = E, I$. DGP for Table F2: trivariate version of the static factor model (8) with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = (1,1,1)'$ and $\boldsymbol{\gamma} = q^{-1}(1,1,1)'$, where q reflects the signal-to-noise ratio, which we set to 2. DGP for Table F3: Local-level model discussed in section 3.4.2 in which the signal-to-noise ratio $q = \sigma_f^2/\sigma_v^2$ is set to 2. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\iota}_{K\times 1}$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

E Inferring real output from GDP and GDI over a long sample

Table E1: Parameter estimates and normality tests over the postwar period

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.755	0.110
δ	0.304	0.031
α_x	0.493	0.059
α_{ϵ_E}	0.265	0.196
α_{ϵ_I}	0.939	0.024
σ_f^2	0.526	0.054
$\sigma_{v_E}^2$	0.076	0.021
$\sigma_{v_I}^2$	0.093	0.019

Panel B: Normality tests

		statistic	p-value
H_{S_f}	Kt	19.061	0.000
	Sk	1.161	0.281
	GH	20.221	0.000
H_{S_v}	Kt	6.537	0.005
	Sk	3.859	0.145
	GH	10.396	0.011
H_R	Kt	13.266	0.000
	Sk	1.232	0.540
	GH	14.498	0.002

Notes: Data: Quarterly real GDP and GDI from 1952Q1 to 2015Q2. Model: Bivariate cointegrated, dynamic single factor model (12); see section 5 for parameter definitions. In Panel A, estimates are Gaussian ML of the bivariate Gaussian likelihood of the stationary transformation $\Delta y_{Et} + \Delta y_{It}$ and $y_{Et} - y_{It}$ in the time domain. Standard errors are obtained from the asymptotic information matrix, which is computed using its frequency domain closed-form expression. In Panel B, the row labels H_{S_f} and H_{S_v} refer to the score tests in Propositions 4 and 5 corresponding to the S_f and S_v alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in section 3.5.2. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Figure E1: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1952Q1 to 2015Q2.

Figure E1a: Smoothed innovations for the underlying factor

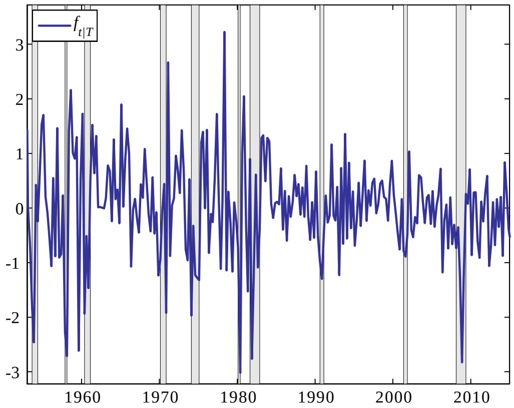


Figure E1b: Smoothed innovations for the measurement errors

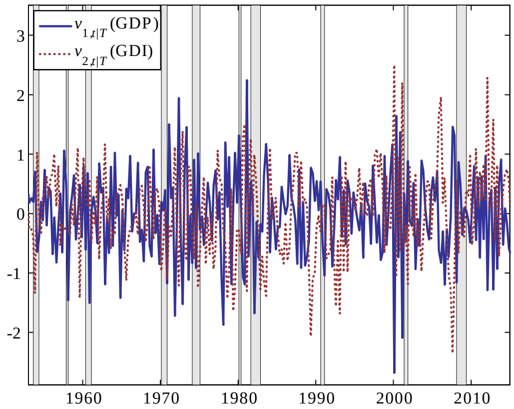


Figure E1c: Influence functions for the underlying factor (kurtosis)

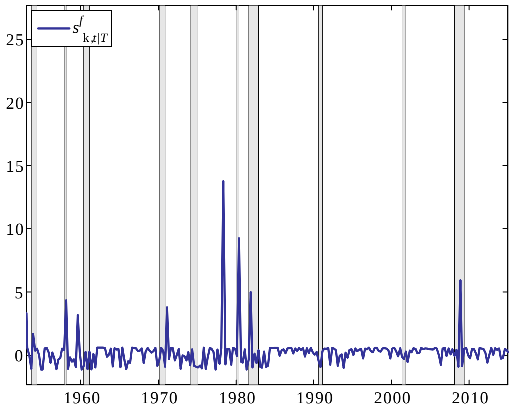


Figure E1d: Influence functions for the measurement errors (kurtosis)

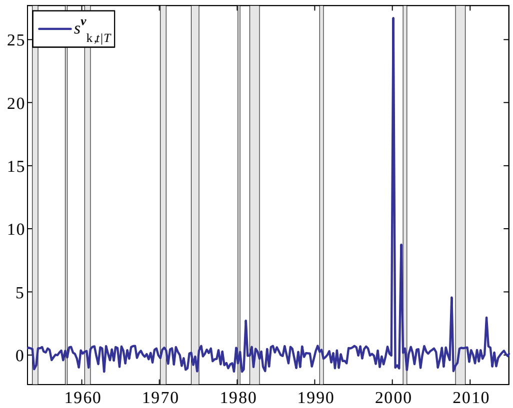


Figure E1e: Influence functions for the underlying factor (skewness)

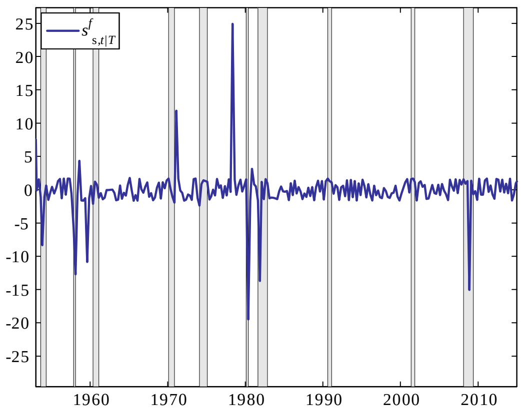
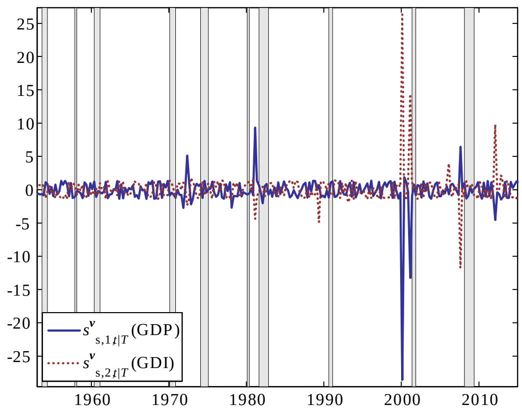


Figure E1f: Influence functions for the measurement errors (skewness)



Notes: Smoothed innovations and influence functions were obtained from fitting the bivariate cointegrated, dynamic single factor model (12) to the quarterly real GDP and GDI from 1952Q1 to 2015Q2; see Table C1 for parameter estimates. Shaded areas represent NBER recessions.