Quantitative Economics 10 (2019), 981–1017 1759-7331/20190981

Normality tests for latent variables

Martín Almuzara
CEMFI

Dante Amengual
CEMFI

Enrique Sentana
CEMFI

We exploit the rationale behind the Expectation Maximization algorithm to derive simple to implement and interpret LM normality tests for the innovations of the latent variables in linear 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 our tests to a common trend 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.

Keywords. Cointegration, gross domestic product, gross domestic income, kurtosis, Kuhn–Tucker test, skewness, supremum test, Wiener–Kolmogorov–Kalman smoother.

JEL classification. C32, C52, E01.

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 classic application of

Martín Almuzara: almuzara@cemfi.edu.es
Dante Amengual: amengual@cemfi.es
Enrique Sentana: sentana@cemfi.es

We are grateful to Gabriele Fiorentini and Javier Mencía, as well as to seminar audiences at Bilkent University, Boston University, European University Institute, Jiao Tong University, Toulouse University, Université de Montréal, the 2016 Barcelona GSE Summer Forum, the 69th ESEM (Geneva), the 2016 NBER-NSF Time Series Conference (New York), the 41st SAEe (Bilbao), the V Encuentro de la SEU (Montevideo), the 7th ICEEE (Messina), the 2017 SETA conference (Beijing), the 4th IAAE conference (Sapporo), the University of Lancaster Macroeconomics and Financial Time Series Analysis Workshop, the 2018 EcoSta conference (Hong Kong), the 2018 AMES (Seoul) and 2018 AMES (Auckland) for helpful comments, discussions, and suggestions. Three anonymous referees have also helped us greatly improve the paper. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Economy and Competitiveness through grant ECO 2014-59262 and the Santander—CEMFI Research Chair is gratefully acknowledged.

Available at http://qeconomics.org. https://doi.org/10.3982/QE859

signal extraction techniques whereby we obtain an improved aggregate (real) production series by combining its expenditure (GDP) and income (GDI) measures, which differ because they are constructed using largely independent data sources (see Landefeld, Seskin, and Fraumeni (2008) for a review).

We will use this model in Section 7, 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

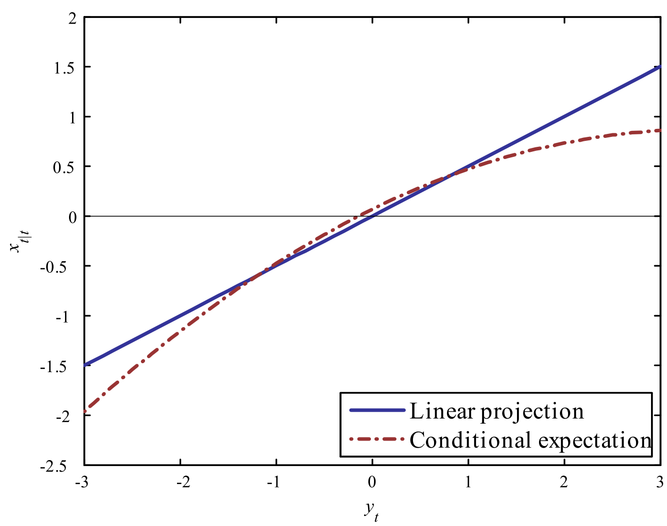


Figure 1. Linear projection versus conditional expectation in a non-Gaussian univariate static factor model. Notes: The observed variable is $y_t = \frac{1}{2}x_t + \frac{\sqrt{3}}{2}\epsilon_t$. We assume that the joint distribution of x_t and ϵ_t is asymmetric Student t with zero mean, identity covariance 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.

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 nonnormality 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 modeling 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 exploiting what we call "the EM principle." Specifically, we generalize Louis' (1982) score formula in order to obtain the first derivatives of the log likelihood with respect to the shape 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), resp.). 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 nonparametric 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 for 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 our normality tests against the Student t first and the GH distribution later in Sections 3 and 4, respectively. Then, in Section 5 we illustrate our procedures with two popular examples, while in Section 6 we discuss the results of our Monte Carlo experiments. Section 7 explores in detail the information about aggregate output in the GDP and GDI measures. Finally, we present our conclusions in Section 8. The Online Supplemental Material (Almuzara, Amengual, and Sentana (2019)) contains proofs and provides additional results.

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

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

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

$$\boldsymbol{\varepsilon}_t | \mathcal{I}_{t-1}, \boldsymbol{\phi} \sim i.i.d.\ D(\mathbf{0}, \mathbf{I}_K, \boldsymbol{\varphi}), \tag{3}$$

where $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varphi}')'$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is a vector of p first and second moment parameters, $\boldsymbol{\pi} : \Theta \to \mathbb{R}^N$ is the mean vector of the observed series, $\mathbf{H} : \Theta \to \mathbb{R}^{N\times M}$, $\mathbf{F} : \Theta \to \mathbb{R}^{M\times M}$ 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 *i.i.d.* innovations driving those variables whose distribution depends on a vector of shape parameters $\boldsymbol{\varphi}$, 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 $\boldsymbol{\theta}$ for which (1) and (2) constitute the true data generating process of $\{\mathbf{y}_t, \boldsymbol{\xi}_t\}$. In this context, static models will be such that $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}$ for all $\boldsymbol{\theta}$.

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

We 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

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

where $\mathbf{M}^{+}(\boldsymbol{\theta}) = [\mathbf{M}'(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})]^{-1}\mathbf{M}'(\boldsymbol{\theta})$ denotes the Moore–Penrose inverse of $\mathbf{M}(\boldsymbol{\theta})$. We also assume no linear combination of all the observables $\{\mathbf{y}_t\}$ has zero variance.

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 (2018) and the references therein). These assumptions are satisfied in virtually all applications of state-space models.

2.2 *Null and alternative hypotheses*

In Section 4, 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

[1]For example, Durbin and Koopman (2012) shifted the transition equation (2) forward by one period, as in Anderson and Moore (1979), and included measurement errors in (1), which they assume are orthogonal to the innovations in the state variables. On the other hand, Komunjer and Ng (2011) substituted 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.

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

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. Nevertheless, for clarity we first present the relevant results regarding testing multivariate normal versus multivariate Student t innovations in the next section, and then generalize them to the GH case.

In many applications, the researcher may only be interested in testing whether the source of nonnormality comes from a subset of the underlying components, which have some meaningful interpretation. In our empirical application, for example, it matters whether the potential nonnormality is a feature of the true GDP or its measurement errors. Given that we can always reorder the vector of structural innovations $\boldsymbol{\varepsilon}_t$ and post-multiply 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 $\boldsymbol{\varepsilon}_t^{\mathrm{GH}} = \mathbf{S}_{\mathrm{RK}}\boldsymbol{\varepsilon}_t$, with $\mathbf{S}_{\mathrm{RK}} = (\mathbf{I}_R, \mathbf{0})$, and to the remaining ones as $\boldsymbol{\varepsilon}_t^{\mathrm{N}}$.

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

1. The joint distribution of all structural innovations is *GH*: $H_J : \boldsymbol{\varepsilon}_t \sim \mathrm{GH}_K(\eta, \psi, \boldsymbol{\beta})$;
2. The joint distribution of the first R structural innovations is *GH* while the rest are Gaussian: $H_S : \boldsymbol{\varepsilon}_t^{\mathrm{GH}} \sim \mathrm{GH}_R(\eta, \psi, \boldsymbol{\beta})$, $\boldsymbol{\varepsilon}_t^{\mathrm{N}} \sim N_{K-R}(\mathbf{0}, \mathbf{I}_{K-R})$.[3]

3. Multivariate normal versus Student t innovations

The multivariate Student t distribution generalizes the multivariate normal distribution through a single additional parameter ν, which is usually known as the degrees of freedom. For convenience, we work with its reciprocal, η, so that Gaussianity requires $\eta \to 0^+$.

3.1 *The score under Gaussianity*

LM tests are usually obtained from the score associated to the (marginal) likelihood function of the observed variables, $f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})$, with $\mathbf{Y}_T = \mathrm{vec}(\mathbf{y}_1, \ldots, \mathbf{y}_T)$, evaluated under the Gaussian null. Unfortunately, the functional form of $f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})$ is generally unknown under the alternative, and 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 a variant of Louis' (1982) score formula, which is based on the so-called "EM principle"; see also Ruud (1991) and Tanner (1996).

Initially, we assume $\boldsymbol{\theta}$ is fixed and known, and later on we consider the effect of estimation of mean and variance parameters. Formally, the EM principle applied to this context says the following.

[3]We might also envisage an alternative situation in which the elements of $\boldsymbol{\varepsilon}_t$ are cross-sectionally independent but non-Gaussian; see Almuzara, Amengual, and Sentana (2017) for details.

Proposition 1. *Let $f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi})$ denote the density of $\mathbf{E}_T = \operatorname{vec}(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)$ with respect to Lebesgue measure on $\mathbb{R}^{\mathrm{KT}}$, which is continuous in $\mathbf{E}_T$ and differentiable in $\boldsymbol{\varphi}$. Then*

$$\frac{\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})}{\partial \boldsymbol{\varphi}} = E\left[\frac{\partial \ln f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\middle|\mathbf{Y}_T, \boldsymbol{\phi}\right]. \tag{5}$$

Remark 1. The identity (5) is different from Louis' (1982) original formula, in that it applies when $K > N$ but only for a subset of the parameters.

Remark 2. Let $f_{(\mathbf{Y},\mathbf{E})}(\mathbf{Y}_T, \mathbf{E}_T|\boldsymbol{\phi})$ denote the joint likelihood function for both observed variables $\{\mathbf{y}_t\}$ and structural innovations $\{\boldsymbol{\varepsilon}_t\}$ of model (1)–(2) for a sample of size T. This joint density will necessarily be singular in linear state-space models because of the restrictions the observed variables $\mathbf{Y}_T$ place on the latent $\mathbf{E}_T$. The same comment applies to the conditional likelihood function of the latent variables given the observed ones, $f_{\mathbf{E}|\mathbf{Y}}(\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\phi})$, which will usually be defined over a manifold of smaller dimension. Since the Kullback inequality implies that $E[\partial \ln f_{\mathbf{E}|\mathbf{Y}}(\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\phi})/\partial \boldsymbol{\varphi}|\mathbf{Y}_T, \boldsymbol{\phi}] = \mathbf{0}$, it follows that we can obtain $\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})/\partial \boldsymbol{\varphi}$ as the expected value of the unobservable score corresponding to $f_{(\mathbf{Y},\mathbf{E})}(\mathbf{Y}_T, \mathbf{E}_T|\boldsymbol{\phi})$ conditional on $\mathbf{Y}_T$ and $\boldsymbol{\phi}$. Therefore, an alternative formulation of (5) is

$$\frac{\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})}{\partial \boldsymbol{\varphi}} = E\left[\frac{\partial \ln f_{(\mathbf{Y},\mathbf{E})}(\mathbf{Y}_T, \mathbf{E}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\middle|\mathbf{Y}_T, \boldsymbol{\phi}\right], \tag{6}$$

where

$$f_{(\mathbf{Y},\mathbf{E})}(\mathbf{Y}_T, \mathbf{E}_T|\boldsymbol{\phi}) = 1_{\{\mathbf{Y}_T = \mathbf{D}_T(\mathbf{E}_T, \boldsymbol{\theta})\}} f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi}),$$

with $\mathbf{Y}_T = \mathbf{D}_T(\mathbf{E}_T, \boldsymbol{\theta})$ denoting the exact relationship between observed variables $\mathbf{Y}_T$ and innovations $\mathbf{E}_T$ implied by model (1)–(2) and $1_{\{.\}}$ the usual indicator function.

Importantly, while the mean and variance parameters enter in the indicator function, the shape parameters do not, so that the right-hand side of (6) is well-defined.[4]

Remark 3. One must be careful in applying the "EM principle" to the score with respect to $\boldsymbol{\theta}$. We deal with this situation in Proposition 3(b) and 6(b) (see the proof of Lemma 7 in the Online Supplemental Material A).

In the case of Student t innovations, we can use the expression provided in Fiorentini, Sentana, and Calzolari (2003) for the score with respect to η under the Gaussian null:

$$\frac{\partial \ln f_{(\mathbf{Y},\mathbf{E})}(\mathbf{Y}_T, \mathbf{E}_T|\boldsymbol{\phi})}{\partial \eta} = \frac{R(R+2)}{4} - \frac{R+2}{2}\varsigma_t^{\mathrm{GH}} + \frac{1}{4}\left(\varsigma_t^{\mathrm{GH}}\right)^2, \tag{7}$$

where $\varsigma_t^{\mathrm{GH}} = \boldsymbol{\varepsilon}_t^{\mathrm{GH}\prime}\boldsymbol{\varepsilon}_t^{\mathrm{GH}}$ and $\boldsymbol{\varepsilon}_t^{\mathrm{GH}} = \mathbf{S}_{\mathrm{RK}}\boldsymbol{\varepsilon}_t$. Thus, we can regard (7) as the M-step in Louis' (1982) formula (5). Next, we can apply the E-step by taking expectations. Specifically, if $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$ denotes the Kalman smoothed values of the innovations at t given $\mathbf{Y}_T$, which

[4]We thank a referee for pointing this out.

contains past, present, and future values of the observed series, and $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})$ the corresponding mean-square error, we have that $\boldsymbol{\varepsilon}_t|\mathbf{Y}_T,\boldsymbol{\theta}\sim N[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}),\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})]$ under the null of normality, so that

$$\frac{\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\phi})}{\partial \eta}=\frac{R(R+2)}{4}-\frac{R+2}{2}E\big[\varsigma_t^{\mathrm{GH}}|\mathbf{Y}_T,\boldsymbol{\theta}\big]+\frac{1}{4}E\big[\big(\varsigma_t^{\mathrm{GH}}\big)^2|\mathbf{Y}_T,\boldsymbol{\theta}\big]$$

only involves the computation of $E[\varsigma_t^{\mathrm{GH}}|\mathbf{Y}_T,\boldsymbol{\theta}]$ and $E[(\varsigma_t^{\mathrm{GH}})^2|\mathbf{Y}_T,\boldsymbol{\theta}]$, whose expressions we derive in the Online Supplemental Material A. Thus, we can show the following:

Proposition 2. *The score of the Student t log-likelihood with respect to the shape parameter η when $\eta=0$ is given by*

$$\bar{s}_{\mathrm{k}t|T}(\boldsymbol{\theta})=\frac{1}{T}\sum_{t=1}^{T}s_{\mathrm{k}t|T}(\boldsymbol{\theta})=\frac{1}{T}\sum_{t=1}^{T}\mathbf{b}'_{\mathrm{k}t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{k}t|T}(\boldsymbol{\theta}),$$

where $\mathbf{m}_{\mathrm{k}t|T}(\boldsymbol{\theta})=[1,\mathbf{m}'_{2t|T}(\boldsymbol{\theta}),\mathbf{m}'_{4t|T}(\boldsymbol{\theta})]'$, $\mathbf{b}_{\mathrm{k}t|T}(\boldsymbol{\theta})=[b_{0t|T}(\boldsymbol{\theta}),\mathbf{b}'_{2t|T}(\boldsymbol{\theta}),\mathbf{b}'_{4t|T}(\boldsymbol{\theta})]'$,

$$\begin{aligned}\mathbf{m}_{2t|T}(\boldsymbol{\theta})&=\operatorname{vec}\big[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})'\big],\\ \mathbf{m}_{4t|T}(\boldsymbol{\theta})&=\operatorname{vec}\big\{\big[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\big]\big[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\big]'\big\}.\end{aligned}\tag{8}$$

and

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

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

3.2 *Asymptotic covariance matrix of the score under Gaussianity*

As is well known, the Kalman smoothed process $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$ will typically be serially correlated in spite of $\boldsymbol{\varepsilon}_t$ being *i.i.d.* Consequently, the same will be true of $s_{\mathrm{k}t|T}(\boldsymbol{\theta})$. In addition, the autocovariances of $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$ change with both t and T. Nevertheless, we show in Online Supplemental Material B that it suffices to compute the autocovariances of powers of $\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})$, which is the Wiener–Kolmogorov filter of $\boldsymbol{\varepsilon}_t$ based on a double-infinite sample of the observable vector $\mathbf{y}_t$, for the purposes of obtaining the asymptotic variance of $\sqrt{T}\bar{s}_{\mathrm{k}t|T}(\boldsymbol{\theta})$. For that reason, we define $\mathbf{m}_{jt}(\boldsymbol{\theta})$ as $\mathbf{m}_{jt|T}(\boldsymbol{\theta})$ in (8) with $\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})$ in place of $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$, $\mathbf{b}_j(\boldsymbol{\theta})$ as $\mathbf{b}_{jt|T}(\boldsymbol{\theta})$ in Proposition 2 with $\boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})$ replacing $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})$,[5] and $\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta})$ as the associated average score.

[5] Under the usual controllability and observability conditions (see, e.g., Harvey (1989)), which we assume henceforth, $\boldsymbol{\Omega}_{t|\infty}(\boldsymbol{\theta})$ will not depend on t in steady state, so we can write $\boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})=\boldsymbol{\Omega}_{t|\infty}(\boldsymbol{\theta})$.

In practice, however, we do not generally know $\boldsymbol{\theta}$. Therefore, we need to obtain the asymptotic covariance matrix of $\sqrt{T}\bar{s}_{kT}(\hat{\boldsymbol{\theta}}_T)$, where $\hat{\boldsymbol{\theta}}_T$ is the Gaussian maximum likelihood estimator of $\boldsymbol{\theta}$, which is the efficient estimator under the null. Importantly, the second part of the following proposition shows that the sampling variability of the Gaussian ML estimators of $\boldsymbol{\theta}$ does not affect the asymptotic variance of the test:

Proposition 3. *Under the null hypothesis of Gaussian innovations*:

(a) $\lim_{T\to\infty} V[\sqrt{T}\bar{s}_{kT}(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{b}_4'(\boldsymbol{\theta})\boldsymbol{\kappa}_4(\boldsymbol{\theta})\mathbf{b}_4(\boldsymbol{\theta}) - \mathbf{b}_2'(\boldsymbol{\theta})\boldsymbol{\kappa}_2(\boldsymbol{\theta})\mathbf{b}_2(\boldsymbol{\theta}) = \mathcal{C}_k(\boldsymbol{\theta})$, *where*

$$\boldsymbol{\kappa}_i(\boldsymbol{\theta}) = \sum_{j=-\infty}^{\infty} \mathrm{cov}\big[\mathbf{m}_{it}(\boldsymbol{\theta}), \mathbf{m}_{it-j}(\boldsymbol{\theta})\big], \tag{9}$$

denotes the autocovariance generating function of $\mathbf{m}_{it}(\boldsymbol{\theta})$ *evaluated at one.*

(b) $\lim_{T\to\infty} \mathrm{cov}[\sqrt{T}\bar{s}_{kT}(\boldsymbol{\theta}), \sqrt{T}\bar{\mathbf{s}}_{MVT}(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{0}$, *where* $\bar{\mathbf{s}}_{MVT}(\boldsymbol{\theta})$ *denotes the average Gaussian score with respect to the conditional mean and variance parameters* $\boldsymbol{\theta}$.

3.3 *The test statistic*

We can easily compute a 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*

$$\mathrm{LM}_T^{\mathrm{Student}}(\hat{\boldsymbol{\theta}}_T) = T\frac{\bar{s}^2_{kt|T}(\hat{\boldsymbol{\theta}}_T)}{\mathcal{C}_k(\hat{\boldsymbol{\theta}}_T)},$$

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.[6] In this respect, it is important to mention that when there is a single

[6]Intuitively, under the null of normality

$$\sqrt{T}\frac{\bar{s}_{kt|T}(\hat{\boldsymbol{\theta}}_T)}{\sqrt{\mathcal{C}_k(\hat{\boldsymbol{\theta}}_T)}}$$

restriction, as in our case, those one-sided tests would be asymptotically locally most powerful.

4. Multivariate normal versus *GH* innovations

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

We can gain some intuition about the *GH* distribution by considering Blæsild's (1981) interpretation as a location-scale mixture of normals in which the mixing variable is a Generalized Inverse Gaussian (*GIG*). Specifically, if $\boldsymbol{\varepsilon}$ is a *GH* vector, then it can be expressed as

$$\boldsymbol{\varepsilon} = \boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\zeta^{-1} + \zeta^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\boldsymbol{\varepsilon}^{\circ}, \tag{10}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^K$, $\boldsymbol{\Upsilon}$ is a symmetric positive definite matrix of order K, $\boldsymbol{\varepsilon}^{\circ} \sim N(\mathbf{0}, \mathbf{I}_K)$ and the positive mixing variable ζ is an independent *GIG* with parameters $-\nu$, γ, and δ, or $\zeta \sim \mathrm{GIG}(-\nu, \gamma, \delta)$ for short, where $\nu \in \mathbb{R}$ and $\gamma, \delta \in \mathbb{R}^+$ (see Jørgensen (1982) and Johnson, Kotz, and Balakrishnan (1994) for further details). Obviously, the distribution of $\boldsymbol{\varepsilon}$ becomes a simple scale mixture of normals, and thereby spherical, when $\boldsymbol{\beta}$ is zero. By restricting $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$, Mencía and Sentana (2012) derived 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, $\eta = -0.5\nu^{-1}$ and $\psi = (1+\gamma)^{-1}$, whose product $\tau = \eta\psi$ effectively controls excess kurtosis in the vicinity of the Gaussian null.

4.2 *The score under Gaussianity*

As in Section 3, there is no analytical expression for the log-likelihood function under the alternative, so once again we resort to the generalized Louis' (1982) formula. However, we 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 $\tau \to 0$.

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 in our context, we need the following:

will be asymptotically distributed as a standard normal. Therefore, the one-sided nature of the alternative hypothesis implies that the relevant critical value for size $\varkappa$ is given by the $(1-\varkappa)$th quantile of a standard normal instead of the usual $(1-\varkappa/2)$th one.

Proposition 5. *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*

$$\begin{aligned} \bar{s}_{\mathrm{GH}T}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \frac{1}{T}\sum_{t=1}^{T}\bigl[s_{\mathrm{k}t|T}(\boldsymbol{\theta}) + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta})\bigr], \\ \mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{s}t|T}(\boldsymbol{\theta})\mathbf{m}_{\mathrm{s}t|T}(\boldsymbol{\theta}), \end{aligned} \tag{11}$$

where $\mathbf{m}_{\mathrm{s}t|T}(\boldsymbol{\theta}) = [\mathbf{m}'_{1t|T}(\boldsymbol{\theta}), \mathbf{m}'_{3t|T}(\boldsymbol{\theta})]'$, $\mathbf{b}_{\mathrm{s}t|T}(\boldsymbol{\theta}) = [\mathbf{b}'_{1t|T}(\boldsymbol{\theta}), \mathbf{b}'_{3t|T}(\boldsymbol{\theta})]'$,

$$\begin{aligned} \mathbf{m}_{1t|T}(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}), \\ \mathbf{m}_{3t|T}(\boldsymbol{\theta}) &= \mathrm{vec}\bigl\{\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\bigl[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})\bigr]'\bigr\}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{b}_{1t|T}(\boldsymbol{\theta}) &= \bigl[c_3 + \mathrm{tr}\bigl(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\bigr)\bigr]\mathbf{S}'_{\mathrm{RK}} + 2\mathbf{S}'_{\mathrm{RK}}\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}), \\ \mathbf{b}_{3t|T}(\boldsymbol{\theta}) &= \mathbf{S}'_{\mathrm{RK}}\boldsymbol{\ell}_R \otimes \mathbf{S}'_{\mathrm{RK}}, \end{aligned}$$

with $c_3 = -(R+2)$ *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, $\mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta})$.

4.3 *Asymptotic covariance matrix of the score under Gaussianity*

If we denote by $\bar{s}_{\mathrm{s}T}(\boldsymbol{\theta})$ the average score with $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})$ replaced by $\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})$, respectively, arguments analogous to those in Section 3.2 allow us to prove the following result:

Proposition 6. *Under the null hypothesis of Gaussian innovations*:

(a) $\sqrt{T}\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta})$ *and* $\sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}T}(\boldsymbol{\theta})$ *are asymptotically independent, and*

$$\lim_{T\to\infty} V\bigl[\sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}T}(\boldsymbol{\theta})|\boldsymbol{\theta}\bigr] = \mathbf{b}'_3(\boldsymbol{\theta})\boldsymbol{\kappa}_3(\boldsymbol{\theta})\mathbf{b}_3(\boldsymbol{\theta}) - \mathbf{b}'_1(\boldsymbol{\theta})\boldsymbol{\kappa}_1(\boldsymbol{\theta})\mathbf{b}_1(\boldsymbol{\theta}) = \mathcal{C}_{\mathrm{s}}(\boldsymbol{\theta}),$$

with $\boldsymbol{\kappa}_i(\boldsymbol{\theta})$ *defined in* (9).

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

The second part of this proposition, combined with the second part of Proposition 3, implies that the scores of the conditional mean and variance parameters $\boldsymbol{\theta}$ and the scores of the shape parameters $\boldsymbol{\varphi}$ are asymptotically independent under the null of Gaussianity, so that we need not worry about parameter uncertainty, at least in large samples. Interestingly, this implication is closely related to Proposition 3 in Fiorentini and Sentana (2007), which contains an analogous result for multivariate, dynamic location-scale models with non-Gaussian innovations. It is also related to Bontemps and Meddahi (2005), who show that univariate normality tests based on third and higher order Hermite polynomials are insensitive to parameter uncertainty, too.

In the Online Supplemental Material C, we provide a numerically reliable algorithm for computing the asymptotic covariance matrices $\mathcal{C}_{\mathrm{s}}(\boldsymbol{\theta})$ and $\mathcal{C}_{\mathrm{k}}(\boldsymbol{\theta})$ for any state space model.

4.4 *The test statistic*

If we combine Propositions 5 and 6, we can easily show that the LM test statistic for a given value of $\boldsymbol{\beta}$ will be given by

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

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

But 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. Remarkably, we can maximize $\mathrm{LM}_T^{\mathrm{GH}}(\boldsymbol{\theta}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ in closed form, and also obtain the asymptotic distribution of the resulting sup test statistic. Specifically:

Proposition 7. *The supremum with respect to $\boldsymbol{\beta}$ of the LM tests based on* (11) *is equal to*

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

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 $\mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta})$ from the first part of Proposition 6, 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 $\mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta})$, which should be equivalent in large samples to the corresponding LR test (see Proposition 6 in Mencía and Sentana (2012) for a more formal argument). 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 two χ^2 p-values.

5. Further discussion

In Section 7, we will use our methods for improving GDP measurement. But since they apply far more generally, in this section we first describe how to implement our testing procedures in a generic model. Next, we illustrate them with two popular textbook examples: a static factor model and the so-called local-level dynamic model. Finally, we use these examples to explicitly compare our proposed testing procedures to previous suggestions.

5.1 *Practical implementation*

Assume the researcher has already (i) specified the model and (ii) computed if necessary the Gaussian maximum likelihood estimates, $\hat{\boldsymbol{\theta}}_T$.

STEP 1: *Compute the smoothed influence functions*. Propositions 2 and 5 provide the explicit expressions required to compute the contribution to the score of each smoothed innovation, namely $s_{\mathrm{k}t|T}(\boldsymbol{\theta})$ and $\mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta})$, respectively. One simply needs as inputs the smoothed innovations $\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta})$, which are required to compute $\mathbf{m}_{jt|T}(\boldsymbol{\theta})$, for $j = 1, \dots, 4$, and the mean-square error for the vector of innovations being tested, $\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})$, which are necessary to compute the vector of coefficients $\mathbf{b}_{jt|T}(\boldsymbol{\theta})$, for $j = 0, \dots, 4$. Importantly, both of these quantities can be easily obtained from a standard Kalman filter-smoother.

STEP 2: *Obtain the asymptotic variance of the test statistics*. Although this can be done in different ways, in what follows we describe a numerically reliable and computationally efficient algorithm that avoids simulations.

STEP 2.1: *Obtain the* Varma *representation of the Wiener–Kolmogorov filter for the innovations*. It turns out that the Wiener–Kolmogorov filter always has a finite-order Varma representation with scalar autoregressive part for all the models in this paper. This feature follows from the fact that their autocovariance generating functions are rational polynomials. Specifically, there exist positive integers p and q, a set of scalars $\phi_1, \dots, \phi_p \in \mathbb{R}$ and a set of matrices $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_q \in \mathbb{R}^{(M+K)\times K}$ such that

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

This representation, in which importantly the Var component is scalar, is useful to the extent that the coefficients $\phi_1, \dots, \phi_p$ and matrices $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_q$ can be obtained in terms of the parameterization of the state matrices $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{M}$. This can be done using symbolic software such as Mathematica. We refer the reader to Lemma 4 in the Online Supplemental Material A.

STEP 2.2: *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 $\boldsymbol{\Theta}(z) = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 z + \cdots \boldsymbol{\Theta}_q z^q$. In Section C of the Online Supplemental Material, we provide a detailed algorithm to compute the autocovariance function of $\mathbf{x}_t$, from which we can compute the autocovariances of $(\hat{\boldsymbol{\xi}}'_{t-1|\infty}, \hat{\boldsymbol{\varepsilon}}'_{t|\infty})'$.

STEP 2.3: *Compute the expressions that appear in Propositions* 3 *and* 6. To do so, for example, one can obtain the autocovariance function of $\mathbf{m}_{h,t|\infty}(\boldsymbol{\theta})$ for $h = 1, \dots, 4$ from the expressions in (i), (ii), (iii), and (iv) in the proof of Proposition 6. Next, one can cumulate the autocovariance matrices of $\mathbf{m}_{h,t|\infty}(\boldsymbol{\theta})$ for $h = 1 \dots 4$ until some convergence criterion is satisfied. This gives a numerical approximation to $\boldsymbol{\kappa}_h(\hat{\boldsymbol{\theta}}_T)$ for $h = 1, \dots, 4$. Finally, one computes $\mathbf{b}_h(\hat{\boldsymbol{\theta}}_T)$, which only requires knowledge of the contemporaneous covariance matrix of the Wiener–Kolmogorov filter because $\boldsymbol{\Omega}_\infty = \mathbf{I}_K - \boldsymbol{\Gamma}_0$.

Codes for all the different steps above, as well as detailed derivations for the expressions in STEP 2 are available upon request.

5.2 *Static factor models*

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

$$\begin{gathered}\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c} f_t + \mathbf{v}_t, \\ \begin{pmatrix} f_t \\ \mathbf{v}_t \end{pmatrix} \Big| \mathcal{I}_{t-1}, \boldsymbol{\phi} \sim i.i.d.\ D\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix}, \boldsymbol{\eta}\right],\end{gathered} \tag{12}$$

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 $\boldsymbol{\theta} = (\boldsymbol{\pi}', \mathbf{c}', \boldsymbol{\gamma}')'$, with $\boldsymbol{\gamma} = \operatorname{vecd}(\boldsymbol{\Gamma})$.

We can easily express model (12) as in (1)–(2) with $\boldsymbol{\xi}_t = (f_t, \mathbf{v}_t')'$, $\mathbf{H}(\boldsymbol{\theta}) = (\mathbf{c}, \mathbf{I}_N)$, $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}$,

$$\mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \operatorname{diag}^{1/2}(\boldsymbol{\gamma}) \end{pmatrix}$$

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

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

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

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

so like in any other static model, $\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})$ will be white noise, with covariance matrix

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

In addition,

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

which has rank N rather than $N+1$, so that the conditional density will be degenerate. Hence, we will have that under the null,

$$\boldsymbol{\varepsilon}_t | \mathbf{Y}_T, \boldsymbol{\theta} \sim N\big[\boldsymbol{\varepsilon}_{t|t}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{\infty}(\boldsymbol{\theta})\big],$$

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 at 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 nonnormality 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 functions:

$$\left.\begin{aligned} H_3(f_t) &= f_t^3 - 3f_t, \\ H_4(f_t) &= f_t^4 - 6f_t^2 + 3, \end{aligned}\right\} \tag{13}$$

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 values of the elements of (13) are 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

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

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 $\mathbf{v}_t$.

5.3 *The local-level model*

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

$$\begin{aligned} y_t &= \pi + x_t + v_t, \\ x_t &= x_{t-1} + f_t, \end{aligned}$$

$$\begin{pmatrix} f_t \\ v_t \end{pmatrix} \Big| \mathcal{I}_{t-1}, \boldsymbol{\phi} \sim i.i.d.\, D\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}, \boldsymbol{\varphi}\right],$$

where x_t is the "signal" component, v_t the orthogonal "nonsignal" component, and $\boldsymbol{\theta}$ refers to the model parameters that characterize the autocovariance structure of the observed series. Once again, we can easily express this model as in (1)–(2) with $\boldsymbol{\xi}_t = (f_t, v_t')'$,

$\mathbf{H}(\boldsymbol{\theta}) = (1, 1)$,

$$\mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}$$

and $\boldsymbol{\varepsilon}_t = (f_t^*, v_t^*)'$, where $f_t^* = \sigma_f^{-1} f_t$ and $v_t^* = \sigma_v^{-1} v_t$.

Since there are only two shocks, we could look at (i) a test of joint normality, (ii) a test of normality of the "signal" with the maintained hypothesis of normality for the "nonsignal," and (iii) vice versa.

For the sake of brevity, let us focus on the nonsignal component. Proposition 5 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

$$E\big[s_{\eta t}^{S_v}(\boldsymbol{\theta}, \mathbf{0}) | \mathbf{Y}_T\big] = \frac{1}{2}\sqrt{\frac{3}{2}}\big[1 - \omega_{\mathrm{v}t|T}(\boldsymbol{\theta})\big]^2 - \sqrt{\frac{3}{2}}\big[1 - \omega_{\mathrm{v}t|T}(\boldsymbol{\theta})\big] v_{t|T}^{*2}(\boldsymbol{\theta}) + \frac{1}{2}\sqrt{\frac{1}{6}} v_{t|T}^{*4}(\boldsymbol{\theta}). \quad (14)$$

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

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

which in turns means that

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

Hence, expression (14) is proportional to the fourth order Hermite polynomial of the standardized variable $v_{t|T}^*(\boldsymbol{\theta})/\sqrt{1 - \omega_{\mathrm{v}t|T}(\boldsymbol{\theta})}$. 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 *i.i.d.* series. Unlike in the static model considered in Section 5.2, though, the elements of (14) are serially correlated.

5.4 *Comparison with alternative approaches*

5.4.1 *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

$$\mathrm{sk}_{\varepsilon_i} = m_{\varepsilon_i 3}/m_{\varepsilon_i 2}^{3/2},$$

where

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

is the jth centered sample moment of the smoothed innovations of either the signal ($i = 1$) or the noise ($i = 2$). Under normality, the asymptotic variance of $\mathrm{sk}_{\varepsilon_i}$ will be given by $\zeta_{\varepsilon_i}(\boldsymbol{\theta}, 3)$, where

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

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

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

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

It is interesting to compare these tests to our LM tests based on Propositions 4 and 7. The procedures proposed by Harvey and Koopman (1992) can be regarded as moment tests of

$$E\left[f_{t|T}^{*3}(\boldsymbol{\theta})\right] = 0, \qquad E\left[f_{t|T}^{*4}(\boldsymbol{\theta}) - 3\right] = 0,$$
$$E\left[v_{t|T}^{*3}(\boldsymbol{\theta})\right] = 0, \qquad E\left[v_{t|T}^{*4}(\boldsymbol{\theta}) - 3\right] = 0,$$

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 Propositions 3 and 6. 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)).[7] In fact, we can show that their tests and ours are asymptotically equivalent under the null hypothesis in the local level model in Section 5.3.

5.4.2 *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 $M \geq N$.

As a result, we will be able to write

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

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,

$$\mathbf{w}_t = \left[\mathbf{I}_N + \sum_{j=1}^{q_y} \mathbf{B}_j(\boldsymbol{\theta})L^j\right]^{-1} \left[\mathbf{I}_N - \sum_{j=1}^{q_y} \mathbf{A}_j(\boldsymbol{\theta})L^j\right]\left[\mathbf{y}_t - \boldsymbol{\pi}(\boldsymbol{\theta})\right]. \tag{15}$$

[7] 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)).

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 *i.i.d.* normal, then $\mathbf{y}_t$ will be a Gaussian process and, therefore, $\mathbf{w}_t|\mathcal{I}_{t-1}$ will be *i.i.d.* 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.[8] 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:

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

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 $\boldsymbol{\varphi}=\mathbf{0}$, but whose log-likelihood function and score we can obtain in closed form for every possible value of $\boldsymbol{\theta}$ when $\boldsymbol{\varphi}\neq\mathbf{0}$. 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 as

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

and their (square) Euclidean norm as

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

we can write the influence functions underling their test as

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

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 $\boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\theta})$ (see footnote 4).

[8]Although we would expect it to be closer to a normal than $\boldsymbol{\varepsilon}_t$ because of the averaging implicit in (15).

6. 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:[9]

1. the cointegrated single factor model we use in our empirical application in Section 7,
2. the illustrative local level model in Section 5.3, and
3. a multivariate version of this local level model in which there is a single integrated common trend, but the number of observed series is 10, each of which containing an *i.i.d.* idiosyncratic component.

6.1 *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*: $\boldsymbol{\varepsilon}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$ (alternative J);
2. The distribution of the innovations to the signal component is *GH* while the idiosyncratic shocks are Gaussian: $f_t \sim \mathrm{GH}(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ (alternative S_f);
3. The joint distribution of the innovations to the idiosyncratic variables is *GH* while the common component is Gaussian: $\mathbf{v}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ (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 $\boldsymbol{\beta} = -\boldsymbol{\ell}_{K\times 1}$. 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 exploiting the location-scale mixture of normal representation of the GH distribution we discussed in Section 4.1.

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 sum in expression (9) when the additional terms lead to increments lower than 10^{-5}.[10]

Given that in all the 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

[9]Results for a trivariate version of the static factor model (12) can be found in Section 4.2.2 of Almuzara, Amengual, and Sentana (2017).

[10]In Online Supplemental Material E, we report analogous results but using a HAC estimator to compute the asymptotic variances of the influence functions underlying our test statistics. As expected, the results are far less reliable than when we use the theoretical expressions.

orthogonality of the scores between mean and variance parameters on the one hand and shape parameters on the other in Propositions 3 and 6. 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).

In all the tables, the row labels H_J, H_{S_f}, and $H_{S_\mathbf{v}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f and $S_\mathbf{v}$ alternative hypotheses, while Red denotes the reduced form tests discussed in Section 5.4.2. 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.

6.2 *Small sample properties*

6.2.1 *Cointegrated dynamic factor model* We simulate data from the model (16) that we use in our empirical application in Section 7. We calibrate it to $\rho_x = 0.5$, $\rho_{\epsilon_E} = \rho_{\epsilon_I} = 0$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = q_I = 1$, where $q_i = \sigma_f^2/[(1-\rho_x^2)\sigma_{\epsilon_i}^2]$ represents the signal-to-noise ratio for y_{it} for $i = E, I$.[11]

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 7. The results make clear that the parametric bootstrap works remarkably well for both sample sizes.[12]

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 nontrivial power, especially under alternative J, they are clearly dominated by the tests aimed at the structural innovations.

6.2.2 *Univariate local level model* Table 3 contains the results for samples of size $T = 250$ of the local-level model in Section 5.3 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 their original tests.

Our results confirm the asymptotic equivalence between their tests and the less powerful two-sided versions of ours (not reported). More generally, we essentially reach the same conclusions for size and power as in the previous example.

6.2.3 *Multivariate local level model* To assess the performance of our tests when the cross-section dimension is moderately large, in Table 4 we provide results for a ten-variate model with a single common trend and uncorrelated idiosyncratic terms. Specifically, we assume $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = \boldsymbol{\ell}_{10}$ and $\boldsymbol{\gamma} = q^{-1}\boldsymbol{\ell}_{10}$, where the signal-to-noise ratio q is set

[11]We set $\rho_{\epsilon_i} = 0$ and impose it at the estimation stage to the effect of implementing the bootstrap within the Monte Carlo simulation in a reasonable amount of time.

[12]Given the number of Monte Carlo replications, the 95% asymptotic confidence intervals for the rejection probabilities under the null are (0.80, 1.20), (4.57, 5.43), and (9.41, 10.59) at the 1, 5, and 10% levels.

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_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

Note: Results based on 10,000 samples of size $T = 100$ from model (16) with $\rho_x = 0.5$, $\rho_{\epsilon_E} = \rho_{\epsilon_I} = 0$, $\sigma_f^2 = 1$, and $\sigma_{v_i}^2$ chosen such that $q_E = q_I = 1$, where $q_i = \sigma_f^2/[(1-\rho_x^2)\sigma_{\epsilon_i}^2]$ 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 $\boldsymbol{\varepsilon}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$ (i.e., $R = 3$), $f_t \sim \mathrm{GH}(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ ($R = 1$), and $\mathbf{v}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ ($R = 2$), respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in Section 5.4.2. 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 $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. 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.

to 2, as in the univariate version. We also maintain $T = 250$. Once again, we reach analogous conclusions for size and power as in the other two examples. The main difference is that rejection rates are almost 100% under S_J and $S_{\mathbf{v}}$ because the number of non-normal innovations is substantially larger than in the univariate case. Moreover, the precision with which the common factor is filtered is much higher than in the previous example because, ceteris paribus, the increase in the cross-sectional dimension N increases the observability of the latent variables. As a result, we obtain rejection rates close to the nominal ones in cases in which the maintained assumption of normality is indeed satisfied.

7. Inferring real output from GDP and GDI

7.1 *The model*

As we mentioned in the Introduction, in theory the expenditure (GDP) and income (GDI) measures of output should be equal, but they differ because they are calculated from different sources. 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

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	$S_{\mathbf{v}}$
	Kt	0.83	4.67	9.72	88.54	9.89	76.00	96.80	23.30	86.98
H_J	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
	Kt	1.07	4.81	9.79	34.44	22.74	8.27	64.40	48.53	22.74
H_{S_f}	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
	Kt	0.86	4.78	9.78	81.86	5.60	79.33	91.86	6.87	88.03
$H_{S_{\mathbf{v}}}$	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
	Kt	0.93	4.68	9.61	85.85	11.43	66.66	96.22	27.25	80.75
Red	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

Note: Results based on 10,000 samples of size $T = 250$ from model (16) with $\rho_x = 0.5$, $\rho_{\epsilon_E} = \rho_{\epsilon_I} = 0$, $\sigma_f^2 = 1$, and $\sigma_{v_i}^2$ chosen such that $q_E = q_I = 1$, where $q_i = \sigma_f^2/[(1-\rho_x^2)\sigma_{\epsilon_i}^2]$ 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 $\boldsymbol{\varepsilon}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$ (i.e., $R = 3$), $f_t \sim \mathrm{GH}(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ ($R = 1$), and $\mathbf{v}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ ($R = 2$), respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in Section 5.4.2. 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 $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. 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.

more reliable GDP growth figures by combining the two measures (see, e.g., Nalewaik (2010, 2011), Greenaway-McGrevy (2011), and especially Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016), which provides the background for the Philadelphia Fed GDPplus measure). Some national statistical offices compute a simple equally weighted average of the different aggregate series, and in fact, BEA began providing this 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 also matter because the contemporaneously filtered GDP series and its successive updates as future data becomes available will depend on the specification of the underlying processes. The secular growth in GDP and GDI has understandably led all previous studies to apply a signal-extraction framework to their growth rates, but doing so rules out by construction the possibility of saying anything about the level of U.S. output, which is of considerable interest on its own. In addition, taken literally, the absence of cointegration between the expenditure and income measures, with cointegrating vector $(1, -1)$, implies an implausible diverging statistical discrepancy. Figure 2(a) 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,

TABLE 3. 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	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
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

Note: Results based on 10,000 samples of size $T = 250$ from the local-level model discussed in Section 5.3 in which the signal-to-noise ratio $q = \sigma_f^2/\sigma_v^2$ is set to 2. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t \sim \text{GH}(\eta, \psi, \boldsymbol{\beta})$ (i.e., $R = 2$), $f_t \sim \text{GH}(\eta, \psi, \beta)$, $v_t \sim N(0, 1)$ $(R = 1)$, and $v_t \sim \text{GH}(\eta, \psi, \beta)$, $f_t \sim N(0, 1)$ $(R = 1)$, respectively. The row labels H_J, H_{S_f}, and H_{S_v} refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and S_v alternative hypotheses, Red denotes the reduced form tests discussed in Section 5.4.2, while HK denotes the original Harvey and Koopman (1992) tests discussed in Section 5.4.1. 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 $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. 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.

their $(1, -1)$ cointegration relationship is evident. In turn, Figure 2(b) shows that their first differences are also highly correlated, but with a rich dynamic bivariate structure. Finally, Figure 2(c) makes clear that the statistical discrepancy is a persistent but stationary series whose movements are unrelated to the business cycle.

In view of the previous considerations, we prefer to formulate and estimate a model with covariance stationary measurement errors and an integrated common trend in the (log) levels of the two output measures.[13] Specifically, if y_{Et} and y_{It} denote (log) GDP and

[13]Arguably, a sufficiently flexible specification for the measurement errors in first-differences by means of high-order autoregressive moving average processes may mitigate and, eventually, eliminate the consequences of ignoring cointegration, at least for the growth rates (see Almuzara, Fiorentini, and Sentana (2018)).

Table 4. Monte Carlo rejection rates (in %) under null and alternative hypotheses for the multivariate local-level 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	$S_{\mathbf{v}}$
	Kt	0.91	4.79	9.51	100.00	13.63	100.00	100.00	34.12	100.00
H_J	Sk	1.04	5.29	10.53	98.31	9.91	96.99	99.97	44.21	99.12
	GH	1.17	4.91	10.46	100.00	10.24	100.00	100.00	45.78	100.00
	Kt	0.96	5.16	10.20	67.44	64.58	4.98	84.78	86.99	8.44
H_{S_f}	Sk	1.08	5.36	9.60	28.13	58.96	5.10	86.72	95.72	5.88
	GH	0.95	5.60	10.14	62.40	59.98	5.44	90.96	95.81	7.95
	Kt	0.67	5.05	9.83	100.00	5.35	100.00	100.00	5.34	100.00
$H_{S_{\mathbf{v}}}$	Sk	0.87	5.27	10.13	97.70	4.98	97.48	99.90	5.26	99.48
	GH	0.93	5.33	10.06	100.00	4.90	100.00	100.00	5.29	100.00
	Kt	0.95	4.86	9.45	100.00	14.48	100.00	100.00	36.05	100.00
Red	Sk	1.10	5.18	10.23	98.18	10.84	96.63	99.93	46.46	98.15
	GH	1.10	4.91	10.46	100.00	11.10	100.00	100.00	48.75	100.00

Note: Results based on 10,000 samples of size $T = 250$ from a 10-variate version of the local-level model with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = \boldsymbol{\ell}_{10}$, and $\boldsymbol{\gamma} = q^{-1}\boldsymbol{\ell}_{10}$, 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 $\boldsymbol{\varepsilon}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$ (i.e., $R = 11$), $f_t \sim \mathrm{GH}(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ $(R = 1)$ and $\mathbf{v}_t \sim \mathrm{GH}(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ $(R = 10)$, respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 7 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 $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. 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.

GDI, respectively, the model that we consider is

$$
\begin{aligned}
\begin{pmatrix} y_{Et} \\ y_{It} \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} x_t + \begin{pmatrix} \epsilon_{Et} \\ \epsilon_{It} \end{pmatrix}, \\
(1 - \rho_x L)(\Delta x_t - \mu) &= f_t, \\
(1 - \rho_{\epsilon_E} L)(\epsilon_{Et} - \delta/2) &= v_{Et}, \\
(1 - \rho_{\epsilon_I} L)(\epsilon_{It} + \delta/2) &= v_{It}, \\
\begin{pmatrix} f_t \\ v_{Et} \\ v_{It} \end{pmatrix} \Bigg| \mathcal{I}_{t-1}, \boldsymbol{\phi} &\sim i.i.d.\, D\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & 0 & 0 \\ 0 & \sigma_{v_E}^2 & 0 \\ 0 & 0 & \sigma_{v_I}^2 \end{pmatrix}, \boldsymbol{\varphi}\right],
\end{aligned}
\tag{16}
$$

where x_t is the "true GDP" common factor, whose rate of growth follows an Ar(1) process with mean μ, autoregressive coefficient ρ_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 ρ_{ϵ_E} and ρ_{ϵ_I}, and innovation variances $\sigma_{v_E}^2$ and

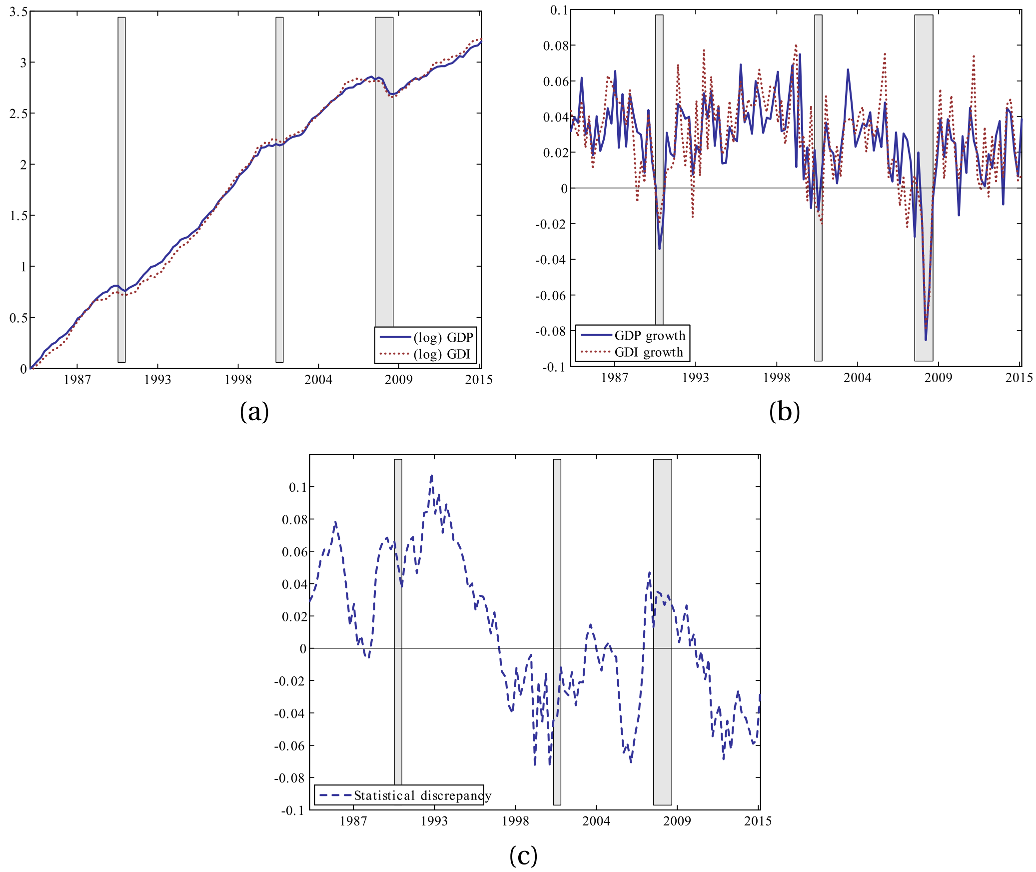


FIGURE 2. Expenditure (GDP) and income (GDI) measures of real output. (a) Quarterly real (log) GDP and (log) GDI. (b) Quarterly real GDP and GDI growth. (c) Statistical discrepancy. Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2015Q2. Statistical discrepancy is defined as $\log(\text{GDP}) - \log(\text{GDI})$. Shaded areas represent NBER recessions.

$\sigma^2_{v_I}$.[14] Our specification of the serial correlation structure of the latent series follows from the empirical analysis in earlier versions of Fiorentini and Sentana (2019), who found evidence in favor of AR(1) processes for both the first difference of the common factor and the levels of the measurement errors. Importantly, our model allows for systematic biases in the measurement errors through δ, the difference between those biases

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

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mu(1-\rho_x) & 1+\rho_x & -\rho_x & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_{\epsilon_E} & 0 \\ 0 & 0 & 0 & 0 & \rho_{\epsilon_I} \end{pmatrix}, \quad \mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 \\ \sigma_f & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_{v_E} & 0 \\ 0 & 0 & \sigma_{v_I} \end{pmatrix}$$

and $\boldsymbol{\varepsilon}_t = (f_t, v_{Et}, v_{It})'$.

determining the mean of the statistical discrepancy while their levels fixing the initial conditions.[15]

7.2 *Estimation under the null and normality tests*

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 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 stationarity-inducing 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. On the other hand, 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 we documented in Section 6.

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 3(d) and 3(f) 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

[15]For identification purposes, though, we assume without loss of generality that the magnitude of those biases is the same for the two output series. We also assume that the two measurement errors are uncorrelated, which guarantees the nonparametric identification of the signal from the noise (see Almuzara, Fiorentini, and Sentana (2018) for further details). The fact that the two measures of output are obtained from independent sources provides some plausibility to this assumption (but see Aruoba et al. (2016)).

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
ρ_x	0.536	0.105
ρ_{ϵ_E}	−0.672	0.152
ρ_{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
	Kt	0.646	0.211
H_{S_f}	Sk	1.540	0.215
	GH	2.186	0.237
	Kt	5.901	0.008
$H_{S_{\mathbf{v}}}$	Sk	7.914	0.019
	GH	13.815	0.002
	Kt	1.585	0.104
Red	Sk	1.478	0.478
	GH	3.063	0.299

Note: Data: Quarterly real GDP and GDI from 1984Q3 to 2007Q2. Model: Bivariate cointegrated, dynamic single factor model (16); see Section 7 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 7 corresponding to the S_f $(R = 1)$ and $S_{\mathbf{v}}$ $(R = 2)$ alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in Section 5.4.2. 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.

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 1 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

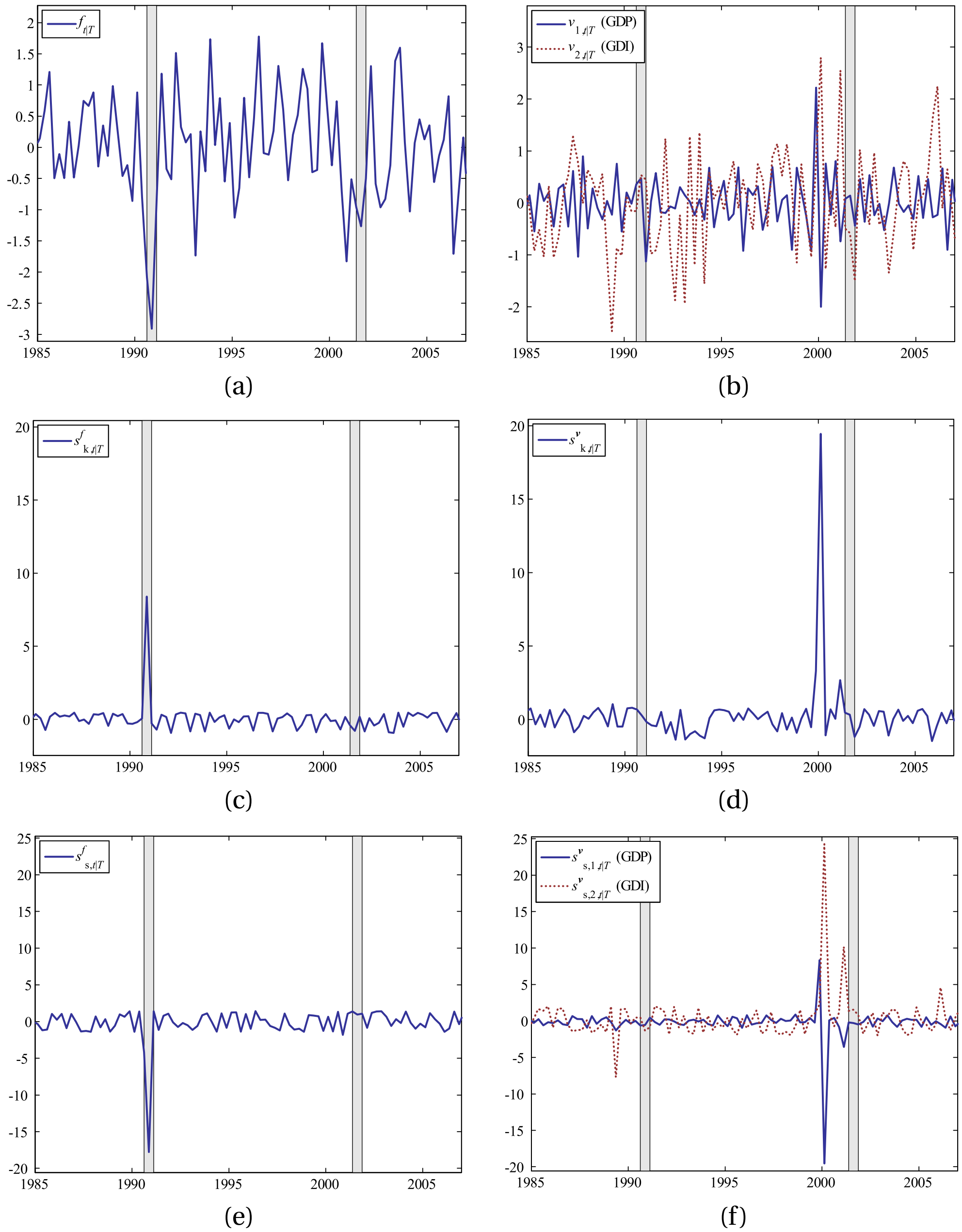


FIGURE 3. Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q3 to 2007Q2. (a) Smoothed innovations for the underlying factor. (b) Smoothed innovations for the measurement errors. (c) Influence functions for the underlying factor (kurtosis). (d) Influence functions for the measurement errors (kurtosis). (e) Influence functions for the underlying factor (skewness). (f) Influence functions for the measurement errors (skewness). Notes: Smoothed innovations and influence functions were obtained by fitting the bivariate cointegrated, dynamic single factor model (16) to the quarterly real GDP and GDI from 1984Q3 to 2007Q2; see Table 5 for parameter estimates. Shaded areas represent NBER recessions.

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
ρ_x	0.643	0.080
ρ_{ϵ_E}	−0.384	0.204
ρ_{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
$H_{S_\mathbf{v}}$	Kt	8.210	0.002
	Sk	4.398	0.111
	GH	12.607	0.004
Red	Kt	20.828	0.000
	Sk	7.818	0.020
	GH	28.645	0.000

Note: Data: Quarterly real GDP and GDI from 1984Q3 to 2015Q2. Model: Bivariate cointegrated, dynamic single factor model (16); see Section 7 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 7 corresponding to the S_f $(R = 1)$ and $S_\mathbf{v}$ $(R = 2)$ alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in Section 5.4.2. 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.

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 nonnormality. 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 1970s, early 1980s, in which the nor-

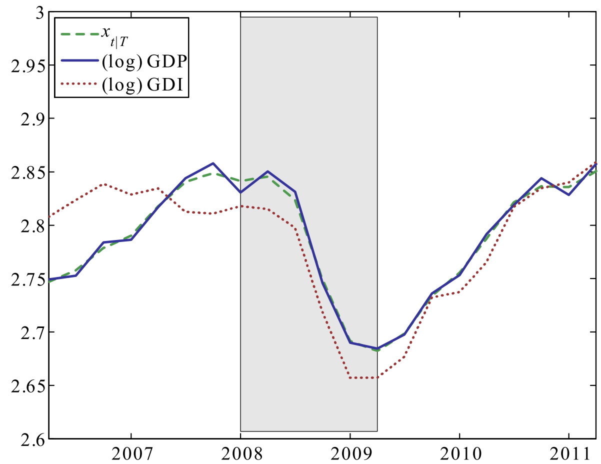


FIGURE 4. GDP, GDI, and smoothed estimate of real output around the Great Recession. Notes: The smoothed estimate $x_{t|T}$ was obtained by fitting the bivariate cointegrated, dynamic single factor model (16) to the quarterly real GDP and GDI from 1984Q3 to 2015Q2; see Table 5 for parameter estimates. The shaded area represents the NBER recession.

mality of the "true GDP" innovations is clearly rejected (see the Online Supplemental Material F for details).

7.3 *The model under the alternative*

Given those rejections, the natural next step is to estimate the parameters and obtain smoothed versions of the latent variables under the alternative distributions that we have considered. In view of the fact that the rejection of the null comes from both skewness and kurtosis, we consider an asymmetric Student t, a popular member of the asymmetric *GH* distribution, as DGP for the structural innovations. To estimate the model, we rely on a Metropolis-within-Gibbs algorithm which exploits the interpretation of the asymmetric Student t as a location-scale mixture of normals in (10). We estimate this model with 500,000 draws for the parameters and 250,000 for the latent variables, which correspond to 1 in 20 and 1 in 40 of the 10^7 original simulations (see the Online Supplemental Material D for further details on the posterior simulator).

For the sake of brevity, we focus on the shape parameters, which are reported in Figure 6, with the left and right panels corresponding to the posterior distributions for the samples 1984Q3–2007Q2 and 1984Q3–2015Q2, respectively. Interestingly, when we exclude the Great Recession from the sample, the 95% credible intervals of all the skewness parameters include the origin. In the longer sample, in contrast, the asymmetry coefficient of the latent "true GDP" series becomes statistically significantly different from zero, which is in line with the evidence obtained from our proposed score test statistics in the previous section. Similarly, there is a shift in the mode and median of the reciprocal of the degrees of freedom (top panels) toward a lower number when we use the longer sample. The results in Figure F2 in the Online Supplemental Material F confirm the agreement between our proposed tests and the posterior intervals.

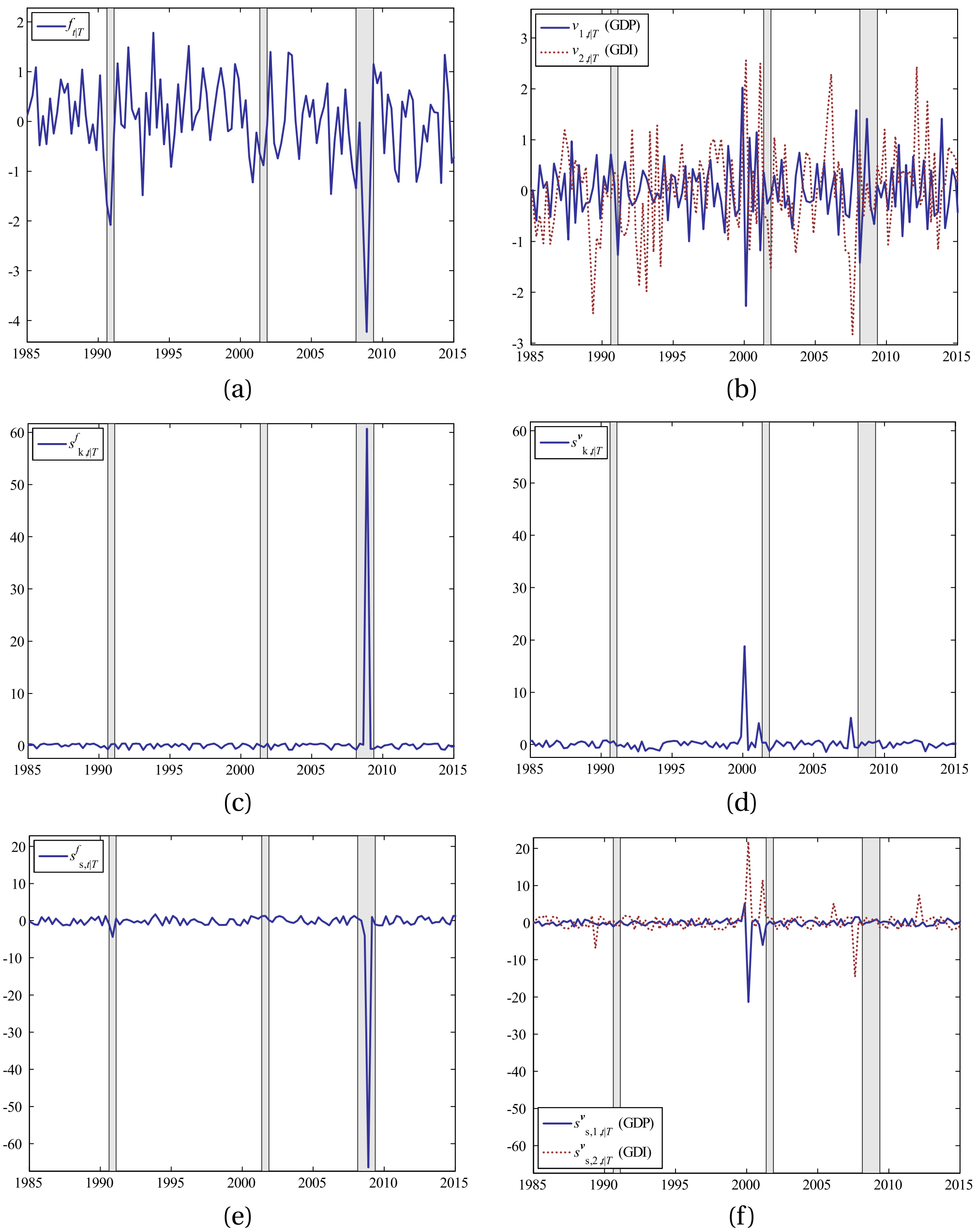


FIGURE 5. Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1984Q3 to 2015Q2. (a) Smoothed innovations for the underlying factor. (b) Smoothed innovations for the measurement errors (c) Influence functions for the underlying factor (kurtosis). (d) Influence functions for the measurement errors (kurtosis). (e) Influence functions for the underlying factor (skewness). (f) Influence functions for the measurement errors (skewness). Notes: Smoothed innovations and influence functions were obtained by fitting the bivariate cointegrated, dynamic single factor model (16) to the quarterly real GDP and GDI from 1984Q3 to 2015Q2; see Table 5 for parameter estimates. Shaded areas represent NBER recessions.

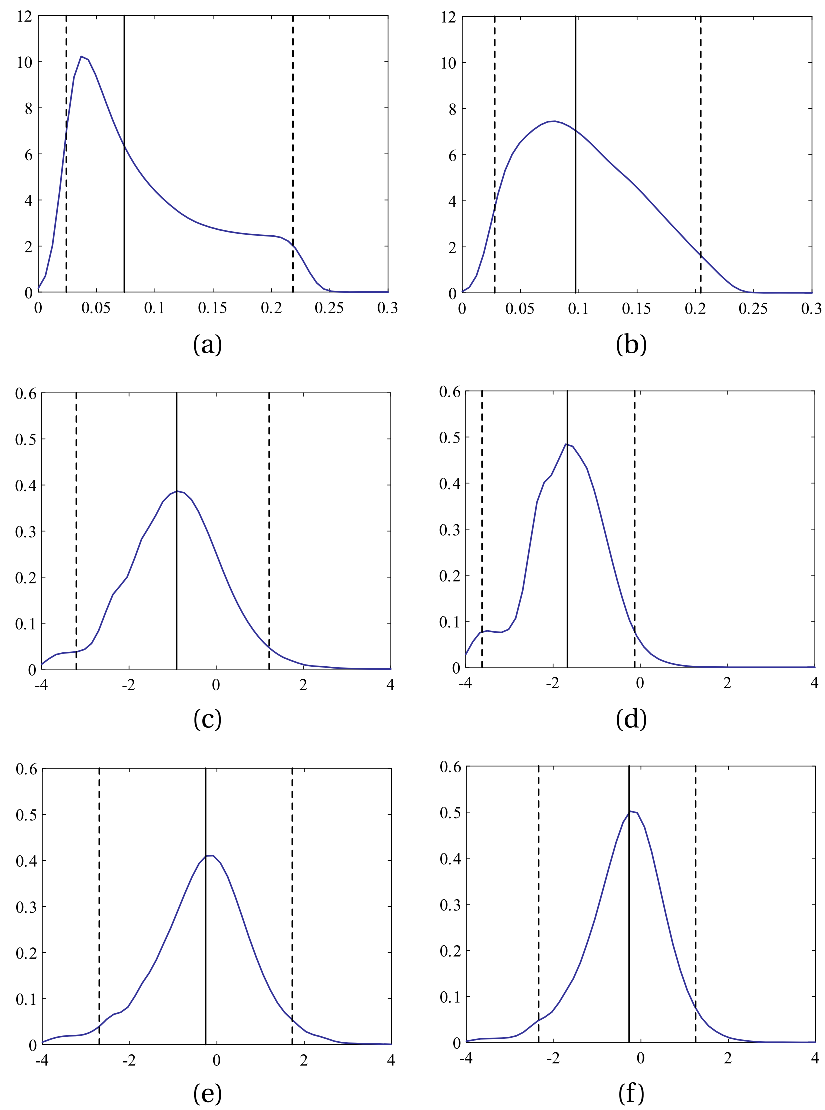


Figure 6. Posterior densities of shape parameters under the asymmetric Student t alternative. (a) η, 1984Q3 to 2007Q2. (b) η, 1984Q3 to 2015Q2. (c) β_x, 1984Q3 to 2007Q2. (d) β_x, 1984Q3 to 2015Q2. (e) β_{v_E}, 1984Q3 to 2007Q2. (f) β_{v_E}, 1984Q3 to 2015Q2. (g) β_{v_I}, 1984Q3 to 2007Q2. (h) β_{v_I}, 1984Q3 to 2015Q2. Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2007Q2 (2015Q2) in left (right) panels. Model: Bivariate cointegrated, dynamic single factor model (16) with multivariate asymmetric Student t innovations; see Section 7 for parameter definitions. η refers to the reciprocal of degrees of freedom while β_x (β_{v_E}) [β_{v_I}] refers to the skewness parameter of the "true GDP" (expenditure) [income] measure. Solid vertical lines refer to the median values while dashed lines report the 2.5% and 97.5% quantiles.

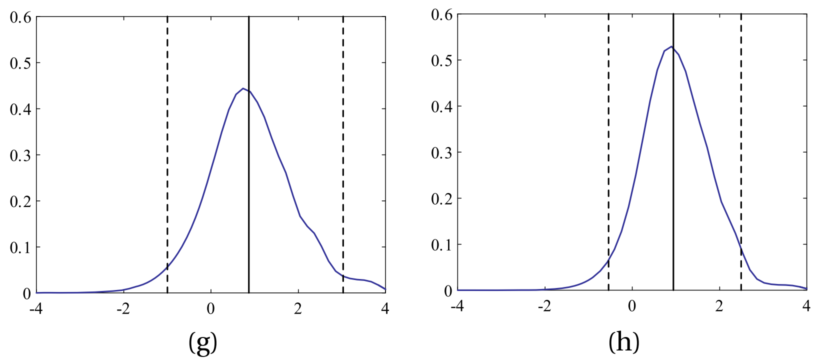


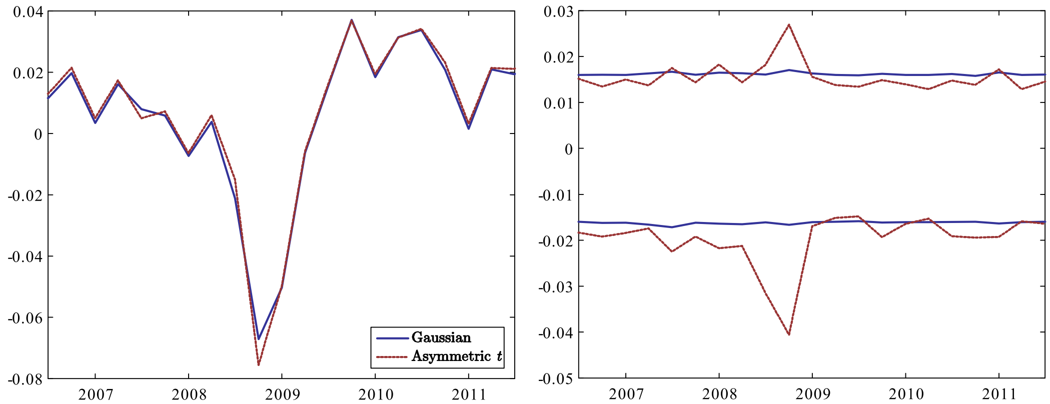
Figure 6. *Continued.*

Finally, in Figure 7 we compare $\Delta x_{t|T}$ under the null and under asymmetric t innovations. In order to account for parameter uncertainty in both models, we also estimate the Gaussian specification using a simplified version of the MCMC algorithm which imposes $\eta = 0$ but uses the same number of draws. The top panel (Figure 7(a)) reports the median of the posteriors, while the bottom one (Figure 7(b)) reports the centered 95% error bands, computed by subtracting the median from the quantiles 97.5% and 2.5%. As can be seen, the median values are quite similar across distributions, but the drop in 2008Q4 seems to be sharper under asymmetric Student t innovations. Perhaps more interestingly, while we find that the asymmetric t seems to generate narrower (wider) intervals on the right (left) of the distribution in normal times, their magnitudes increase substantially during the Great Recession, exacerbating the asymmetry of the error band too. Importantly, this pattern starts to appear—albeit moderately—a few quarters before 2008Q4. In contrast, the Gaussian error bands are symmetric and almost constant irrespective of whether the economy is in a recession or not.

8. Conclusions

We exploit the EM principle to derive simple to implement and interpret LM-type tests of normality in all or a subset of the innovations to the latent variables in state space models against *GH* 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 LR 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 previously proposed tests. For all the models that we consider, our results detect a pile-up problem, whereby 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 the reliability of our tests under the null. In terms of power, we find that the most powerful test for any given alternative is usually the score test we have designed against it. We also find that while the tests that are based on the reduced form innovations have nontrivial power, they are clearly dominated by our proposed tests, which aim at the structural innovations.



(a) (b)

Figure 7. Smoothed "true GDP" growth under Gaussian and asymmetric Student t innovations. (a) Posterior median of $\Delta x_{t|T}$. (b) Posterior 95% error bands for $\Delta x_{t|T}$. Notes: Data: Quarterly real GDP and GDI from 1984Q3 to 2015Q2. Model: Bivariate cointegrated, dynamic single factor model (16) with multivariate asymmetric Student t innovations; see Section 7 for parameter definitions. Results are based on 25,000 draws from the posterior simulator. Error bands refer to the 2.5% and 97.5% quantiles from which the median values where subtracted.

When we apply our tests to a common trend model which combines the levels of 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 nonnormal regardless of the period. For that reason, we develop a nonlinear, simulation-based filtering procedure that improves over the Kalman filter, and highlights the importance of taking nonnormality into account during turbulent periods such as the Great Recession.

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, Monfort, Renault, and Trognon (1987). While the linearity implicit in (1)–(2) helps us obtain closed-form expressions for all the relevant quantities, it is not a requirement for applying our methodology in different contexts. Analyzing other latent variable models in which non-Gaussianity might be relevant constitutes a very interesting avenue for future research.

References

Almuzara, M., D. Amengual, and E. Sentana (2019), "Supplement to 'Normality tests for latent variables'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE859. [984]

Almuzara, M., D. Amengual, and E. Sentana (2017), "Normality tests for latent variables." CEMFI Working Paper No. 1708. [986, 999]

Almuzara, M., G. Fiorentini, and E. Sentana (2018), "US aggregate output measurement: A common trend approach." Report. [1003, 1006]

Amengual, D. and E. Sentana (2015), "Is a normal copula the right copula?" CEPR Discussion Paper 10809. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=10809. [984, 1000]

Anderson, B. D. O. and J. B. Moore (1979), *Optimal Filtering*. Prentice-Hall, NJ. [985]

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. [1002, 1006]

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. [983]

Barndorff-Nielsen, O. (1977), "Exponentially decreasing distributions for the logarithm of particle size." *Proceedings of the Royal Society*, 353, 401–419. [985]

Blæsild, P. (1981), "The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data." *Biometrika*, 68, 251–263. [985, 990]

Bontemps, C. and N. Meddahi (2005), "Testing normality: A GMM approach." *Journal of Econometrics*, 124, 149–186. [983, 991, 997]

Chesher, A. and M. Irish (1987), "Residual analysis in the grouped data and censored normal linear model." *Journal of Econometrics*, 34, 33–62. [1014]

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. [983]

Durbin, J. and S. J. Koopman (2012), *Time Series Analysis by State Space Methods*, second edition. Oxford University Press, Oxford. [983, 985]

Fiorentini, G., A. Galesi, and E. Sentana (2016), "A spectral EM algorithm for dynamic factor models." *Journal of Econometrics*, 205, 249–279. [985]

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. [987, 989]

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. [997]

Fiorentini, G. and E. Sentana (2007), "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models." CEMFI Working Paper 0713. [991]

Fiorentini, G. and E. Sentana (2015), "Tests for serial dependence in static, non-Gaussian factor models." In *Unobserved Components and Time Series Econometrics* (S. J. Koopman and N. Shephard, eds.), 118–189, Oxford University Press. [983]

Fiorentini, G. and E. Sentana (2019), "Dynamic specification test for dynamic factor models." *Journal of Applied Econometrics*, 34, 325–346. [1005]

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. [989]

Gouriéroux, C., A. Monfort, E. Renault, and A. Trognon (1987), "Generalized residuals." *Journal of Econometrics*, 34, 5–32. [1014]

Greenaway-McGrevy, R. (2011), "Is GDP or GDI a better measure of output? A statistical approach." Bureau of Economic Analysis WP, 2011–08. [1002]

Grimm, B. T. (2007), "The statistical discrepancy." Bureau of Economic Analysis, Washington D.C. [1001]

Harvey, A. C. (1989), *Forecasting, Structural Models and the Kalman Filter*. Cambridge University Press, Cambridge. [985, 988]

Harvey, A. C. and S. J. Koopman (1992), "Diagnostic checking of unobserved-components time series models." *Journal of Business and Economic Statistics*, 10, 377–389. [983, 995, 996, 997, 1000, 1003]

Horowitz, J. and N. E. Savin (2000), "Empirically relevant critical values for hypothesis tests: A bootstrap approach." *Journal of Econometrics*, 95, 375–389. [999]

Jarque, C. M. and A. Bera (1980), "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals." *Economics Letters*, 6, 255–259. [984, 995, 997]

Johannes, M. and N. Polson (2009), "Particle filtering." In *Handbook of Financial Time Series* (T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, eds.), 1015–1029, Springer-Verlag. [982]

Johnson, N., S. Kotz, and N. Balakrishnan (1994), *Continuous Univariate Distributions*. Wiley, New York, NY. [990]

Jørgensen, B. (1982), *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer-Verlag, New York, NY. [990]

Komunjer, I. and S. Ng (2011), "Dynamic identification of dynamic stochastic general equilibrium models." *Econometrica*, 79, 1995–2032. [985]

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. [982]

Lomnicki, Z. (1961), "Tests for departure from normality in the case of linear stochastic processes." *Metrika*, 4, 37–62. [997]

Louis, T. A. (1982), "Finding the observed information matrix when using the EM algorithm." *Journal of the Royal Statistical Society Series B*, 44, 226–233. [983, 986, 987, 990]

Mencía, J. (2012), "Testing nonlinear dependence in the hedge fund industry." *Journal of Financial Econometrics*, 10, 545–587. [982]

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. [984, 990, 992, 995, 998]

Nalewaik, J. (2010), "The income- and expenditure-side measures of output growth." *Brookings Papers on Economic Activity*, 1, 71–106. [1002, 1006]

Nalewaik, J. (2011), "The income- and expenditure-side measures of output growth—An update through 2011Q2." *Brookings Papers on Economic Activity*, 2, 385–402. [1002]

Pagan, A. and F. Vella (1989), "Diagnostic tests for models based on individual data: A survey." *Journal of Applied Econometrics*, 4, S29–S59. [1014]

Rubin, D. and D. Thayer (1982), "EM algorithms for ML factor analysis." *Psychometrika*, 47, 69–76. [983]

Ruud, P. (1991), "Extensions of estimation methods using the EM algorithm." *Journal of Econometrics*, 49, 305–341. [986]

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. [1002]

Stone, R., D. G. Champernowne, and J. E. Meade (1942), "The precision of national income estimates." *Review of Economic Studies*, 9, 111–125. [1002]

Tanner, M. (1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, third edition. Springer-Verlag. [986]

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. [983]

Weale, M. (1992), "Estimation of data measured with error and subject to linear restrictions." *Journal of Applied Econometrics*, 7, 167–174. [1002]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 23 March, 2017; final version accepted 1 November, 2018; available online 21 November, 2018.