Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Moment tests of independent components	
Article Sub-Title		
Article CopyRight	The Author(s) (This will be the copyright line in the final PDF)	
Journal Name	SERIEs	

Corresponding Author	Family Name	**Sentana**
	Particle	
	Given Name	**Enrique**
	Suffix	
	Division	
	Organization	CEMFI
	Address	Casado del Alisal 5, 28014, Madrid, Spain
	Phone	
	Fax	
	Email	sentana@cemfi.es
	URL	
	ORCID	http://orcid.org/0000-0003-2328-909X
Author	Family Name	**Amengual**
	Particle	
	Given Name	**Dante**
	Suffix	
	Division	
	Organization	CEMFI
	Address	Casado del Alisal 5, 28014, Madrid, Spain
	Phone	
	Fax	
	Email	amengual@cemfi.es
	URL	
	ORCID	
Author	Family Name	**Fiorentini**
	Particle	
	Given Name	**Gabriele**
	Suffix	
	Division	
	Organization	Università di Firenze and RCEA
	Address	Viale Morgagni 59, 50134, Florence, Italy
	Phone	
	Fax	
	Email	gabriele.fiorentini@unifi.it
	URL	

ORCID

Schedule	Received	18 February 2021
	Revised	
	Accepted	22 September 2021

Abstract	We propose simple specification tests for independent component analysis and structural vector autoregressions with non-Gaussian shocks that check the normality of a single shock and the potential cross-sectional dependence among several of them. Our tests compare the integer (product) moments of the shocks in the sample with their population counterparts. Importantly, we explicitly consider the sampling variability resulting from using shocks computed with consistent parameter estimators. We study the finite sample size of our tests in several simulation exercises and discuss some bootstrap procedures. We also show that our tests have non-negligible power against a variety of empirically plausible alternatives.

Keywords (separated by '-')	Covariance - Co-skewness - Co-kurtosis - Finite normal mixtures - Normality tests - Pseudo-maximum likelihood estimators - Structural vector autoregressions
JEL Classification (separated by '-')	C32 - C46 - C52
Footnote Information	We would like to thank Giuseppe Cavaliere and Aapo Hyvärinen, as well as participants at ESEM 21 for some useful comments and suggestions. Two anonymous referees have also provided very valuable feedback. Of course, the usual caveat applies. The first and third authors acknowledge financial support from the Spanish Ministry of Economy, Industry & Competitiveness through Grant ECO 2017-89689 and the Santander CEMFI Research Chair.

SERIEs
https://doi.org/10.1007/s13209-021-00247-3

SERIEs Journal of the Spanish Economic Association

ORIGINAL ARTICLE

Moment tests of independent components

Dante Amengual[1] · Gabriele Fiorentini[2] · Enrique Sentana[1]

Received: 18 February 2021 / Accepted: 22 September 2021

Abstract
We propose simple specification tests for independent component analysis and structural vector autoregressions with non-Gaussian shocks that check the normality of a single shock and the potential cross-sectional dependence among several of them. Our tests compare the integer (product) moments of the shocks in the sample with their population counterparts. Importantly, we explicitly consider the sampling variability resulting from using shocks computed with consistent parameter estimators. We study the finite sample size of our tests in several simulation exercises and discuss some bootstrap procedures. We also show that our tests have non-negligible power against a variety of empirically plausible alternatives.

Keywords Covariance · Co-skewness · Co-kurtosis · Finite normal mixtures · Normality tests · Pseudo-maximum likelihood estimators · Structural vector autoregressions

JEL Classification C32 · C46 · C52

We would like to thank Giuseppe Cavaliere and Aapo Hyvärinen, as well as participants at ESEM 21 for some useful comments and suggestions. Two anonymous referees have also provided very valuable feedback. Of course, the usual caveat applies. The first and third authors acknowledge financial support from the Spanish Ministry of Economy, Industry & Competitiveness through Grant ECO 2017-89689 and the Santander CEMFI Research Chair.

✉ Enrique Sentana
sentana@cemfi.es

Dante Amengual
amengual@cemfi.es

Gabriele Fiorentini
gabriele.fiorentini@unifi.it

1 CEMFI, Casado del Alisal 5, 28014 Madrid, Spain

2 Università di Firenze and RCEA, Viale Morgagni 59, 50134 Florence, Italy

1 Introduction

The literature on structural vector autoregressions (SVAR) is vast. Popular identification schemes include short- and long-run homogenous restrictions [see, e.g. Sims (1980), Blanchard and Quah (1989)], sign restrictions [see, e.g. Faust (1998), Uhlig (2005)], time-varying heteroskedasticity (Sentana and Fiorentini 2001) or external instruments [see, e.g. Mertens and Ravn (2012), Stock and Watson (2018) or Dolado et al. (2020)]. Recently, identification through independent non-Gaussian shocks has become increasingly popular after Lanne et al. (2017) and Gouriéroux et al. (2017). The signal processing literature on independent component analysis (ICA) popularised by Comon (1994) shares the same identification scheme. Specifically, if in a static model the $N \times 1$ observed random vector $\boldsymbol{y}$—the so-called signals or sensors—is the result of an affine combination of N unobserved shocks $\boldsymbol{\varepsilon}^*$—the so-called components or sources—whose mean and variance we can set to $\mathbf{0}$ and $\boldsymbol{I}_N$ without loss of generality, namely

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{\varepsilon}^*, \tag{1}$$

then the matrix $\boldsymbol{C}$ of loadings of the observed variables on the latent ones can be identified (up to column permutations and sign changes) from an $i.i.d.$ sample of observations on $\boldsymbol{y}$ provided the following assumption holds:[1]

Assumption 1 :Identification

(1) the N shocks in (1) are cross-sectionally independent,
(2) at least $N-1$ of them follow a non-Gaussian distribution, and
(3) $\boldsymbol{C}$ is invertible.

Failure of any of the three conditions in Assumption 1 results in an underidentified model. The best known counterexample is a multivariate Gaussian model for $\boldsymbol{\varepsilon}^*$, in which we can identify $V(\boldsymbol{y}) = \boldsymbol{C}\boldsymbol{C}'$ but not $\boldsymbol{C}$ without additional structural restrictions despite the fact that the elements of $\boldsymbol{\varepsilon}^*$ are cross-sectionally independent. Intuitively, the problem is that any rotation of the structural shocks $\boldsymbol{\varepsilon}^{**} = \boldsymbol{Q}\boldsymbol{\varepsilon}^*$, where $\boldsymbol{Q}$ is an orthogonal matrix, generates another set of N observationally equivalent, cross-sectionally independent shocks with standard normal marginal distributions. A less well-known counterexample would be a non-Gaussian spherical distribution for $\boldsymbol{\varepsilon}^*$, such as the standardised multivariate Student t. In this case, the lack of identifiability of $\boldsymbol{C}$ is due to the fact that $\boldsymbol{\varepsilon}^*$ and $\boldsymbol{\varepsilon}^{**}$ share not only their mean vector ($\mathbf{0}$) and covariance matrix ($\boldsymbol{I}$), but also the same nonlinear dependence structure.

The purpose of our paper is to propose simple to implement and interpret specification tests that check the normality of a single element of $\boldsymbol{\varepsilon}^*$ and the potential cross-sectional dependence among several of them. In very simple terms, our tests compare the integer (product) moments of the shocks in the sample with their population counterparts. Specifically, in the Gaussian tests we compare the marginal third and fourth moments of a single shock to 0 and 3, respectively. In turn, in the case

[1] The same result applies to situations in which $\dim(\boldsymbol{\varepsilon}^*) \leq \dim(\boldsymbol{y})$ provided that $\boldsymbol{C}$ has full column rank.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

of two or more shocks, we assess the statistical significance of their second, third and fourth cross-moments, which should be equal to the product of the corresponding marginal moments under independence. Many of these moments tests can be formally justified as Lagrange multiplier tests against specific parametric alternatives [see, e.g. Mencía and Sentana (2012)], but in this paper we do not pursue this interpretation. Like Almuzara et al. (2019), though, we focus on the latent shocks rather than the observed variables in view of the fact that identifying Assumption 1 is written in terms of $\boldsymbol{\varepsilon}^*$ rather than y.

If we knew the true values of $\boldsymbol{\mu}$ and $\boldsymbol{C}$, $\boldsymbol{\mu}_0$ and $\boldsymbol{C}_0$ say, with $rank(\boldsymbol{C}_0) = N$, our tests would be straightforward, as we could trivially recover the latent shocks from the observed signals without error. In practice, though, both $\boldsymbol{\mu}$ and $\boldsymbol{C}$ are unknown, so we need to estimate them before computing our tests.

Although many estimation procedures for those parameters have been proposed in the literature [see, e.g. Moneta and Pallante (2020) and the references therein], in this paper we consider the discrete mixtures of normals-based pseudo-maximum likelihood estimators (PMLEs) in Fiorentini and Sentana (2020) for three main reasons. First, they are consistent for the model parameters under standard regularity conditions provided that Assumption 1 holds regardless of the true marginal distributions of the shocks. Second, they seem to be rather efficient, the rationale being that finite normal mixtures can provide good approximations to many univariate distributions. And third, the influence functions on which they are based are the scores of the pseudo-log-likelihood, which we can easily compute in closed form. As we shall see, these influence functions play a very important role in adjusting the asymptotic variances of the different tests we propose so that they reflect the sampling variability resulting from computing the shocks with consistent but noisy parameter estimators.

In this respect, we derive computationally simple closed-form expressions for the asymptotic covariance matrices of the sample moments underlying our tests under the relevant null adjusted for parameter uncertainty. Importantly, we do so not only for static ICA model (1) but also for a SVAR, which is far more relevant in economics.

In many empirical finance applications of SVARs, the number of observations is sufficiently large for asymptotic approximations to be reliable. In contrast, the limiting distributions of our tests may be a poor guide for the smaller samples typically used in macroeconomic applications. For that reason, we thoroughly study the finite sample size of our tests in several Monte Carlo exercises. We also discuss some bootstrap procedures that seem to improve their reliability. Finally, we show that our tests have non-negligible power against a variety of empirically plausible alternatives in which the cross-sectional independence of the shocks no longer holds.

The rest of the paper is organised as follows. Section 2 discusses the model and the estimation procedure. Then, we present our general moment tests in Sect. 3 and particularise them to assess normality and independence in Sect. 4. Next, Sect. 5 contains the results of our Monte Carlo experiments. We present our conclusions and suggestions for further research in Sect. 6 and relegate some technical material and additional simulations to several appendices.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

2 Structural vector autoregressions

2.1 Model specification

Consider the following N-variate SVAR process of order p:

$$y_t = \boldsymbol{\tau} + \sum_{j=1}^{p} \boldsymbol{A}_j \boldsymbol{y}_{t-j} + \boldsymbol{C}\boldsymbol{\varepsilon}_t^*, \quad \boldsymbol{\varepsilon}_t^* | I_{t-1} \sim i.i.d.\ (\boldsymbol{0}, \boldsymbol{I}_N), \tag{2}$$

where I_{t-1} is the information set, $\boldsymbol{C}$ the matrix of impact multipliers and $\boldsymbol{\varepsilon}_t^*$ the " structural" shocks, which are normalised to have zero means, unit variances and zero covariances.

Let $\boldsymbol{\varepsilon}_t = \boldsymbol{C}\boldsymbol{\varepsilon}_t^*$ denote the reduced form innovations, so that $\boldsymbol{\varepsilon}_t | I_{t-1} \sim i.i.d.\ (\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}'$. As we mentioned in introduction, a Gaussian (pseudo) log-likelihood is only able to identify $\boldsymbol{\Sigma}$, which means the structural shocks $\boldsymbol{\varepsilon}_t^*$ and their loadings in $\boldsymbol{C}$ are only identified up to an orthogonal transformation. Specifically, we can use the so-called LQ matrix decomposition[2] to relate the matrix $\boldsymbol{C}$ to the Cholesky decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_L \boldsymbol{\Sigma}_L'$ as

$$\boldsymbol{C} = \boldsymbol{\Sigma}_L \boldsymbol{Q}, \tag{3}$$

where $\boldsymbol{Q}$ is an $N \times N$ orthogonal matrix, which we can model as a function of $N(N-1)/2$ parameters $\boldsymbol{\omega}$ by assuming that $|\boldsymbol{Q}| = 1$.[3] Notice that if $|\boldsymbol{Q}| = -1$ instead, we can change the sign of the i^{th} structural shock and its impact multipliers in the i^{th} column of the matrix $\boldsymbol{C}$ without loss of generality as long as we also modify the shape parameters of the distribution of ε_{it}^* to alter the sign of all its nonzero odd moments.

In this context, Lanne et al. (2017) show that statistical identification of both the structural shocks and $\boldsymbol{C}$ (up to column permutations and sign changes) is possible under ICA identification Assumption 1, which we maintain in what follows. Popular examples of univariate non-normal distributions are the Student t and the generalised error (or Gaussian) distribution, which includes normal, Laplace and uniform as special cases, as well as symmetric and asymmetric finite normal mixtures.

2.2 Pseudo-maximum likelihood estimators

2.2.1 The criterion function

Let $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\boldsymbol{A}_1), \ldots, vec'(\boldsymbol{A}_p), vec'(\boldsymbol{C})]' = (\boldsymbol{\tau}', \boldsymbol{a}_1', \ldots, \boldsymbol{a}_p', \boldsymbol{c}') = (\boldsymbol{\tau}', \boldsymbol{a}', \boldsymbol{c}')$ denote the structural parameters characterising the first two conditional moments of

[2] The LQ decomposition is intimately related to the QR decomposition. Specifically, $\boldsymbol{Q}'\boldsymbol{\Sigma}_L'$ provides the QR decomposition of the matrix $\boldsymbol{C}'$, which is uniquely defined if we restrict the diagonal elements of $\boldsymbol{\Sigma}_L$ to be positive [see, e.g. Golub and van Loan (2013) for further details].

[3] See section 10 of Magnus et al. (2021) for a detailed discussion of three ways of explicitly parametrising a rotation (or special orthogonal) matrix: (i) as the product of Givens matrices that depend on $N(N-1)/2$ Tait-Bryan angles, one for each of the strict upper diagonal elements; (ii) by using the so-called Cayley transform of a skew-symmetric matrix; and (c) by exponentiating a skew-symmetric matrix.

$\boldsymbol{y}_t$. In addition, we assume $\varepsilon^*_{it}|I_{t-1} \sim i.i.d.\ D(0, 1, \boldsymbol{\varrho}_i)$, where $\boldsymbol{\varrho}_i$ is a $q_i \times 1$ vector of variation-free shape parameters, so that in principle different shocks could follow different distributions. For simplicity of notation, though, we maintain that the univariate distributions of the shocks belong to the same family. We can then collect all the shape parameters in the $q \times 1$ vector $\boldsymbol{\varrho} = (\boldsymbol{\varrho}'_1, \ldots, \boldsymbol{\varrho}'_N)'$, with $q = \sum_{i=1}^{N} q_i$, so that $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho}')'$ is the $[N + (p+1)N^2 + q] \times 1$ vector containing all the model parameters.

Given the linear mapping between structural shocks and reduced form innovations, the contribution to the conditional log-likelihood function from observation $\boldsymbol{y}_t$ ($t = 1, \ldots, T$) for those parameter configurations for which $\boldsymbol{C}$ has full rank will be given by

$$
\begin{aligned}
l(\boldsymbol{y}_t; \boldsymbol{\phi}) &= -\ln|\boldsymbol{C}| + \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}] = -\ln|\boldsymbol{C}| + \ln f[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \cdots \\
&\quad + \ln f[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N] = l_t(\boldsymbol{\phi}),
\end{aligned} \tag{4}
$$

where $f[\varepsilon^*_{it}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]$ is the univariate log-likelihood function for the i^{th} structural shock, $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}) = \boldsymbol{C}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \boldsymbol{y}_t - \boldsymbol{\tau} - \boldsymbol{A}_1\boldsymbol{y}_{t-1} - \cdots - \boldsymbol{A}_p\boldsymbol{y}_{t-p}$ are the reduced-form innovations.

2.2.2 The score vector

Let $\boldsymbol{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ and partition it into two blocks, $\boldsymbol{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $\boldsymbol{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$, respectively. Fiorentini and Sentana (2021) show that the scores can be written as

$$
\boldsymbol{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = [\boldsymbol{Z}_{lt}(\boldsymbol{\theta}), \boldsymbol{Z}_{st}(\boldsymbol{\theta})]\begin{bmatrix}\boldsymbol{e}_{lt}(\boldsymbol{\phi}) \\ \boldsymbol{e}_{st}(\boldsymbol{\phi})\end{bmatrix} = \boldsymbol{Z}_{dt}(\boldsymbol{\theta})\boldsymbol{e}_{dt}(\boldsymbol{\phi}), \tag{5}
$$

$$
\boldsymbol{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \boldsymbol{e}_{rt}(\boldsymbol{\phi}), \tag{6}
$$

where

$$
\boldsymbol{Z}_{lt}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_N \\ \boldsymbol{y}_{t-1} \otimes \boldsymbol{I}_N \\ \vdots \\ \boldsymbol{y}_{t-p} \otimes \boldsymbol{I}_N \\ \boldsymbol{0}_{N^2 \times N} \end{pmatrix} \boldsymbol{C}^{-1\prime}, \tag{7}
$$

$$
\boldsymbol{Z}_{st}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{0}_{N \times N^2} \\ \boldsymbol{0}_{N^2 \times N^2} \\ \vdots \\ \boldsymbol{0}_{N^2 \times N^2} \\ \boldsymbol{I}_{N^2} \end{pmatrix} (\boldsymbol{I}_N \otimes \boldsymbol{C}^{-1\prime}), \tag{8}
$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$$\boldsymbol{e}_{lt}(\boldsymbol{\phi}) = -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} = -\left\{ \begin{array}{c} \frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \\ \vdots \\ \frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \end{array} \right\}, \tag{9}$$

$$\begin{aligned} \boldsymbol{e}_{st}(\boldsymbol{\phi}) &= -vec\left\{ \boldsymbol{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \right\} \\ &= -vec\left\{ \begin{array}{ccc} 1 + \frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \varepsilon_{1t}^*(\boldsymbol{\theta}) & \ldots & \frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \varepsilon_{1t}^*(\boldsymbol{\theta}) & \ldots & 1 + \frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array} \right\} \end{aligned} \tag{10}$$

and

$$\boldsymbol{e}_{rt}(\boldsymbol{\phi}) = \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}} = \left\{ \begin{array}{c} \frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array} \right\} = \left\{ \begin{array}{c} \boldsymbol{e}_{r_1 t}(\boldsymbol{\phi}) \\ \boldsymbol{e}_{r_2 t}(\boldsymbol{\phi}) \\ \vdots \\ \boldsymbol{e}_{r_N t}(\boldsymbol{\phi}) \end{array} \right\} \tag{11}$$

by virtue of the cross-sectional independence of the shocks, so that the derivatives involved correspond to the underlying univariate densities.

2.2.3 The asymptotic distribution

For simplicity, we assume henceforth that SVAR model (2) generates a covariance stationary process.[4] Consider the reparametrisation $\boldsymbol{C} = \boldsymbol{J}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix whose elements contain the scale of the structural shocks, while the columns of $\boldsymbol{J}$, whose diagonal elements are normalised to 1, measure the relative impact of each of the structural shocks on all the remaining variables. Proposition 3 in Fiorentini and Sentana (2020) shows that the parameters $\boldsymbol{a}_i = vec(\boldsymbol{A}_i)$ and $\boldsymbol{j} = veco(\boldsymbol{J})$ are consistently estimated regardless of the true distribution.[5] As a result, the pseudo-true values of those parameters will coincide with the true ones, i.e. $\boldsymbol{a}_{i\infty} = \boldsymbol{a}_{i0}$ and $\boldsymbol{j}_\infty = \boldsymbol{j}_0$. In contrast, $\boldsymbol{\tau}$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$ will generally be inconsistently estimated, so $\boldsymbol{\tau}_\infty \neq \boldsymbol{\tau}_0$ and $\boldsymbol{\psi}_\infty \neq \boldsymbol{\psi}_0$.

Nevertheless, Fiorentini and Sentana (2020) prove that the unrestricted PMLEs of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ which simultaneously estimate $\boldsymbol{\varrho}$ will be consistent too when the univariate distributions used for estimation purposes are discrete mixtures of normals, in which

4 If the autoregressive polynomial $(\boldsymbol{I}_N - \boldsymbol{A}_1 L - \ldots - \boldsymbol{A}_p L^p)$ had some unit roots, $\boldsymbol{y}_t$ would be a (co-) integrated process, and the estimators of the conditional mean parameters would have non-standard asymptotic distributions, as some of them would converge at the faster rate T. In contrast, the distribution of the ML estimators of the conditional variance parameters would remain standard [see, e.g. Phillips and Durlauf (1986)].

5 See Magnus and Sentana (2020) for some useful properties of the $veco(.)$ and $vecd(.)$ operators.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

case $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ and $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}_t^*$. For that reason, in what follows we focus on the finite normal mixtures-based PMLEs of the original parameters $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{a}', \boldsymbol{c}')$.

Still, the potential misspecification of this distributional assumption implies that the asymptotic covariance matrix of the corresponding PMLEs must be based on the usual sandwich formula. Let

$$\mathcal{A}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\partial \boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)/\partial \boldsymbol{\phi}')|\boldsymbol{\varphi}_0] \tag{12}$$

and

$$\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V[\boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] \tag{13}$$

denote the (-) expected value of the log-likelihood Hessian and the variance of the score, respectively, where $\boldsymbol{\varrho}_\infty$ are the pseudo-true values of the shape parameters of the distributions of the shocks assumed for estimation purposes, $\boldsymbol{\upsilon}$ contains the potentially infinite-dimensional shape parameters of the true distributions of the shocks, and $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \boldsymbol{\upsilon})$. The asymptotic distribution of the pseudo-ML estimators of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$, under standard regularity conditions will be given by

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_\infty) \to N[\mathbf{0}, \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)].$$

In what follows, we shall make extensive use of the detailed expressions for the conditional expected value of the Hessian and covariance matrix of the score for finite normal mixtures-based PMLEs in Amengual et al. (2021b).

3 Specification tests based on integer product moments

3.1 The influence functions

As we have stressed earlier, the parametric identification of the structural shocks $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ and their impact coefficients $\boldsymbol{C}$ that appear in SVAR (2) critically hinges on the validity of identifying Assumption 1. As a consequence, it would be desirable that empirical researchers estimating those models reported specification tests that would check those assumptions. Given that rank failures in $\boldsymbol{C}$ will be inextricably linked to singular dynamic systems,[6] we focus on testing that at most one of the structural shocks is Gaussian and that all the structural shocks are indeed independent of each other.

As is well known, stochastic independence between the elements of a random vector is equivalent to the joint distribution being the product of the marginal ones. In turn, this factorisation implies lack of correlation between not only the levels but also any set of single-variable measurable transformations of those elements. Thus, a rather

[6] The rationale is as follows. If $rank(\boldsymbol{C}_0) < N$, then $rank[V(\boldsymbol{y}_t)] < N$, and the same will be true of the sample covariance matrix. Therefore, sampling variability plays no role in determining whether $rank(\boldsymbol{C}_0) = N$ in model (1). Exactly the same argument applies to dynamic system (2).

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

intuitive way of testing for independence without considering any specific parametric alternative can be based on individual moment conditions of the form

$$m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = \prod_{i=1}^{N} \varepsilon_{it}^{*h_i}(\boldsymbol{\theta}) - \prod_{i=1}^{N} E[\varepsilon_{it}^{*h_i}(\boldsymbol{\theta}_0)], \tag{14}$$

where $\boldsymbol{h} = \{h_1, ..., h_N\}$, with $h_i \in \mathbb{Z}_{0+}$, denotes the index vector characterising a specific product moment. While the influence function in (14) will generally require the estimation of $E[\varepsilon_{it}^{*h_i}(\boldsymbol{\theta}_0)]$ for some of the shocks, the constant term $\prod_{i=1}^{N} E[\varepsilon_{it}^{*h_i}(\boldsymbol{\theta}_0)]$ is either 0 or 1 for the second, third and fourth cross-moments we study in this paper in view of the standardised nature of the shocks, so we do not need to worry about it. Amengual et al. (2021b) discuss in detail how to deal with the estimation of the required $E[\varepsilon_{it}^{*h_i}(\boldsymbol{\theta}_0)]$ in the general case.

Although we have motivated (14) as the basis for our tests of independence, by setting all the elements of $\boldsymbol{h}$ but one to 0, we can also use this expression to look at the marginal moments of a single shock. In this paper, we focus on $h_i = 3$ and 4 because most common departures from normality of the shocks will be reflected in coefficients of skewness or kurtosis different from 0 and 3, respectively.

3.2 The moment tests

Let $\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ denote a $K \times 1$ vector containing a collection of influence functions $m_{\boldsymbol{h}^k}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ of form (14) for different index vectors $\boldsymbol{h}^1, \ldots, \boldsymbol{h}^k, , \ldots, \boldsymbol{h}^K$. The following result, which specialises the general expressions in Newey (1985) and Tauchen (1985) to our context, derives the asymptotic distribution of the scaled sample average of $\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ when we evaluate the structural shocks at the PMLEs $\hat{\boldsymbol{\theta}}_T$ rather than at $\boldsymbol{\theta}_0$:

Proposition 1 *Under Assumption 1 and standard regularity conditions*

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T} \boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)] \to N[0, \mathcal{W}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)],$$

where

$$\begin{aligned}
\mathcal{W}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = {} & \mathcal{V}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) + \mathcal{J}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
& \mathcal{J}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
& + \mathcal{F}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty, \boldsymbol{v}_0)\mathcal{J}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) + \mathcal{J}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
& \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{F}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \\
\mathcal{V}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = {} & V\left\{ \boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \middle| \boldsymbol{\varphi} \right\}, \\
\mathcal{J}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = {} & E\left\{ \frac{\partial \boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]}{\partial \boldsymbol{\phi}'} \middle| \boldsymbol{\varphi} \right\}, \\
\mathcal{F}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = {} & cov\left\{ \frac{\partial \boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]}{\partial \boldsymbol{\phi}'}, \boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}) \middle| \boldsymbol{\varphi} \right\}
\end{aligned}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

and $\mathcal{A}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ are defined in (12) and (13), respectively.

In the next subsections, we provide detailed expressions for $\mathcal{V}(\boldsymbol{\phi}; \boldsymbol{\varphi})$, $\mathcal{J}(\boldsymbol{\phi}; \boldsymbol{\varphi})$ and $\mathcal{F}(\boldsymbol{\phi}; \boldsymbol{\varphi})$ which exploit that the true shocks are cross-sectionally and serially independent under the null hypothesis of correct specification of static ICA model (1) or dynamic SVAR model (2).

3.2.1 Covariance across influence functions

Consider a generic element of the matrix $cov\{\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})], \boldsymbol{m}'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]|\boldsymbol{\varphi}\}$, say

$$\begin{aligned} cov\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})], m_{\boldsymbol{h}'}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]|\boldsymbol{\varphi}\} = & E\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]m_{\boldsymbol{h}'}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]|\boldsymbol{\varphi}\} \\ & -E\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]|\boldsymbol{\varphi}\}E\{m_{\boldsymbol{h}'}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]|\boldsymbol{\varphi}\}. \end{aligned}$$

If we exploit the cross-sectional independence of the shocks under the null hypothesis, which implies that at the true values

$$E\left(\prod_{i=1}^N \varepsilon_{it}^{*h_i}\right) = \prod_{i=1}^N E(\varepsilon_{it}^{*h_i}),$$

obtain

$$cov\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], m_{\boldsymbol{h}'}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]|\boldsymbol{\varphi}_0\} = \prod_{i=1}^N E\left[\varepsilon_{it}^{*(h_i+h_i')}\right] - \prod_{i=1}^N E(\varepsilon_{it}^{*h_i})E(\varepsilon_{it}^{*h_i'}). \tag{15}$$

3.2.2 The expected Jacobian

Straightforward application of the chain rule implies that

$$\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]}{\partial \boldsymbol{\phi}} = \frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]}{\partial \boldsymbol{\varepsilon}'} \frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}}.$$

On this basis, the following proposition characterises the expected Jacobian matrix for any $\boldsymbol{h}$:

Proposition 2 *Suppose that model (2) satisfies Assumption 1. Then, the expected Jacobian matrix of $m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ evaluated at the true values is given by*

$$\begin{aligned} \mathrm{J}_{\boldsymbol{h}\boldsymbol{\tau}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) &= E\left[\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\tau}'}\middle| \boldsymbol{\varphi}_0\right] \\ &= -E\left[\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\middle| \boldsymbol{\varphi}_0\right] \boldsymbol{C}_0^{-1}, \\ \mathrm{J}_{\boldsymbol{h}\boldsymbol{a}_i}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) &= E\left[\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{a}_i'}\middle| \boldsymbol{\varphi}_0\right] \end{aligned}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$$= -E\left[\left.\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\right| \boldsymbol{\varphi}_0\right]\left[E(\boldsymbol{y}_{t-i}^{\prime}|\, \boldsymbol{\varphi}_0) \otimes \boldsymbol{C}_0^{-1}\right]$$

and

$$\begin{aligned}\mathcal{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) &= E\left[\left.\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{c}^{\prime}}\right| \boldsymbol{\varphi}_0\right] \\ &= -E\left\{\left.\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\left[\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_0) \otimes \boldsymbol{C}_0^{-1}\right]\right| \boldsymbol{\varphi}_0\right\}.\end{aligned}$$

As for $\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]/\partial \boldsymbol{\varepsilon}^{*\prime}$, if we denote all the distinct second, third and fourth moments by

$$\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = \begin{pmatrix} \boldsymbol{m}^{cv}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \\ \boldsymbol{m}^{cs}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \\ \boldsymbol{m}^{ck}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_N[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \\ \boldsymbol{T}_N[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \\ \boldsymbol{Q}_N[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \end{pmatrix}, \tag{16}$$

where $\boldsymbol{D}_N$, $\boldsymbol{T}_N$ and $\boldsymbol{Q}_N$ are the duplication, triplication and quadruplication matrices, respectively [see Meijer (2005) for details], the results we derive in "Appendix B.1" provide an easy way to compute all those derivatives recursively.

3.2.3 The covariance with the score

Let $\boldsymbol{\ell}_N$ denote a vector of N ones and $I(.)$ the usual indicator function. The following proposition provides the last ingredient of the adjusted covariance matrix in Proposition 1.

Proposition 3 *Suppose that model* (2) *satisfies Assumption* 1. *Then, the covariance between the influence function $m_{\boldsymbol{h}}(\cdot)$ and the pseudo-log-likelihood scores evaluated at the (pseudo) true values is given by*

$$cov\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], \boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0\} = \mathcal{F}_{\boldsymbol{h}}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = E[\mathcal{F}_{\boldsymbol{h}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0)], \tag{17}$$

where

$$\mathcal{F}_{\boldsymbol{h}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = \begin{bmatrix} \mathcal{F}_{\boldsymbol{h}l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{F}_{\boldsymbol{h}s}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{F}_{\boldsymbol{h}r}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_{lt}^{\prime}(\boldsymbol{\theta}_0) & \boldsymbol{0} \\ \boldsymbol{Z}_s^{\prime}(\boldsymbol{\theta}_0) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_q \end{bmatrix},$$

$\mathcal{F}_{\boldsymbol{h}l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0)$ is a $1 \times N$ vector whose entries are such that for any i with $h_i > 0$,

$$\mathcal{F}_{\boldsymbol{h}s(i,i^{\prime})}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -cov\left\{m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], I(i = i^{\prime}) + \left.\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_0); \boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*}\varepsilon_{i^{\prime}t}^*(\boldsymbol{\theta}_0)\right| \boldsymbol{\varphi}_0\right\}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

and zero otherwise, $\mathcal{F}_{\boldsymbol{hs}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0)$ *is a* $1 \times N^2$ *vector whose entries are such that for any* i *with* $h_i > 0$ *and* i' *with* $h_{i'} > 0$

$$\mathrm{F}_{\boldsymbol{h}s(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -cov\left\{ m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], I(i=i') + \frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_0); \boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*} \varepsilon_{i't}^*(\boldsymbol{\theta}_0) \right| \boldsymbol{\varphi}_0 \right\}$$

and zero otherwise, and finally

$$\mathcal{F}_{\boldsymbol{hr}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = \mathrm{F}'_{\boldsymbol{hr}}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0)\boldsymbol{\ell}_N,$$

with $\mathrm{F}_{\boldsymbol{hr}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0)$ *another block diagonal matrix of order* $N \times q$ *with typical block of size* $1 \times q_i$*,*

$$\mathrm{F}_{\boldsymbol{h}r(i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = cov\left\{ m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], \frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_0); \boldsymbol{\varrho}_{i\infty}]}{\partial \boldsymbol{\varrho}_i'} \right| \boldsymbol{\varphi}_0 \right\}$$

and zero otherwise.

4 Particular cases

4.1 Testing normality

As we have mentioned before, we can use (14) to test the null hypothesis that a single structural shock is Gaussian by comparing its third and fourth sample moments with 0 and 3, respectively, which are the population values of those moments under the null of normality. Nevertheless, many authors [see, e.g. Bontemps and Meddahi (2005) and the references therein] convincingly argue that it is generally more appropriate to look at the sample averages of the third and fourth Hermite polynomials instead. In particular, one should consider $H_3(\varepsilon_{it}^*) = \varepsilon_{it}^{*3} - 3\varepsilon_{it}^*$ and $H_4(\varepsilon_{it}^*) = \varepsilon_{it}^{*4} - 6\varepsilon_{it}^{*2} + 3$ rather than ε_{it}^{*3} and ε_{it}^{*4} only. The reason is that Hermite polynomials have two main advantages. First, given that

$$\frac{\partial H_3(\varepsilon_{it}^*)}{\partial \varepsilon_i^*} = 3H_2(\varepsilon_{it}^*) \quad \text{and} \quad \frac{\partial H_4(\varepsilon_{it}^*)}{\partial \varepsilon_i^*} = 4H_3(\varepsilon_{it}^*),$$

the results in Proposition 2 immediately imply that their expected Jacobians will be 0 under the null of normality, so they are immune to the sampling uncertainty resulting from using estimated shocks. Second, $H_3(\varepsilon_{it}^*)$ and $H_4(\varepsilon_{it}^*)$ are orthogonal under the Gaussian null, which means that the joint test is simply the sum of two asymptotically independent components: one for skewness and another one for kurtosis.

The properties of the estimators that we use, though, mean that the usual implementation of the Jarque and Bera (1980) test, which simply looks at the sample averages of $\varepsilon_{it}^{*3}(\hat{\boldsymbol{\theta}}_T)$ and $\varepsilon_{it}^{*4}(\hat{\boldsymbol{\theta}}_T)$, yields numerically the same statistics as the tests based on the Hermite polynomials despite the fact that it ignores the terms involving ε_{it}^* and ε_{it}^{*2}. The intuition is as follows. Proposition 1 in Fiorentini and Sentana (2020) states that

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

the PMLEs of the unconditional mean and variance of a univariate finite mixture of normals numerically coincide with the sample mean and variance (with denominator T) of the observed series. Given that log-likelihood function (4) for any given values of $\boldsymbol{a}$ and $\boldsymbol{j}$ is effectively the sum of N such univariate log-likelihoods with parameters that are variation-free, the estimated shocks will be such that

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}^{*}(\hat{\boldsymbol{\theta}}_T) = 0 \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}^{*2}(\hat{\boldsymbol{\theta}}_T) - 1 = 0 \quad \forall i \tag{18}$$

regardless of the sample size. This property also has interesting implications for the independence tests that we will consider in the next section because, in effect, each estimated shock will be standardised in the sample.

Finally, it is important to emphasise that the non-normality of a single shock does not guarantee the identification of the model parameters, in the same way as its normality does not imply they are underidentified. As we shall see in the Monte Carlo section, though, researchers can get an informative guide to the validity of Assumption 1 by looking at the normality tests for all the individual shocks.

4.2 Testing independence

At first sight, the arguments in the previous section might suggest that the sample covariances between the estimated shocks will also be 0 by construction. However, this is not generally true. The finite normal mixture PMLEs guarantee the univariate standardisation of each shock, but it does not imply their orthogonality in any given sample, unlike what would happen with a Gaussian likelihood function in which enough a priori restrictions were imposed on $\boldsymbol{C}$ to render the model exactly identified. Intuitively, the parameter values that maximise (4) are trying to make the estimated shocks stochastically independent, not merely orthogonal [see Herwartz (2018)].

For that reason, the first test for independence that we consider will be based on the second cross-moment condition

$$E(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}) = 0, \; i \neq i' \tag{19}$$

In other words, we are simply assessing if the sample correlation between the i^{th} and i'^{th} estimated shocks is significantly different from zero in the usual statistical sense.

Nevertheless, we can also go beyond linear dependence and look at moments that characterise the co-skewness across the structural shocks. These can be of two types:

$$E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}) - E(\varepsilon_{it}^{*2})E(\varepsilon_{i't}^{*}) = E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}) = 0, \; i \neq i', \tag{20}$$

and

$$E(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) - E(\varepsilon_{it}^{*})E(\varepsilon_{i't}^{*})E(\varepsilon_{i''t}^{*}) = E(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) = 0, \; i \neq i' \neq i'', \tag{21}$$

depending on whether they involve two or three different shocks.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Finally, we can also look at the different co-kurtosis among the shocks, which may involve a pair of shocks, namely

$$E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) - E(\varepsilon_{it}^{*2})E(\varepsilon_{i't}^{*2}) = E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) - 1 = 0,\ i \neq i', \tag{22}$$

and

$$E(\varepsilon_{it}^{*3}\varepsilon_{i't}^{*}) - E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*}) = E(\varepsilon_{it}^{*3}\varepsilon_{i't}^{*}) = 0,\ i \neq i', \tag{23}$$

three shocks

$$E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) - E(\varepsilon_{it}^{*2})E(\varepsilon_{i't}^{*})E(\varepsilon_{i''t}^{*}) = E(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) = 0,\ i \neq i' \neq i'', \tag{24}$$

and even four shocks

$$E(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}\varepsilon_{i'''t}^{*}) - E(\varepsilon_{it}^{*})E(\varepsilon_{i't}^{*})E(\varepsilon_{i''t}^{*})E(\varepsilon_{i'''t}^{*}) = E(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}\varepsilon_{i'''t}^{*}) = 0,$$
$$i \neq i' \neq i'' \neq i'''. \tag{25}$$

Thus, we substantially expand the set of moments researchers can use to test for the independence of the components relative to Hyvärinen (2013), who only suggested looking at the co-kurtosis terms in (22). The above moment conditions also augment those considered by Lanne and Luoto (2021), who focus on (19), (22) and (23), together with $E(\varepsilon_{it}^{*}) = 0$ and $E(\varepsilon_{it}^{*2}) = 1$.

4.2.1 Covariance across influence functions

Next, we derive in detail the nonzero elements of the covariance matrix of the second, third and fourth moments in (16).

It is easy to see that under the null hypothesis of independence, the only nonzero elements of the covariance matrix of $\boldsymbol{m}^{cv}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$ are

$$V(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}) = 1.$$

In turn, in the case of $\boldsymbol{m}^{cs}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$ and $\boldsymbol{m}^{ck}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$, the nonzero elements are

$$\begin{aligned} V(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) &= 1, \\ V(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}) &= E(\varepsilon_{it}^{*4}), \\ cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}, \varepsilon_{i't}^{*2}\varepsilon_{it}^{*}) &= E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}), \end{aligned}$$

and

$$\begin{aligned} V(\varepsilon_{it}^{*}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}\varepsilon_{i'''t}^{*}) &= 1, \\ V(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*}) &= E(\varepsilon_{it}^{*4}), \\ V(\varepsilon_{it}^{*3}\varepsilon_{i't}^{*}) &= E(\varepsilon_{it}^{*6}), \end{aligned}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$$
\begin{aligned}
V(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) &= E(\varepsilon_{it}^{*4})E(\varepsilon_{i't}^{*4})-1,\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*},\varepsilon_{i't}^{*2}\varepsilon_{it}^{*}\varepsilon_{i''t}^{*}) &= E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),\\
cov(\varepsilon_{it}^{*3}\varepsilon_{i't}^{*},\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) &= E(\varepsilon_{it}^{*5})E(\varepsilon_{i't}^{*3}),\\
cov(\varepsilon_{it}^{*3}\varepsilon_{i't}^{*},\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) &= E(\varepsilon_{it}^{*5})E(\varepsilon_{i't}^{*3}),\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}\varepsilon_{i''t}^{*},\varepsilon_{i't}^{*2}\varepsilon_{it}^{*}\varepsilon_{i''t}^{*}) &= E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2},\varepsilon_{it}^{*2}\varepsilon_{i''t}^{*2}) &= E(\varepsilon_{it}^{*4})-1,\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i''t}^{*},\varepsilon_{i't}^{*2}\varepsilon_{i''t}^{*}) &= 1,
\end{aligned}
$$

respectively, which can be consistently estimated from $\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)$ under standard regularity conditions.

Finally, the nonzero covariance terms across the different elements of $\boldsymbol{m}^{cv}(\boldsymbol{\varepsilon}_t^*)$, $\boldsymbol{m}^{cs}(\boldsymbol{\varepsilon}_t^*)$ and $\boldsymbol{m}^{ck}(\boldsymbol{\varepsilon}_t^*)$ are

$$
\begin{aligned}
cov(\varepsilon_{it}^{*}\varepsilon_{i't}^{*},\varepsilon_{it}^{*2}\varepsilon_{i't}^{*}) &= E(\varepsilon_{it}^{*3}),\\
cov(\varepsilon_{it}^{*}\varepsilon_{i't}^{*},\varepsilon_{it}^{*3}\varepsilon_{i't}^{*}) &= E(\varepsilon_{it}^{*4}),\\
cov(\varepsilon_{it}^{*}\varepsilon_{i't}^{*},\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) &= E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*},\varepsilon_{it}^{*3}\varepsilon_{i't}^{*}) &= E(\varepsilon_{it}^{*5}),\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*},\varepsilon_{i't}^{*3}\varepsilon_{it}^{*}) &= E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*4}),\ \text{and}\\
cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*},\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) &= E(\varepsilon_{it}^{*4})E(\varepsilon_{it}^{*3}).
\end{aligned}
$$

4.2.2 The expected Jacobian

Straightforward calculations allow us to show that the expected Jacobian of the covariances across shocks in (19) will be given by

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0},\ \mathrm{J}_{\boldsymbol{ha}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-(\boldsymbol{e}_{i'}'\otimes\boldsymbol{c}_0^{i\cdot})-(\boldsymbol{e}_i'\otimes\boldsymbol{c}_0^{i'\cdot}),$$

where $\boldsymbol{e}_i$ is the i^{th} canonical vector and $\boldsymbol{c}^{i\cdot}$ denotes the i^{th} row of $\boldsymbol{C}^{-1}$.

Analogously, for the third cross-moments in (20), we will have

$$
\begin{aligned}
\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-\boldsymbol{c}_0^{i'\cdot},\ \mathrm{J}_{\boldsymbol{ha}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)&=-[E(\boldsymbol{y}_{t-k}'|\boldsymbol{\varphi}_0)\otimes\boldsymbol{c}_0^{i'\cdot}]\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)\\
&=-E(\varepsilon_{it}^{*3})(\boldsymbol{e}_i'\otimes\boldsymbol{c}_0^{i'\cdot}),
\end{aligned}
$$

while for those in (21) we get

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0},\mathrm{J}_{\boldsymbol{ha}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}.$$

In turn, for the fourth moments in (22), we will have

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0},\ \mathrm{J}_{\boldsymbol{ha}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-2(\boldsymbol{e}_i'\otimes\boldsymbol{c}_0^{i\cdot}+\boldsymbol{e}_{i'}'\otimes\boldsymbol{c}_0^{i'\cdot}),$$

while for (23) we get

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E(\varepsilon_{it}^{*3})\boldsymbol{c}_0^{i'\cdot},\ \mathrm{J}_{\boldsymbol{h a}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E(\varepsilon_{it}^{*3})[E(\boldsymbol{y}_{t-k}'|\boldsymbol{\varphi}_0)\otimes\boldsymbol{c}_0^{i'\cdot}]$$

and

$$\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-3(\boldsymbol{e}_{i'}'\otimes\boldsymbol{c}_0^{i\cdot})-E(\varepsilon_{it}^{*4})(\boldsymbol{e}_i'\otimes\boldsymbol{c}_0^{i'\cdot}).$$

Similarly, the expected Jacobian of (24) involves

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0},\ \mathrm{J}_{\boldsymbol{h a}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-(\boldsymbol{e}_{i'}'\otimes\boldsymbol{c}_0^{i''\cdot})-(\boldsymbol{e}_{i''}'\otimes\boldsymbol{c}_0^{i'\cdot}).$$

Finally, when we look at (25), we unsurprisingly end up with

$$\mathrm{J}_{\boldsymbol{h\tau}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0},\mathrm{J}_{\boldsymbol{h a}_k}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}\ \text{and}\mathrm{J}_{\boldsymbol{hc}}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=\boldsymbol{0}.$$

4.2.3 The covariance with the score

As we have seen before, we need to explicitly compute the expressions in Proposition 3 to obtain (17). Fortunately, some of those expressions simplify considerably for the cross-moments we use to test independence. Intuitively, the reason is that the independence of the shocks implies that when $\boldsymbol{h}$ is such that $h_i=1$, we will have

$$E\left[\frac{\partial\ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon_i^*}\varepsilon_{i't}^{*h_{i'}}\varepsilon_{i''t}^{*h_{i''}}\right]=0$$

and

$$E\left[\frac{\partial\ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon_i^*}\varepsilon_{it}^*\varepsilon_{i't}^{*h_{i'}}\varepsilon_{i''t}^{*h_{i''}}\right]=-E(\varepsilon_{i't}^{*h_{i'}})E(\varepsilon_{i''t}^{*h_{i''}})$$

for $i\neq i',i''$.

As a result, (17) will be zero for the second moments $E(\varepsilon_{it}^*\varepsilon_{i't}^*)$, except for $\mathrm{F}_{\boldsymbol{h}s(i,i')}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)$, which will be 1 when $i'\neq i$.

In addition, if we exploit the independence between i and i' and the fact that $E(\varepsilon_{i't}^{*2})=1$, we can easily prove that the only nonzero covariance elements for the co-skewness influence functions $E(\varepsilon_{it}^{*2}\varepsilon_{i't}^*)$ will be

$$\mathrm{F}_{\boldsymbol{h}l(i')}(\boldsymbol{\varrho}_\infty,\boldsymbol{\varphi}_0)=1,\ \mathrm{F}_{\boldsymbol{h}s(i,i')}(\boldsymbol{\varrho}_\infty,\boldsymbol{\varphi}_0)=-E\left[\frac{\partial\ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon_i^*}\varepsilon_{it}^{*2}\right],$$

$$\mathrm{F}_{\boldsymbol{h}s(i',i)}(\boldsymbol{\varrho}_\infty,\boldsymbol{\varphi}_0)=E(\varepsilon_{it}^{*3}),$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$$\mathrm{F}_{\boldsymbol{hs}(i',i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon^*_{i't}; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_{i'}} \varepsilon^{*2}_{i't}\right] \text{ and} \mathrm{F}_{\boldsymbol{hr}(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$$
$$= E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}'_i} \varepsilon^*_{it}\right],$$

while all of them are zero for $E(\varepsilon^*_{it}\varepsilon^*_{i't}\varepsilon^*_{i''t})$.

Similarly, we can also prove that for the co-kurtosis influence functions $E(\varepsilon^{*2}_{it}\varepsilon^{*2}_{i't})$, the only nonzero terms are

$$\mathrm{F}_{\boldsymbol{hl}(i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i} \varepsilon^{*2}_{it}\right], \mathrm{F}_{\boldsymbol{hs}(i,i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -1 - E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i} \varepsilon^{*3}_{it}\right],$$
$$\mathrm{F}_{\boldsymbol{hs}(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E(\varepsilon^{*3}_{i't})E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i} \varepsilon^{*2}_{it}\right] \text{ and } \mathrm{F}_{\boldsymbol{hr}(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}'_i} \varepsilon^*_{it}\right].$$

In turn, we end up with

$$\mathrm{F}_{\boldsymbol{hl}(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = E(\varepsilon^{*3}_{it}), \mathrm{F}_{\boldsymbol{hs}(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i} \varepsilon^{*3}_{it}\right],$$
$$\mathrm{F}_{\boldsymbol{hs}(i',i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0 = E(\varepsilon^{*4}_{it}), \mathrm{F}_{\boldsymbol{hs}(i',i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E[\varepsilon^{*3}_{it}]E\left[\frac{\partial \ln f(\varepsilon^*_{i't}; \boldsymbol{\varrho}_{i'\infty})}{\partial \varepsilon^*_{i'}} \varepsilon^{*2}_{i't}\right]$$

and

$$\mathrm{F}_{\boldsymbol{hr}(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = E(\varepsilon^{*3}_{it})E\left[\frac{\partial \ln f(\varepsilon^*_{i't}; \boldsymbol{\varrho}_{i'\infty})}{\partial \boldsymbol{\varrho}'_{i'}} \varepsilon^*_{i't}\right]$$

for the covariances of the co-kurtosis terms $E(\varepsilon^{*3}_{it}\varepsilon^*_{i't})$ with the scores.

In contrast, the only nonzero covariance of the co-kurtosis influence functions $E(\varepsilon^*_{it}\varepsilon^*_{i't}\varepsilon^{*2}_{i''t})$ with the scores will be $\mathrm{F}_{\boldsymbol{hs}(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = 1$ when $i' \neq i$.

Finally, all the covariances of the scores with $E(\varepsilon^*_{it}\varepsilon^*_{i't}\varepsilon^*_{i''t}\varepsilon^*_{i'''t})$ will be 0 too.

4.3 Combining our tests

Interestingly, we can use the expressions previously derived to prove that under the joint null hypothesis of mutually independent shocks and the normality of one of them, the two separate tests that we have discussed in Sects. 4.1 and 4.2 are asymptotically independent, so effectively the joint test would simply be the sum of those two components.

In addition, we can also prove that a test that jointly assessed the independence and normality of all the shocks would be asymptotically equivalent under the null to a multivariate Hermite-based test of multivariate normality [see Amengual et al. (2021a)] applied to the reduced form residuals once one eliminates the moment condition related to the covariance of the shocks, whose asymptotic variance when evaluated at the PMLEs would be zero under the null.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

5 Monte Carlo analysis

In this section, we assess the finite sample size and power of the normality and independence tests discussed in Sects. 4.1 and 4.2 by means of several Monte Carlo simulation exercises. In addition, we provide some evidence on the effects that dependence across shocks induces on the estimators of the impact multipliers.

5.1 Design and computational details

For the sake of brevity, we focus on the bivariate case in the main text.[7] Specifically, we generate samples of size T from the following bivariate static process

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} \varepsilon^*_{1t} \\ \varepsilon^*_{2t} \end{pmatrix} \tag{26}$$

with $\tau_1 = 1$, $\tau_2 = -1$, $c_{11} = 1$, $c_{12} = .5$, $c_{21} = 0$ and $c_{22} = 2$. However, our PML estimation procedure does not exploit the restriction that the loading matrix of the shocks is upper triangular. Importantly, given that we can easily prove from (4) that the estimated shocks are numerically invariant to affine transformations of the y's, and that the same is true of the different test statistics, the results that we report below do not depend on our choice of $\boldsymbol{\tau}$ and $\boldsymbol{C}$.

We consider both $T = 250$, which is realistic in most macroapplications with monthly or quarterly data, and $T = 1000$, which is representative of financial applications with daily data. The precise DGPs we consider for the shocks are described in Sect. 5.1.2.

5.1.1 Estimation details

To estimate the parameters of the model above, we assume that ε^*_{1t} and ε^*_{2t} follow two serially and cross-sectionally independent standardised discrete mixture of two normals, or $\varepsilon^*_{it} \sim DMN(\delta_i, \varkappa_i, \lambda_i)$ for short, so that

$$\varepsilon^*_{it} = \begin{cases} N[\mu^*_1(\boldsymbol{\varrho}_i), \sigma^{*2}_1(\boldsymbol{\varrho}_i)] \text{ with probability } \lambda_i \\ N[\mu^*_2(\boldsymbol{\varrho}_i), \sigma^{*2}_2(\boldsymbol{\varrho}_i)] \text{ with probability } 1-\lambda_i \end{cases} \tag{27}$$

with

$$\begin{aligned} \mu^*_1(\boldsymbol{\varrho}_i) &= \delta_i(1-\lambda_i), \\ \mu^*_2(\boldsymbol{\varrho}_i) &= -\delta_i\lambda_i, \\ \sigma^{*2}_1(\boldsymbol{\varrho}_i) &= \frac{1-\lambda_i(1-\lambda_i)\delta_i^2}{\lambda_i+(1-\lambda_i)\varkappa_i}, \\ \sigma^{*2}_2(\boldsymbol{\varrho}_i) &= \varkappa_i\sigma^{*2}_1(\boldsymbol{\varrho}_i), \end{aligned}$$

[7] Nevertheless, we include simulation results for a trivariate model in "Appendix C".

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

and $\boldsymbol{\varrho}_i = (\delta_i, \varkappa_i, \lambda_i)'$. Hence, we can interpret $\varkappa_i$ as the ratio of the two variances and δ_i as the parameter that regulates the distance between the means of the two underlying components.[8]

As a consequence, the contribution of observation t to pseudo-log-likelihood function (4) will be

$$l[\varepsilon^*_{it}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i] = \ln\{\lambda_i \cdot \phi[\varepsilon^*_{it}(\boldsymbol{\theta}); \mu^*_1(\boldsymbol{\varrho}_i), \sigma^{*2}_1(\boldsymbol{\varrho}_i)] + (1-\lambda_i)\cdot \phi[\varepsilon^*_{it}(\boldsymbol{\theta}); \mu^*_2(\boldsymbol{\varrho}_i), \sigma^{*2}_2(\boldsymbol{\varrho}_i)]\},$$

where $\phi(\varepsilon; \mu, \sigma^2)$ denotes the probability density function of a Gaussian random variable with mean μ and variance σ^2 evaluated at ε. Importantly, we maximise the log-likelihood with respect to the two elements of $\boldsymbol{\tau}$, the four elements of $\boldsymbol{C}$ and the six shape parameters subject to the nonlinear constraint $\delta_i^2 < \lambda_i^{-1}(1-\lambda_i)^{-1}$, which we impose to guarantee the strict positivity of $\sigma^{*2}_1(\boldsymbol{\varrho}_i)$. Without loss of generality, we also restrict $\varkappa_i \in (0, 1]$ as a way of labelling the components, which in turn ensures the strict positivity of $\sigma^{*2}_2(\boldsymbol{\varrho}_i)$. Finally, we impose $\lambda_i \in (0, 1)$ to avoid degenerate mixtures.[9]

We maximise the log-likelihood subject to these three constraints on the shape parameters using a derivative-based quasi-Newton algorithm, which converges quadratically in the neighbourhood of the optimum. To exploit this property, we start the iterations by obtaining consistent initial estimators of $\boldsymbol{\tau}$ and $\boldsymbol{C}$, $\overline{\boldsymbol{\tau}}_{FICA}$ and $\overline{\boldsymbol{C}}_{FICA}$ say, using the *FastICA* algorithm of G ävert, Hurri, Särelä, and Hyvärinen.[10]In addition, we obtain initial values of the shape parameters of each shock by performing 20 iterations[11]of the expectation maximisation (EM) algorithm in Dempster et al. (1977) on each of the elements of $\overline{\boldsymbol{\varepsilon}}^*_{t,FICA} = \overline{\boldsymbol{C}}^{-1}_{FICA}\left(\boldsymbol{y}_t - \overline{\boldsymbol{\tau}}_{FICA}\right)$.

As we mentioned in Sect. 2.2, Assumption 1 only guarantees the identification of $\boldsymbol{C}$ up to sign changes and column permutations. Although in empirical applications a researcher would carefully chose the appropriate ordering and interpretation of the structural shocks, this leeway may have severe consequences when analysing Monte Carlo results. For that reason, we systematically choose a unique global maximum from the different observationally equivalent permutations and sign changes of the columns of the matrix $\boldsymbol{C}$ using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne et al. (2017). In addition, we impose that $diag(\boldsymbol{C})$ is positive by simply changing the sign of all the elements of the relevant columns. Naturally, we apply the same changes to the shape parameters estimates and the sign of δ_i.

8 We can trivially extend this procedure to three or more components if we replace the normal random variable in the first branch of (27) by a k-component normal mixture with mean and variance given by $\mu^*_1(\boldsymbol{\varrho})$ and $\sigma^{*2}_1(\boldsymbol{\varrho})$, respectively, so that the resulting random variable will be a $(k+1)$ -component Gaussian mixture with zero mean and unit variance.

9 Specifically, we impose $\varkappa_i \in [\underline{\varkappa}, 1]$ with $\underline{\varkappa} = .0001$, and $\lambda_i \in [\underline{\lambda}, \overline{\lambda}]$ with $\underline{\lambda} = 2/T$ and $\overline{\lambda} = 1 - 2/T$.

10 See Hyvärinen (1999) and https://research.ics.aalto.fi/ica/fastica/ for details on the *FastICA* package.

11 As is well known, the EM algorithm progresses very quickly in early iterations but tends to slow down significantly as it gets close to the optimum. After some experimentation, we found that 20 iterations achieve the right balance between CPU time and convergence of the parameters.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

5.1.2 DGPs under the null and the alternative

The four bivariate DGPs for the standardised shocks that we consider under the null of independence are:

DGP 1: A normal distribution and a discrete mixture of two normals with kurtosis coefficient 4 and skewness coefficients equal to $-.5$, i.e. $\varepsilon^*_{1t} \sim N(0, 1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$.

DGP 1D: The VAR(1) model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} 1/2 & 1/4 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} \varepsilon^*_{1t} \\ \varepsilon^*_{2t} \end{pmatrix}$$

with exactly the same shocks and values of $\boldsymbol{\tau}$ and $\boldsymbol{C}$ as in DGP 1.[12]

DGP 2: Independent discrete mixtures of two normals with kurtosis coefficient 4 and skewness coefficients equal to .5 and $-.5$, respectively. In other words, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$.

DGP 3: A Student t with 10 degrees of freedom (and kurtosis coefficient equal to 4), and an asymmetric t with kurtosis and skewness coefficients equal to 4 and $-.5$, respectively, so that $\beta = -1.354$ and $\nu = 18.718$ in the notation of Mencía and Sentana (2012).

The left panels of Fig. 1a–c display the density functions of these distributions over a range of ± 4 standard deviations with the standard normal as a benchmark, while the right panels zoom in on the left-tail.

In turn, under the alternative of cross-sectionally dependent shocks we simulate from the following three standardised joint distributions:

DGP 4: Bivariate Student t with 6 degrees of freedom.

DGP 5: Bivariate asymmetric t with skewness vector $\boldsymbol{\beta} = -5\boldsymbol{\ell}_2$ and degrees of freedom parameter $\nu = 16$ [see Mencía and Sentana (2012) for details].

DGP 6: Bivariate mixture of two zero-mean normal vectors with covariance matrices

$$\boldsymbol{\Omega}_1 = \begin{pmatrix} 1/[\lambda + \varkappa_1(1-\lambda)] & 0 \\ 0 & 1/[\lambda + \varkappa_2(1-\lambda)] \end{pmatrix},$$

$$\boldsymbol{\Omega}_2 = \begin{pmatrix} \varkappa_1/[\lambda + \varkappa_1(1-\lambda)] & 0 \\ 0 & \varkappa_2/[\lambda + \varkappa_2(1-\lambda)] \end{pmatrix},$$

which we denote by $DMN_{LL}(\varkappa_1, \varkappa_2, \lambda)$ [see Lanne and Lütkepohl (2010) for details]. Specifically, we set $\varkappa_1 = 0.1$, $\varkappa_2 = 0.2$ and $\lambda = 1/5$.

The left panels of Fig. 2 display the joint densities for these distributions, while their contours are presented in the right panels.

To gauge the finite sample size and power of our proposed independence tests, we generate 20, 000 samples for each of the designs under the null and 5000 for those

[12] Given that Monte Carlo simulations involving a regular bootstrap are very costly in terms of CPU time, we have only compared the results of a VAR(1) with those of a static model for DGP 1.

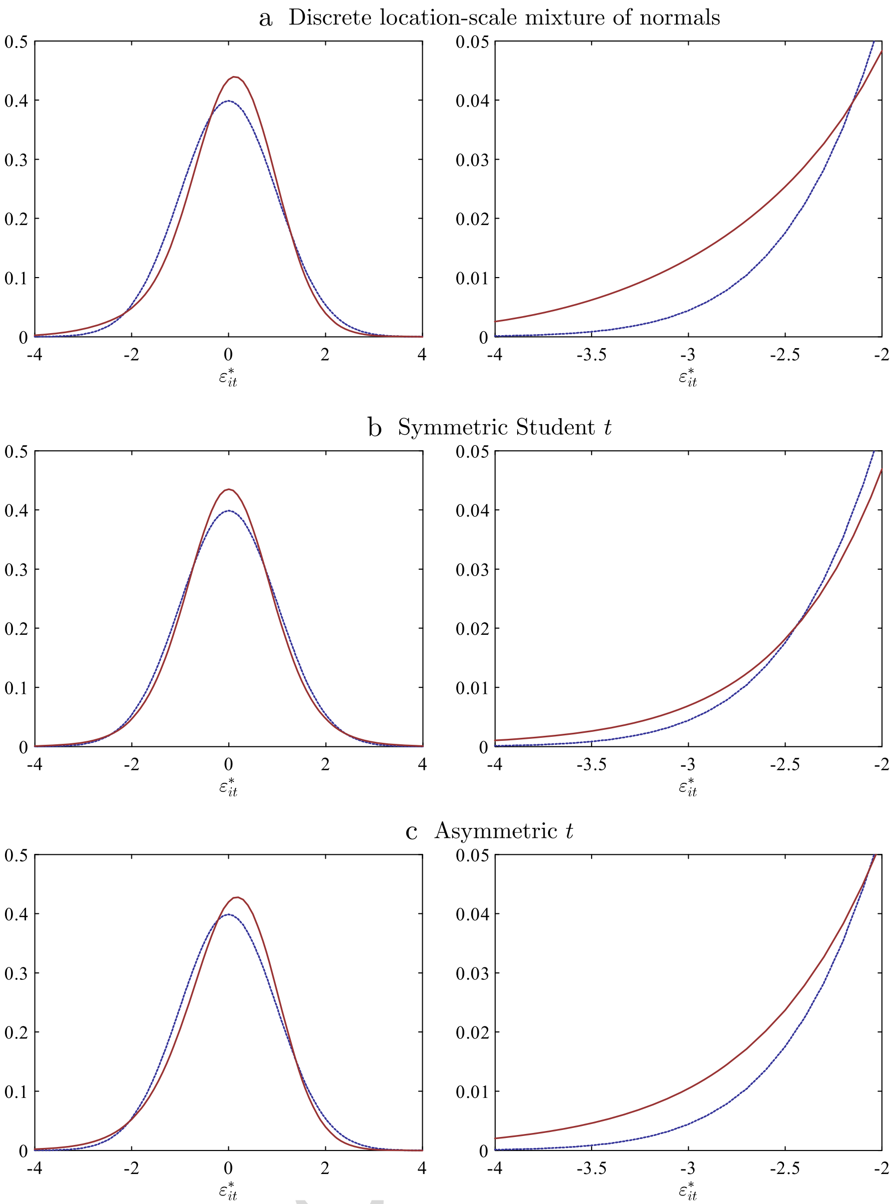

Fig. 1 Univariate densities of the independent shocks. Notes: dashed lines represent the standard normal distribution. **a** Plots a standardised discrete mixture of two normals with skewness and kurtosis coefficients of $-.5$ and 4, respectively (with parameters $\delta = -.859$, $\varkappa = .386$ and $\lambda = 1/5$); **b** Plots a standardised symmetric Student t with the same kurtosis (i.e. 10 degrees of freedom), while **c** plots a standardised asymmetric t with skewness and kurtosis as the one in (**a**) [i.e. with $\beta = -1.354$ and $\nu = 18.718$, see Mencía and Sentana (2012) for details]

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

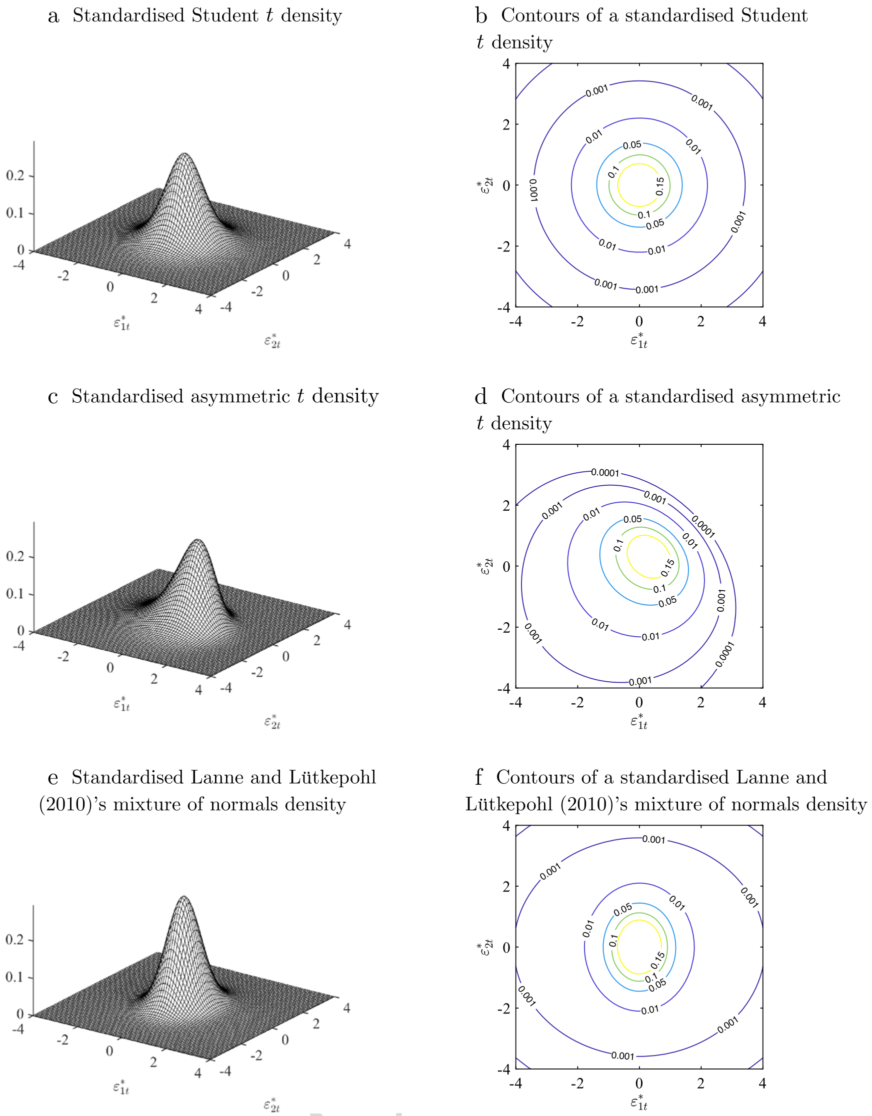

Fig. 2 Densities and contours of the bivariate distributions under the alternative hypotheses. Notes: **a**, **b** plot a bivariate Student t with 6 degrees of freedom; **c**, **d** a standardised bivariate asymmetric t with $\boldsymbol{\beta} = -5\boldsymbol{\ell}_N$ and $\nu = 16$ [see Mencía and Sentana (2012) for details], while **e**, **f** plot a standardised mixture of two bivariate normals with joint mixing Bernoulli with $\lambda = 1/5$ and scale parameters $\varkappa_1 = .1$ and $\varkappa_2 = .2$ [see Sect. 5.1.2 and Lanne and Lütkepohl (2010) for details]

under the alternative. Additionally, we evaluate the small sample size and power of the normality tests presented in Sect. 4.1 using the results from the simulation designs DGP 1 and 1D (null), and DGP 2 and DGP 3 (alternative).

5.1.3 Bootstrap procedures

The theoretical results in Beran (1988) imply that if the usual Gaussian asymptotic approximation provides a reliable guide to the finite sample distribution of the sample version of the moments being tested, the bootstrapped critical values should not only be valid, but also their errors should be of a lower order of magnitude under additional regularity conditions that guarantee the validity of a higher-order Edgeworth expansion. [13]For that reason, we also analyse the performance of applying the bootstrap to the testing procedures we have described in Sects. 4.1 and 4.2.

In the case of our tests for independence, for each Monte Carlo sample, we can easily generate another N_{boot} bootstrap samples of size T that impose the null with probability approaching 1 as T increases as follows. [14]First, we generate NT draws R_{is} from a discrete uniform distribution between 1 and T, which we then use to construct

$$\tilde{\boldsymbol{y}}_s = \hat{\boldsymbol{\tau}}_T + \hat{\boldsymbol{C}}_T \tilde{\boldsymbol{\varepsilon}}_s^*,$$

where $\tilde{\varepsilon}_{is}^* = \hat{\varepsilon}_{iR_{is}}^*$ and $\hat{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T) = \hat{\boldsymbol{C}}_T^{-1}\left(\boldsymbol{y}_t - \hat{\boldsymbol{\tau}}_T\right)$ are the estimated residuals in any given sample.

As for the normality tests, whose null hypothesis is that a single shock ε_{it}^* is Gaussian, we adopt a partially parametric resampling scheme in which the draws of the i^{th} shock $\tilde{\varepsilon}_{is}^*$ are independently simulated from a $N(0, 1)$ distribution, while the draws for the remaining shocks $\tilde{\varepsilon}_{ks}^*$ $(k \neq i)$ are obtained nonparametrically as in the previous paragraph.

Although these bootstrap procedures are simple and fast for any given sample, they quickly become prohibitively expensive in a Monte Carlo exercise as T increases. For this reason, for the designs with $T = 1000$ we rely on the warp-speed method of Giacomini et al. (2013).

[13] Therefore, if the true shocks had unbounded variance, the bootstrap would not work, but neither would the asymptotic approximation.

[14] To see this, notice that under the null,

$$E\left(\prod_{i=1}^N \tilde{\varepsilon}_{is}^{*j_i}\right) = \prod_{i=1}^N E(\varepsilon_{is}^{*j_i}),$$

while under the alternative,

$$E\left(\prod_{i=1}^N \tilde{\varepsilon}_{is}^{*j_i}\right) = \frac{T-1}{T}\prod_{i=1}^N E(\varepsilon_{is}^{*j_i}) + \frac{1}{T}E\left(\prod_{i=1}^N \varepsilon_{is}^{*j_i}\right)$$

where the second term in the right-hand side accounts for the probability of sampling contemporaneous residuals in a sample of size T. Clearly, the second expression converges to the first one as T goes to infinity.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

5.2 Simulation results

5.2.1 Testing normality

Table 1 reports Monte Carlo rejection rates of the normality tests proposed in Sect. 4.1 for DGP 1, 1D, 2 and 3. As can be seen, the null of normality is correctly rejected a large number of times when it does not hold, even in samples of length 250. The only possible exception is the skewness component of the Jarque-Bera test when applied to the symmetric Student t shock in DGP 3. Given that the population third moment is zero in this case, the only source of power is the fact that the sample variability of H_3 is larger for this shock than its theoretical value under Gaussianity.

On the other hand, the first three rows of the panels DGP 1 and 1D, which are the ones with a Gaussian shock, show that the normality tests tend to be oversized at the usual nominal levels, especially for samples of length 250.[15] For that reason, we generate $N_{Boot} = 399$ bootstrap samples at each Monte Carlo replication, as described in Sect. 5.1.3. Table 2 shows that the standard bootstrap version of our tests is pretty accurate for both the third and fourth moment tests. Unlike what we observed in Table 1, though, the size-adjusted power is slightly lower for DGP 1D than for DGP 1.

However, as mentioned at the end of Sect. 4.1, researches may only get a reliable guide to the validity of Assumption 1 by looking at the normality tests for all the individual shocks, the objective being to get at least $N - 1$ rejections. To shed some light on this issue, in Table 3 we report contingency tables which fully characterise the extent to which simultaneous rejections of the individual normality tests occur. As can be seen, our proposed normality tests tend to be rather informative when used in this way.

5.2.2 Testing independence

In Tables 4 ($T = 250$) and 5 ($T = 1000$) we report the Monte Carlo rejection rates of the tests we have proposed in Sect. 4.2 under the null of independence. Specifically, we look at the second, third and fourth moment individual tests in $\boldsymbol{m}^{cv}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$, $\boldsymbol{m}^{cs}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ and $\boldsymbol{m}^{ck}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$, and also at the joint tests for the two co-skewness moments, the three co-kurtosis moments and the combined six moments, including the correlation between the shocks. The left panels of those tables report rejection rates using asymptotic critical values, while the right panels show the bootstrap-based ones for $T = 250$ and the warp-speed bootstrap-based ones for $T = 1000$.[16]

We can see in Table 4 some small to moderate finite sample size distortion when $T = 250$, although in several cases they are corrected by the bootstrap. The only exceptions seem to be DGP 1 and 1D, in which some small distortions remain even with this procedure. Given that in these designs there is only one non-Gaussian shock, a plausible explanation is that the identification of $\boldsymbol{C}$ may be weaker, a conjecture we

[15] Given 20,000 Monte Carlo replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (.86,1.14), (4.70,5.30) and (9.58,10.42) at the 1, 5 and 10% levels, respectively.

[16] All our $i.i.d.$ designs are such that the individual moment tests converge in distribution to a χ_1^2 random variable, and the joint ones to χ_2^2, χ_3^2 and χ_6^2 variables, respectively.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 1 Monte Carlo size and power of normality tests

Nominal size	Asymptotic critical values					
	Sample size $T = 250$			Sample size $T = 1000$		
	10%	5%	1%	10%	5%	1%
	DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN					
$H_3(\varepsilon^*_{1t})$	13.58	7.70	2.45	11.03	5.96	1.32
$H_4(\varepsilon^*_{1t})$	12.37	6.86	2.85	10.38	5.32	1.38
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	13.03	8.17	3.67	10.56	5.76	1.67
$H_3(\varepsilon^*_{2t})$	83.40	77.93	64.27	99.93	99.88	99.50
$H_4(\varepsilon^*_{2t})$	70.78	64.44	51.80	99.26	98.79	96.80
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	85.73	81.33	71.52	99.95	99.94	99.90
	DGP 1D VAR(1)—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN					
$H_3(\varepsilon^*_{1t})$	15.08	8.83	2.78	11.15	5.65	1.19
$H_4(\varepsilon^*_{1t})$	13.28	7.47	2.94	10.82	5.62	1.50
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	14.72	8.96	4.07	11.02	5.91	1.71
$H_3(\varepsilon^*_{2t})$	82.51	77.12	63.70	99.91	99.86	99.60
$H_4(\varepsilon^*_{2t})$	70.17	63.90	51.70	99.29	98.73	96.84
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	85.33	80.75	70.99	99.96	99.94	99.89
	DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN					
$H_3(\varepsilon^*_{1t})$	84.36	78.73	64.33	99.88	99.81	99.39
$H_4(\varepsilon^*_{1t})$	70.53	64.07	51.13	99.22	98.63	95.84
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	86.54	81.92	71.58	99.98	99.95	99.77
$H_3(\varepsilon^*_{2t})$	85.14	79.63	65.82	99.92	99.84	99.50
$H_4(\varepsilon^*_{2t})$	70.86	64.31	51.46	99.41	98.81	95.97
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	87.34	82.88	72.26	100.00	99.98	99.82
	DGP 3—Shocks: ε^*_{1t} asymmetric t & ε^*_{2t} Student t					
$H_3(\varepsilon^*_{1t})$	84.93	79.50	65.37	99.98	99.92	99.76
$H_4(\varepsilon^*_{1t})$	58.58	52.38	42.24	95.10	93.04	87.73
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	82.72	77.21	65.27	99.97	99.91	99.69
$H_3(\varepsilon^*_{2t})$	33.97	25.62	14.52	36.43	28.41	16.68
$H_4(\varepsilon^*_{2t})$	60.68	54.21	42.13	96.98	95.35	90.70
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	60.83	54.14	42.38	95.77	93.85	88.56

Monte Carlo empirical rejection rates of normality tests; 20,000 replications. DMN denotes discrete mixture of two normals. Details on the data generating processes: DGP 1 and 1D, $\varepsilon^*_{1t} \sim N(0, 1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$; DGP 2, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$; and DGP 3, $\varepsilon^*_{1t} \sim At(-1.354, 18.718)$ and $\varepsilon^*_{2t} \sim t(10)$ [see Mencía and Sentana (2012) for details]. Asymptotic critical values: $H_3(\cdot) \sim \chi^2_1$, $H_4(\cdot) \sim \chi^2_1$ and $H_3(\cdot)$ & $H_4(\cdot) \sim \chi^2_2$

will revisit in the next section. For the other DGPs, the results in Table 4 clearly show that the usual bootstrap version of the tests, which is the relevant one in empirical applications, has much better size properties.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 2 Monte Carlo size and power of normality tests with bootstrap: sample size $T = 250$

	Asymptotic critical values			Bootstrap (399 samples) critical values		
Nominal size	10%	5%	1%	10%	5%	1%
	DGP 1—Shocks: ε_{1t}^* normal & ε_{2t}^* DMN					
	Size (ε_{1t}^* normal)					
$H_3(\varepsilon_{1t}^*)$	13.58	7.70	2.45	9.13	4.59	0.98
$H_4(\varepsilon_{1t}^*)$	12.37	6.86	2.85	9.46	4.80	1.18
$H_3(\varepsilon_{1t}^*)$ & $H_4(\varepsilon_{1t}^*)$	13.03	8.17	3.67	9.31	4.70	1.22
	Power (ε_{2t}^* DMN)					
$H_3(\varepsilon_{1t}^*)$	83.40	77.93	64.27	79.94	73.33	55.47
$H_4(\varepsilon_{1t}^*)$	70.78	64.44	51.80	67.75	60.56	38.23
$H_3(\varepsilon_{1t}^*)$ & $H_4(\varepsilon_{1t}^*)$	85.73	81.33	71.52	82.76	75.81	53.79
	DGP 1D VAR(1)—Shocks: ε_{1t}^* normal & ε_{2t}^* DMN					
	Size (ε_{1t}^* normal)					
$H_3(\varepsilon_{1t}^*)$	15.08	8.83	2.80	9.36	4.50	0.91
$H_4(\varepsilon_{1t}^*)$	13.28	7.47	2.94	9.22	4.47	1.10
$H_3(\varepsilon_{1t}^*)$ & $H_4(\varepsilon_{1t}^*)$	14.72	8.96	4.07	8.90	4.31	1.04
	Power (ε_{2t}^* DMN)					
$H_3(\varepsilon_{2t}^*)$	82.51	77.12	63.70	77.24	69.93	51.99
$H_4(\varepsilon_{2t}^*)$	70.17	63.90	51.70	65.57	57.57	36.00
$H_3(\varepsilon_{1t}^*)$ & $H_4(\varepsilon_{1t}^*)$	85.33	80.75	70.99	80.26	72.73	50.63

Monte Carlo empirical rejection rates of normality tests; 20,000 replications. DMN denotes discrete mixture of two normals. Data generated according to DGP 1 and DGP 1D, i.e. $\varepsilon_{1t}^* \sim N(0,1)$ and $\varepsilon_{2t}^* \sim DMN(-.859, .386, 1/5)$. Testing for univariate normality of ε_{1t}^* provides size figures while doing the same but with ε_{2t}^* delivers power measures. Asymptotic critical values: $H_3(\cdot) \sim \chi_1^2$ and $H_4(\cdot) \sim \chi_1^2$. We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement the bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 3 Contingency tables of the normality test based on $H_3(\varepsilon^*_{it})$ & $H_4(\varepsilon^*_{it})$

Sample Size $T = 250$ Bootstrap (399 samples)					Sample Size $T = 1000$ Warp-speed bootstrap				
DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN									
		ε^*_{2t} (Alt.)					ε^*_{2t} (Alt.)		
		Yes	No				Yes	No	
ε^*_{1t}	Yes	2.62	2.08	*4.70*	ε^*_{1t}	Yes	5.01	0.04	*5.05*
(Null)	No	73.19	22.11	*95.30*	(Null)	No	94.92	0.03	*94.95*
		75.81	*24.19*				*99.93*	*0.07*	
DGP 1D VAR(1)—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN									
		ε^*_{2t} (Alt.)					ε^*_{2t} (Alt.)		
		Yes	No				Yes	No	
ε^*_{1t}	Yes	2.27	2.04	*4.31*	ε^*_{1t}	Yes	4.43	0.05	*4.48*
(Null)	No	70.47	25.23	*95.69*	(Null)	No	95.48	0.04	*95.52*
		72.73	*27.27*				*99.91*	*0.09*	
DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN									
		ε^*_{2t} (Alt.)					ε^*_{2t} (Alt.)		
		Yes	No				Yes	No	
ε^*_{1t}	Yes	55.89	18.40	*74.29*	ε^*_{1t}	Yes	99.94	0.02	*99.96*
(Alt.)	No	18.97	6.74	*25.71*	(Alt.)	No	0.04	0.00	*0.04*
		74.86	*25.14*				*99.98*	*0.02*	
DGP 3—Shocks: ε^*_{1t} asymmetric t & ε^*_{2t} Student t									
		ε^*_{2t} (Alt.)					ε^*_{2t} (Alt.)		
		Yes	No				Yes	No	
ε^*_{1t}	Yes	28.07	34.51	*62.58*	ε^*_{1t}	Yes	92.97	6.69	*99.66*
(Alt.)	No	17.74	19.68	*37.42*	(Alt.)	No	0.33	0.01	*0.34*
		45.81	*54.19*				*93.30*	*6.70*	

Monte Carlo empirical rejection rates of normality tests; 20,000 replications. Yes/No refers to rejections of the Gaussian null. DMN denotes discrete mixture of two normals. Details on the data generating processes: DGP 1 and 1D, $\varepsilon^*_{1t} \sim N(0, 1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$; DGP 2, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$; and DGP 3, $\varepsilon^*_{1t} \sim At(-1.354, 18.718)$ and $\varepsilon^*_{2t} \sim t(10)$ [see Mencía and Sentana (2012) for details]. We describe the sampling procedure we use to implement both the standard bootstrap and Giacomini et al. (2013)'s warp-speed bootstrap in Sect. 5.1.3

As can be seen in Table 5, finite sample sizes improve considerably for $T = 1000$. Indeed, the bootstrap versions of the tests seem unnecessary for this sample size because the empirical rejection rates based on asymptotic critical values become generally very close to the nominal ones, though the warp-speed version performs comparably well.

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 4 Monte Carlo size of independence moment tests: sample size $T = 250$

	Asymptotic critical values			Bootstrap (399 samples) critical values			Asymptotic critical values			Bootstrap (399 samples) critical values		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
	DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN						DGP 1D VAR(1)—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	7.12	3.16	0.47	10.11	4.83	0.89	6.82	3.20	0.40	9.13	4.68	0.87
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	7.81	3.49	0.55	8.09	3.85	0.65	7.55	3.49	0.46	9.12	4.38	0.72
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	10.08	4.95	0.86	10.02	4.92	0.97	10.39	5.18	1.03	11.20	5.79	1.39
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	6.26	2.94	0.55	8.43	4.08	0.81	6.46	2.88	0.53	8.51	4.05	0.82
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	8.45	3.94	0.67	10.15	4.98	0.88	7.04	3.11	0.67	9.53	4.81	0.92
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	6.55	2.72	0.76	9.35	4.44	0.87	8.52	4.02	0.67	10.41	5.29	0.98
Co-skewness	8.05	3.74	0.72	8.45	3.89	0.73	8.30	3.98	0.74	10.05	4.87	1.05
Co-kurtosis	5.82	2.86	0.92	8.42	3.99	0.89	5.88	3.09	0.91	9.12	4.50	0.99
Joint test	5.58	3.06	0.92	7.50	3.71	0.83	5.72	3.05	0.78	8.15	4.06	0.80
	DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN						DGP 3—Shocks: ε^*_{1t} asymmetric t & ε^*_{2t} Student t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	7.51	3.40	0.60	10.18	5.13	0.95	6.51	2.96	0.48	9.74	4.67	0.81
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	9.55	5.11	1.31	10.10	5.18	1.30	9.81	5.15	1.11	10.23	5.38	1.27
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	9.13	4.32	0.82	9.96	4.84	0.86	8.38	3.96	0.76	9.05	4.30	0.79
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	7.52	3.75	0.88	9.62	4.86	0.98	6.69	3.43	0.84	9.52	4.78	1.05
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	7.76	3.85	0.87	9.89	4.92	1.00	7.07	3.30	0.70	9.35	4.54	0.87
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	7.48	3.71	1.23	9.86	5.08	1.18	7.08	3.52	1.29	10.06	5.13	1.40
Co-skewness	9.58	5.17	1.46	10.11	5.29	1.31	8.87	4.47	1.04	9.55	4.76	1.03
Co-kurtosis	7.13	4.16	1.55	10.03	5.16	1.16	6.29	3.69	1.47	9.16	4.72	1.21
Joint test	7.71	4.54	1.81	9.63	4.78	1.23	7.07	4.07	1.57	9.07	4.73	1.20

Monte Carlo empirical rejection rates of independence tests; 20,000 replications. DMN denotes discrete mixture of two normals. Details on the data generating processes: DGP 1, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$; DGP 2, $\varepsilon^*_{1t} \sim N(0,1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$; and DGP 3, $\varepsilon^*_{1t} \sim At(-1.354, 18.718)$ and $\varepsilon^*_{2t} \sim t(10)$ [see Mencía and Sentana (2012) for details]. We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement the bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 5 Monte Carlo size of independence moment tests: sample size $T = 1000$

	Asymptotic critical values			Warp-speed bootstrap critical values			Asymptotic critical values			Warp-speed bootstrap critical values		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
	DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN						DGP 1D VAR(1)—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	9.52	4.52	0.94	11.13	5.30	0.83	9.21	4.44	0.93	10.00	5.02	1.06
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	9.80	4.77	0.90	10.06	4.96	0.78	9.41	4.56	0.89	9.95	5.01	1.01
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	10.63	5.49	1.06	10.74	5.33	0.70	10.34	5.29	1.21	10.31	5.38	1.32
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	8.95	4.28	0.77	9.71	4.96	0.78	9.08	4.47	0.92	10.45	5.31	1.16
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	10.02	4.98	1.06	11.12	5.61	1.07	9.65	4.43	0.81	10.83	4.99	0.97
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	8.88	4.42	0.88	9.41	4.72	0.92	9.89	4.89	0.97	10.43	5.25	1.09
Co-skewness	9.90	4.90	0.93	10.33	4.83	0.53	9.60	5.11	1.06	10.03	5.61	1.16
Co-kurtosis	8.65	4.40	1.10	10.15	4.93	0.96	8.70	4.40	1.07	10.40	5.37	1.17
Joint test	8.43	4.26	1.07	10.01	4.67	0.72	8.46	4.33	1.11	10.37	5.26	1.07
	DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN						DGP 3—Shocks: ε^*_{1t} asymmetric t & ε^*_{2t} Student t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	9.41	4.78	0.94	10.16	5.04	0.98	9.28	4.50	0.90	10.84	5.54	1.18
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	9.65	4.69	0.94	10.27	5.12	1.14	10.20	5.15	1.26	10.94	5.52	1.21
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	9.58	4.55	0.93	10.49	4.94	1.16	9.78	5.06	0.97	10.36	5.16	0.89
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	9.10	4.87	1.11	9.93	5.05	1.00	9.16	4.71	1.14	10.89	5.26	1.23
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	9.46	4.81	1.18	10.44	5.24	1.14	9.41	5.10	1.19	10.54	5.58	1.12
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	9.01	4.26	1.05	10.08	5.07	0.81	8.29	4.11	1.27	9.62	4.72	1.18
Co-skewness	9.34	4.69	0.97	10.23	5.23	1.15	9.76	5.07	1.29	10.40	5.29	1.26
Co-kurtosis	8.96	4.87	1.51	10.40	5.01	1.01	8.66	4.89	1.73	10.60	5.62	1.24
Joint test	9.05	5.03	1.58	10.63	5.30	1.33	9.18	5.22	1.74	11.71	6.02	1.33

Monte Carlo empirical rejection rates of independence tests; 20,000 replications. DMN denotes discrete mixture of two normals. Details on the data generating processes: DGP 1, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$; DGP 2, $\varepsilon^*_{1t} \sim N(0, 1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$; and DGP 3, $\varepsilon^*_{1t} \sim At(-1.354, 18.718)$ and $\varepsilon^*_{2t} \sim t(10)$ [see Mencía and Sentana (2012) for details]. We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement Giacomini et al. (2013)'s warp-speed bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 6 Monte Carlo power of independence moment tests: sample size $T = 250$

	Asymptotic critical values			Bootstrap (399 samples) critical values		
Nominal size	10%	5%	1%	10%	5%	1%
	DGP 4—Joint Student t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	6.90	3.32	0.68	10.80	5.36	1.28
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	9.80	5.10	1.10	11.42	6.16	1.22
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	10.02	5.12	1.04	10.94	5.88	1.12
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	8.50	4.84	1.40	11.86	6.00	1.50
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	8.92	5.18	1.70	11.80	6.66	1.84
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	12.04	8.18	3.64	15.02	11.26	3.68
Co-skewness	9.98	5.06	1.26	11.64	5.60	1.38
Co-kurtosis	11.82	7.84	4.10	16.22	9.66	3.20
Joint test	11.80	8.08	4.44	15.12	9.32	3.34
	DGP 5—Joint asymmetric t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	16.00	9.18	3.44	19.90	12.60	4.58
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	25.38	16.34	6.54	25.12	16.06	4.56
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	19.64	12.54	4.58	20.54	12.80	4.56
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	14.46	9.68	3.52	16.94	11.02	3.56
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	14.14	9.02	3.52	17.90	11.44	4.88
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	15.42	10.84	5.60	18.80	13.16	5.12
Co-skewness	23.80	16.08	6.16	23.90	15.06	3.94
Co-kurtosis	16.56	11.82	5.98	21.20	13.70	5.50
Joint test	17.92	11.88	5.80	20.22	11.88	4.28
	DGP 6—Lanne and Lütkepohl (2010)'s mixture					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	37.12	28.50	15.64	39.78	29.00	14.76
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	25.26	17.34	7.80	26.44	18.16	6.50
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	28.00	20.26	9.50	29.44	20.22	7.54
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	28.48	21.00	10.92	30.90	20.48	7.46
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	34.60	26.26	15.26	36.22	25.14	9.14
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	64.14	54.88	38.18	70.82	61.12	26.42
Co-skewness	33.16	24.48	13.32	35.06	23.58	7.72
Co-kurtosis	62.02	53.98	39.84	64.72	49.34	20.26
Joint test	67.02	58.78	43.84	67.02	52.42	22.28

Monte Carlo empirical rejection rates of independence tests; 5000 replications. Details on the data generating processes: DGP 4, joint (standardised) Student t: $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim t(\mathbf{0}, \boldsymbol{I}_2, 6)$; DGP 5, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim At(\mathbf{0}, \boldsymbol{I}_2, -5\boldsymbol{\ell}_2, 16)$ [see Mencía and Sentana (2012) for details]; and DGP 6, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim DMN_{LL}(.1, .2, 1/5)$ (see Sect. 5.1.2 for details). We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement Giacomini et al. (2013)'s warp-speed bootstrap in Sect. 5.1.3

Table 7 Monte Carlo power of independence moment tests: sample size $T = 1000$

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%
	DGP 4—Joint Student t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	15.72	10.04	2.82	17.36	11.26	3.30
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	16.02	9.10	2.86	16.32	9.82	2.86
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	15.74	9.44	2.90	15.98	9.74	3.18
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	18.68	12.44	5.42	20.94	13.02	4.96
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	19.30	12.42	4.94	20.14	12.78	4.48
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	54.78	44.52	27.08	57.74	46.76	26.12
Co-skewness	18.26	11.22	3.76	18.82	11.34	3.72
Co-kurtosis	46.92	38.26	23.36	50.08	40.38	18.28
Joint test	44.50	35.36	21.40	48.50	37.06	16.22
	DGP 5—Joint asymmetric t					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	84.52	81.52	75.24	84.94	81.72	74.14
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	69.28	64.76	56.38	69.78	65.38	55.58
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	98.72	98.28	96.98	98.72	98.24	96.62
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	56.36	50.28	40.08	57.54	50.08	39.96
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	65.62	59.52	48.36	66.02	59.62	45.64
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	88.42	84.16	74.32	90.48	85.66	67.64
Co-skewness	100.00	100.00	99.90	100.00	100.00	99.78
Co-kurtosis	87.32	83.16	74.40	88.00	82.36	66.22
Joint test	100.00	99.94	99.58	100.00	99.94	98.42
	DGP 6—Lanne and Lütkepohl (2010)'s mixture					
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	58.22	51.60	39.84	59.78	52.52	39.84
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	29.00	20.16	9.72	29.88	20.50	9.12
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	33.28	24.64	12.68	32.74	23.92	12.02
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	46.70	38.44	26.34	47.42	37.76	23.24
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	55.76	48.12	34.64	57.80	48.02	28.78
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	99.98	99.86	99.28	99.98	99.88	98.52
Co-skewness	40.46	30.70	16.82	40.76	29.68	14.82
Co-kurtosis	99.80	99.58	98.22	99.80	99.36	94.46
Joint test	99.48	99.08	97.64	99.42	98.68	92.22

Monte Carlo empirical rejection rates of independence tests; 5000 replications. Details on the data generating processes: DGP 4, joint (standardised) Student t: $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim t(\mathbf{0}, \boldsymbol{I}_2, 6)$; DGP 5, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim At(\mathbf{0}, \boldsymbol{I}_2, -5\boldsymbol{\ell}_2, 16)$ [see Mencía and Sentana (2012) for details]; and DGP 6, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim DMN_{LL}(.1, .2, 1/5)$ (see Sect. 5.1.2 for details). We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement Giacomini et al. (2013)'s warp-speed bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Next, we assess the power of the independence tests for $T = 250$ and $T = 1000$ in Tables 6 and 7, respectively. In this respect, we find that the power of our tests against DGP 4 is disappointingly low. A possible explanation is that when the true joint distribution is a symmetric Student t, the dependence between the components is mostly visible in the tails of the distribution. On the other hand, power is mostly coming from co-skewness component (20) in the case of the joint asymmetric t. Still, the test based on the covariance of shocks (19) is also very powerful. Finally, the co-kurtosis test based on (22) is the most powerful single moment test under the Lanne and Lütkepohl (2010) alternative in DGP 6, with the joint tests that include this moment inheriting its power. Nevertheless, the test based on second moment (19) also has non-negligible power for this design.

In summary, although the rejection rates naturally depend on the type of departure from the null and the specific influence function used for testing, the joint test that considers all moments at once seems to be a winner regardless of the sample size.

5.3 Structural parameters estimates

Table 8 reports summary statistics for the Monte Carlo distribution of the PMLEs of the structural parameters. The first thing we would like to highlight is when one of the shocks is Gaussian, the sampling variability and the finite sample bias are noticeably larger than when both shocks are non-Gaussian but independent, which is in line with the conjecture we expressed in the previous section. Still, even in that case the biases are usually small and often negligible. In addition, the Monte Carlo standard deviations of the estimators in Panel B are roughly half those in Panel A, as one would expect.

The situation is completely different when the true shocks are cross-sectionally dependent. Failure of condition 2 in Assumption 1 results into significant biases, mostly in the off-diagonal terms of the impact multiplier matrix. In fact, the Monte Carlo variance of these estimators seems to increase with the sample size. In this respect, it is important to remember that the elements of the $\boldsymbol{C}$ matrix are no longer point identified when the joint distribution of the true shocks is either a symmetric or asymmetric Student t. This is confirmed by the fact that the bias of the estimators is lower for DGP 6, in which the rotations of the shocks are not observationally equivalent [see Lanne and Lütkepohl (2010)].

6 Conclusions and directions for further research

Given that the parametric identification of the structural shocks and their impact coefficients $\boldsymbol{C}$ in SVAR (2) critically hinges on the validity of the identifying restrictions in Assumption 1, it would be desirable that empirical researchers estimating those models reported specification tests that checked those assumptions to increase the empirical credibility of their findings. For that reason, in this paper we propose simple specification tests for independent component analysis and structural vector autoregressions with non-Gaussian shocks that check the normality of a single shock and the potential cross-sectional dependence among several of them. Our tests compare the

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 8 Monte Carlo distribution of parameter estimators

Parameter θ (θ_0)	τ_1 (1)	τ_2 (-1)	c_{11} (1)	c_{21} (0)	c_{12} (.5)	c_{22} (2)	τ_1	τ_2	c_{11}	c_{21}	c_{12}	c_{22}
	Panel A: sample size $T = 250$											
	Under the null of independence						*Under the alternative*					
	DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN						DGP 4—Joint Student *t*					
Mean	1.000	-1.000	0.974	0.087	0.432	1.906	1.000	-1.001	0.975	0.417	0.226	1.741
Std.dev.	0.071	0.126	0.120	0.542	0.310	0.201	0.071	0.127	0.163	0.831	0.465	0.290
	DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN						DGP 5—Joint asymmetric *t*					
Mean	1.001	-1.000	0.983	0.029	0.472	1.951	1.001	-1.000	0.981	0.490	0.200	1.727
Std.dev.	0.071	0.128	0.101	0.384	0.206	0.149	0.072	0.127	0.148	0.831	0.476	0.282
	DGP 3—Shocks: ε^*_{1t} asymmetric *t* & ε^*_{2t} Student *t*						DGP 6—Lanne and Lütkepohl (2010)'s mixture					
Mean	1.000	-0.999	0.979	0.074	0.444	1.922	0.998	-1.002	0.974	0.284	0.322	1.818
Std.dev.	0.071	0.127	0.118	0.505	0.272	0.185	0.071	0.126	0.161	0.780	0.386	0.244

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 8 continued

Parameter θ (θ_0)	τ_1 (1)	τ_2 (−1)	c_{11} (1)	c_{21} (0)	c_{12} (.5)	c_{22} (2)	τ_1	τ_2	c_{11}	c_{21}	c_{12}	c_{22}
	Panel B: sample size $T = 1000$											
	Under the null of independence						*Under the alternative*					
	DGP 1—Shocks: ε^*_{1t} normal & ε^*_{2t} DMN						DGP 4—Joint Student *t*					
Mean	1.000	−1.000	0.993	0.002	0.496	1.985	1.000	−1.000	0.982	0.434	0.219	1.748
Std.dev.	0.035	0.063	0.060	0.223	0.120	0.064	0.036	0.063	0.147	0.824	0.462	0.269
	DGP 2—Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN						DGP 5—Joint asymmetric *t*					
Mean	1.000	−1.000	0.997	−0.000	0.498	1.993	0.999	−1.000	0.994	0.788	−0.020	1.591
Std.dev.	0.036	0.063	0.045	0.142	0.072	0.056	0.035	0.063	0.158	0.888	0.467	0.277
	DGP 3—Shocks: ε^*_{1t} asymmetric *t* & ε^*_{2t} Student *t*						DGP 6—Lanne and Lütkepohl (2010)'s mixture					
Mean	1.000	−1.000	0.994	−0.000	0.497	1.988	1.000	−1.001	0.979	0.140	0.411	1.900
Std.dev.	0.035	0.064	0.055	0.191	0.095	0.060	0.035	0.062	0.132	0.637	0.308	0.176

20,000 (5000) replications under the null (alternative). DMN denotes discrete mixture of two normals. Details on the data generating processes: DGP 1, $\varepsilon^*_{1t} \sim N(0, 1)$ and $\varepsilon^*_{2t} \sim DMN(-.859, .386, 1/5)$; DGP 2, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon^*_{2t} \sim DMN(.859, .386, 1/5)$; DGP 3, $\varepsilon^*_{1t} \sim At(-1.354, 18.718)$ and $\varepsilon^*_{2t} \sim t(10)$ [see Mencía and Sentana (2012) for details]; DGP 4, joint (standardised) Student *t*: $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim t(\mathbf{0}, \boldsymbol{I}_2, 6)$; DGP 5, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim At(\mathbf{0}, \boldsymbol{I}_2, -5\boldsymbol{\ell}_2, 16)$ [see Mencía and Sentana (2012) for details]; and DGP 6, $(\varepsilon^*_{1t}, \varepsilon^*_{2t}) \sim DMN_{LL}(.1, .2, 1/5)$ (see Sect. 5.1.2 for details)

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

integer (product) moments of the shocks in the sample with their population counterparts. Importantly, we explicitly consider the sampling variability resulting from using shocks computed with consistent parameter estimators. We study the finite sample size of our tests in several simulation exercises and discuss some bootstrap procedures. We also show that our tests have non-negligible power against a variety of empirically plausible alternatives.

As we mentioned in introduction, there are many estimators for the parameters of static ICA model (1) in addition to the discrete mixture of normals-based PMLEs we have considered in this paper. For example, even within the same likelihood framework, Fiorentini and Sentana (2020) discuss two other consistent estimators of the conditional mean and variance parameters of the SVAR in (2):

1. The two-step procedure of Gouriéroux et al. (2017), which first estimates the reduced form parameters $\boldsymbol{\tau}$, $\boldsymbol{a}$ and $\boldsymbol{\sigma}_L = vec(\boldsymbol{\Sigma}_L)$ by equation-by-equation OLS, and then the $N(N-1)/2$ free elements $\boldsymbol{\omega}$ of the orthogonal rotation matrix $\boldsymbol{Q}$ in (3) mapping structural shocks and reduced form innovations by non-Gaussian PML.
2. The two-step estimator in Fiorentini and Sentana (2019), which replaces the inconsistent non-Gaussian PMLEs of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ by the sample means and standard deviations of pseudo-standardised shocks computed using $\hat{\boldsymbol{a}}_T$ and $\hat{\boldsymbol{j}}_T$.

Although the specifications tests that we have proposed in this paper could also be applied to shocks computed on the basis of these alternative estimators, the asymptotic covariance matrices that take into account their sampling variability will differ from the ones we have derived in this paper. Given that some researchers may prefer to use one of those two-step estimation methods, obtaining computationally simple expressions for the adjusted covariance matrix would provide a valuable addition to our results.

In fact, the moment conditions that we consider for testing independence could form the basis of a GMM estimation procedure for the model parameters $\boldsymbol{\theta}$ along the lines of Lanne and Luoto (2021), although with a larger set of third and fourth cross-moments. The overidentification restrictions tests obtained as a by-product of this procedure could be used as a specification test of the assumed independence-like restrictions.

Our tests for normality tackle a single shock at a time. Although we could in principle simultaneously test the normality of two or more shocks by combining the corresponding normality tests, the implicit joint null hypothesis would violate the second identification condition in Assumption 1. The asymptotic distribution of such joint tests constitutes a very interesting topic for further research. In addition, we could formally study the limiting probability of finding $N-1$ rejections of the univariate normality tests in those circumstances.

Another important research topic would be the limiting behaviour of the PMLEs of $\boldsymbol{\theta}$ when Assumption 1 does not hold, either because two or more of the shocks are Gaussian or because they are not independent.

Finally, while the integer product moment tests for independence that we have considered are very intuitive, they may have little power against alternatives in which the dependence is mostly visible in certain regions of the domain of the random shocks. With this in mind, in Amengual et al. (2021b) we study moment tests that look at the product of nonlinear transformations of the shocks, such as $I(q_{\alpha i} \leq \varepsilon_{it} \leq q_{\omega i})$, where

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$q_{\alpha i}$ and $q_{\omega i}$ are the α and ω quantiles of the marginal distribution of the i^{th} shock (with $0 \le \alpha < \omega \le 1$), or $I(k_{li} \le \varepsilon_{it} \le k_{ui})$, where $k_{li} < k_{ui}$ are some fixed values, or indeed $\varepsilon_{it} I(k_{li} \le \varepsilon_{it} \le k_{ui})$. Extending this approach in such a way that it leads to a consistent test of independence constitutes another promising research avenue.

A Proofs

Proposition 1

Under standard regularity conditions [see, e.g. Newey and McFadden (1994)], we can linearise the vector of influence functions underlying our tests around $\boldsymbol{\theta}_0$ so that

$$
\begin{aligned}
\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)] &= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)] + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]}{\partial \boldsymbol{\theta}}\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \\
&\quad + o_p(1) \\
&= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)] + \mathcal{J}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1).
\end{aligned}
$$

But since

$$
\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_0) + o_p(1),
$$

we can combine both expressions to write

$$
\begin{aligned}
\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)] = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)] + \mathcal{J}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\sqrt{T}\frac{1}{T} \\
\times\sum_{t=1}^{T}\boldsymbol{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_0) + o_p(1),
\end{aligned}
$$

whence the asymptotic distribution in the proposition follows. □

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Proposition 2

Fiorentini and Sentana (2021) prove in their "Appendix D" that

$$\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\{\boldsymbol{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]\boldsymbol{Z}_{st}'(\boldsymbol{\theta})\},$$

which in our case reduces to

$$\begin{aligned}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= -\boldsymbol{C}^{-1}\left(\boldsymbol{I}_N \;\; \boldsymbol{y}_{t-1}' \otimes \boldsymbol{I}_N \;\; \ldots \;\; \boldsymbol{y}_{t-p}' \otimes \boldsymbol{I}_N \;\; \boldsymbol{0}_{N \times N^2}\right) \\ &\quad -[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N](\boldsymbol{I}_N \otimes \boldsymbol{C}^{-1})\left(\boldsymbol{0}_{N^2 \times N} \;\; \boldsymbol{0}_{N^2 \times N^2} \;\; \ldots \;\; \boldsymbol{0}_{N^2 \times N^2} \;\; \boldsymbol{I}_{N^2}\right)\end{aligned}$$

in view of (7) and (8). Therefore, it immediately follows that

$$\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}'} = -\boldsymbol{C}^{-1} \text{and} \frac{\partial \varepsilon_{it}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}'} = -\boldsymbol{c}^{i\cdot},$$

where

$$\boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{c}^{1\cdot} \\ \vdots \\ \boldsymbol{c}^{i\cdot} \\ \vdots \\ \boldsymbol{c}^{N\cdot} \end{pmatrix}.$$

Similarly,

$$\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{a}_j'} = -(\boldsymbol{y}_{t-j}' \otimes \boldsymbol{C}^{-1}) \text{and} \frac{\partial \varepsilon_{it}^*(\boldsymbol{\theta})}{\partial \boldsymbol{a}_j'} = -(\boldsymbol{y}_{t-j}' \otimes \boldsymbol{c}^{i\cdot}) \text{ for } j = 1, \ldots, p.$$

Finally,

$$\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{c}'} = -[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{C}^{-1}] \text{ and } \frac{\partial \varepsilon_{it}^*(\boldsymbol{\theta})}{\partial \boldsymbol{c}'} = -[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{c}^{i\cdot}].$$

If we combine these expressions with the fact that

$$\frac{\partial m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]}{\partial \varepsilon_i^*} = I(h_i > 0)\frac{h_i}{\varepsilon_{it}^*}\prod_{i'=1}^{N} \varepsilon_{it}^{*h_{i'}},$$

we obtain the desired results. □

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Proposition 3

General expression (17) follows directly from the definition of the scores for $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$ in (5) and (6) and the law of iterated expectations after exploiting the fact that $m_{\boldsymbol{h}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]$, $\boldsymbol{e}_{lt}(\boldsymbol{\phi}_\infty)$, $\boldsymbol{e}_{lt}(\boldsymbol{\phi}_\infty)$ and $\boldsymbol{e}_{rt}(\boldsymbol{\phi}_\infty)$ are $i.i.d.$ processes with zero mean under our assumptions.

In turn, the more detailed expressions exploit the cross-sectional independence of the shocks. For example, consider

$$\mathcal{F}_{\boldsymbol{h}l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = cov\left\{ m_{\boldsymbol{h}}(\boldsymbol{\varepsilon}_t^*), \left[\begin{array}{c} \partial \ln f(\varepsilon_{1t}^*; \boldsymbol{\varrho}_\infty)/\partial\varepsilon_1^* \\ \vdots \\ \partial \ln f(\varepsilon_{Nt}^*; \boldsymbol{\varrho}_\infty)/\partial\varepsilon_N^* \end{array}\right] \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}.$$

It is clear that row i will be zero if $h_i = 0$ because of the cross-sectional independence of the shocks and the fact that $E[\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_\infty)/\partial\varepsilon_i^*|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = 0$.

The same argument applies to the remaining blocks. □

B Additional material

B.1 Some useful results

As mentioned in Sect. 3, the following lemma provides an easy way to recursively compute some of the ingredients of the independence tests:

Lemma 1 *Let* $[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]^{\otimes k} = \underbrace{\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes ... \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}_{k\ \ times}$ *denote the* k^{th}*-order Kronecker power of the* $N \times 1$ *vector* $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$. *Then, for any* $k \geq 2$

$$d\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]^{\otimes k}\} = \{\boldsymbol{I}_N \otimes [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]^{\otimes k-1}\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{N^{K-1}}]d\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]^{\otimes k-1}\}.$$

Proof The result follows immediately from the product rule for differentials [see section 9.14 in Magnus and Neudecker (2019)] after exploiting the fact that $\boldsymbol{K}_{1N} = \boldsymbol{K}_{N1} = \boldsymbol{I}_N$ and

$$\begin{aligned} vec(\boldsymbol{A}_{m\times n} \otimes \boldsymbol{B}_{p\times q}) &= (\boldsymbol{I}_n \otimes \boldsymbol{K}_{qm} \otimes \boldsymbol{I}_p)[vec(\boldsymbol{A}_{m\times n}) \otimes vec(\boldsymbol{B}_{p\times q})] \\ &= \{\boldsymbol{I}_n \otimes [(\boldsymbol{K}_{qm} \otimes \boldsymbol{I}_p)[\boldsymbol{I}_m \otimes vec(\boldsymbol{B}_{p\times q})]\}vec(\boldsymbol{A}_{m\times n}) \\ &= \{[(\boldsymbol{I}_n \otimes \boldsymbol{K}_{qm})[vec(\boldsymbol{A}_{m\times n}) \otimes \boldsymbol{I}_q] \otimes \boldsymbol{I}_p\}vec(\boldsymbol{B}_{p\times q}), \quad \text{(B1)} \end{aligned}$$

[see section 3.7 in Magnus and Neudecker (2019)]. □

A trivial—but useful—consequence of Lemma 1 that we make extensively use in this paper is:

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Corollary 1 *The differentials of the second, third and fourth powers of the structural shocks will be*

$$d[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = [\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),$$

$$\begin{aligned} d[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = &[\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ &+\{[\boldsymbol{I}_{N^2} \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})][\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ &+[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \end{aligned}$$

and

$$\begin{aligned} d[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = &[\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]d\boldsymbol{\varepsilon}_t^* \\ &+\{[\boldsymbol{I}_N^2 \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})][\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ &+\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{N^2}][\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ &+[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \boldsymbol{I}_N]d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}). \end{aligned}$$

Proof To save space, let $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$. The differential of $\boldsymbol{m}^{cv}(\boldsymbol{\varepsilon}_t^*)$, $\mathrm{d}(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)$, follows directly from Lemma 1.

This lemma also implies that the differential of $\boldsymbol{m}^{cs}(\boldsymbol{\varepsilon}_t^*)$ will be

$$\begin{aligned} \mathrm{d}(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= [\mathrm{d}(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes \boldsymbol{\varepsilon}_t^*] + (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= (\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) + (\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) + (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \end{aligned}$$

Expression (B1) then yields

$$\begin{aligned} (\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= \{(\boldsymbol{K}_{1N} \otimes \boldsymbol{I}_{N^2})[\boldsymbol{I}_N \otimes vec\left(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*\right)]\}vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= (\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)\mathrm{d}\boldsymbol{\varepsilon}_t^*, \\ (\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= \{(\boldsymbol{K}_{1N^2} \otimes \boldsymbol{I}_N)[\boldsymbol{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]vec(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= [\boldsymbol{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]vec(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= [\boldsymbol{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]\{(1 \otimes \boldsymbol{K}_{1N})[vec(\boldsymbol{\varepsilon}_t^*) \otimes 1]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= [(\boldsymbol{I}_{N^2} \otimes \boldsymbol{\varepsilon}_t^*)(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_N)]\mathrm{d}\boldsymbol{\varepsilon}_t^* \end{aligned}$$

and

$$\begin{aligned} (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) &= \{(1 \otimes \boldsymbol{K}_{1N^2})[vec\left(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*\right) \otimes 1] \otimes \boldsymbol{I}_N\}vec(d\boldsymbol{\varepsilon}_t^*) \\ &= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_N)d\boldsymbol{\varepsilon}_t^* \end{aligned}$$

because $\boldsymbol{K}_{1N} = \boldsymbol{K}_{N1} = \boldsymbol{I}_N$.

Finally, Lemma 1 implies that the differential of $\boldsymbol{m}^{ck}(\boldsymbol{\varepsilon}_t^*)$ will be

$$\begin{aligned} \mathrm{d}\left(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*\right) &= [\mathrm{d}(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes \boldsymbol{\varepsilon}_t^*] + (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \\ &= (\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) + (\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \\ &\quad +(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) + (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*). \end{aligned}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Once again, expression (B1) yields

$$\begin{aligned}
(\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= \{1 \otimes (\boldsymbol{K}_{1N} \otimes \boldsymbol{I}_{N^3})[\boldsymbol{I}_N \otimes vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*) \\
&= (\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)\mathrm{d}\boldsymbol{\varepsilon}_t^*, \\
(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= \{1 \otimes (\boldsymbol{K}_{1N^2} \otimes \boldsymbol{I}_{N^2})[\boldsymbol{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)]vec(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) \\
&= (\boldsymbol{I}_N^2 \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_N)\mathrm{d}\boldsymbol{\varepsilon}_t^*, \\
(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) &= [\{(1 \otimes \boldsymbol{K}_{1N^2})[vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes 1]\} \otimes \boldsymbol{I}_N^2]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \\
&= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_{N^2})[1 \otimes \{(\boldsymbol{K}_{1N} \otimes \boldsymbol{I}_N)[\boldsymbol{I}_N \otimes vec(\boldsymbol{\varepsilon}_t^*)]\}]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*) \\
&= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_{N^2})(\boldsymbol{I}_N \otimes \boldsymbol{\varepsilon}_t^*)\mathrm{d}\boldsymbol{\varepsilon}_t^*
\end{aligned}$$

and

$$\begin{aligned}
(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) &= [\{(1 \otimes K_{1N^3})[vec\left(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*\right) \otimes 1]\} \otimes \boldsymbol{I}_N]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*) \\
&= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{I}_N)\mathrm{d}\boldsymbol{\varepsilon}_t^*,
\end{aligned}$$

as desired. □

B.2 Univariate discrete mixtures of normals

B.2.1 Moments

The parameters δ, $\varkappa$ and λ of the two-component Gaussian mixture we consider in Sect. 5 determine the higher-order moments of ε_t^* through the relationship

$$E(\varepsilon_t^{*j}|\boldsymbol{\varrho}) = \lambda E(\varepsilon_t^{*j}|s_t = 1; \boldsymbol{\varrho}) + (1-\lambda)E(\varepsilon_t^{*j}|s_t = 2; \boldsymbol{\varrho}),$$

where $s_t \in \{1, 2\}$ is a Bernoulli random variable with $\Pr(s_t = 1) = \lambda$. Specifically,

$$\begin{aligned}
E(\varepsilon_t^*|s_t = k; \boldsymbol{\varrho}) &= \mu_k^*(\boldsymbol{\varrho}), \\
E(\varepsilon_t^{*2}|s_t = k; \boldsymbol{\varrho}) &= \mu_k^{*2}(\boldsymbol{\varrho}) + \sigma_k^{*2}(\boldsymbol{\varrho}), \\
E(\varepsilon_t^{*3}|s_t = k; \boldsymbol{\varrho}) &= \mu_k^{*3}(\boldsymbol{\varrho}) + 3\mu_k^*(\boldsymbol{\varrho})\sigma_k^{*2}(\boldsymbol{\varrho}), \\
E(\varepsilon_t^{*4}|s_t = k; \boldsymbol{\varrho}) &= \mu_k^{*4}(\boldsymbol{\varrho}) + 6\mu_k^{*2}(\boldsymbol{\varrho})\sigma_k^{*2}(\boldsymbol{\varrho}) + 3\sigma_k^{*4}(\boldsymbol{\varrho}).
\end{aligned}$$

Given that $E(\varepsilon_t^*|\boldsymbol{\varrho}) = 0$ and $E(\varepsilon_t^{*2}|\boldsymbol{\varrho}) = 1$ by construction, straightforward algebra shows that the skewness and kurtosis coefficients will be given by

$$E(\varepsilon_{it}^{*3}|\boldsymbol{\varrho}) = -\frac{\delta(\lambda-1)\lambda[\delta^2\{\lambda[2+\lambda(\varkappa-1)] - \varkappa\} + 3(\varkappa-1)]}{\varkappa + (1-\lambda)\varkappa}$$

and

$$E(\varepsilon_{it}^{*4}|\boldsymbol{\varrho}) = \frac{3\lambda - 2\delta^2(3+\delta^2)\lambda^3 + (6\delta^2 + 8\delta^4)\lambda^4 - 9\delta^4\lambda^5 + 3\delta^4\lambda^6}{[\lambda + (1-\lambda)\varkappa]^2}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

$$+\frac{2\delta^2(1-\lambda)\lambda[3-(1-\lambda)\lambda\{6+\delta^2[2-3(1-\lambda)\lambda]\}]\varkappa}{[\lambda+(1-\lambda)\varkappa]^2}$$
$$+\frac{(1-\lambda)\{3-\delta^2(\lambda-1)^2\lambda[6+\delta^2(-1+3\lambda^2)]\}\varkappa^2}{[\lambda+(1-\lambda)\varkappa]^2}.$$

B.2.2 Score with respect to $\boldsymbol{\varrho}$

Regarding the specific elements that appear in (9) and (10), we have

$$\begin{aligned}\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \varepsilon^*_{it}} &= -\frac{1}{f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}\left\{\lambda_i\frac{\phi_{1it}[\varepsilon^*_{it}(\boldsymbol{\theta})-\mu^*_1(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_1(\boldsymbol{\varrho}_i)}\right.\\
&\left.\quad+(1-\lambda_i)\frac{\phi_{2it}[\varepsilon^*_{it}(\boldsymbol{\theta})-\mu^*_2(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_2(\boldsymbol{\varrho}_i)}\right\}\\
&= -\left\{\lambda_i w_{1it}\frac{[\varepsilon^*_{it}(\boldsymbol{\theta})-\mu^*_1(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_1(\boldsymbol{\varrho}_i)}+(1-\lambda_i)w_{2it}\frac{[\varepsilon^*_{it}(\boldsymbol{\theta})-\mu^*_2(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_2(\boldsymbol{\varrho}_i)}\right\},\end{aligned}$$

where we have defined the posterior probabilities of shock i being drawn from component k at time t as $w_{kit}=\phi[\varepsilon^*_{it}(\boldsymbol{\theta});\mu^*_k(\boldsymbol{\varrho}_i),\sigma^{*2}_k(\boldsymbol{\varrho}_i)]/f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]$ to shorten the expressions [see Boldea and Magnus (2009)].

As for the derivatives with respect to the shape parameters in (11), we have

$$\boldsymbol{e}_{r_it}(\boldsymbol{\phi})=\left[\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \delta_i},\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \varkappa_i},\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \lambda_i}\right]',$$

with

$$\begin{aligned}&\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \delta_i}=\lambda_i(1-\lambda_i)\\
&\times\left\{w_{1it}\left(\frac{\delta_i\lambda_i}{\sigma^{*2}_1(\boldsymbol{\varrho}_i)[\varkappa_i+(1-\lambda_i)\varkappa_i]}-\frac{[1+\delta_i(1-\lambda_i)\varepsilon_{it}]}{1-\delta_i^2\lambda_i(1-\lambda_i)}\frac{[\varepsilon_{it}-\mu^*_1(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_1(\boldsymbol{\varrho}_i)}\right)\right.\\
&\left.+\,w_{2it}\left(\frac{\delta_i(1-\lambda_i)\varkappa_i}{\sigma^{*2}_2(\boldsymbol{\varrho}_i)[\varkappa_i+(1-\lambda_i)\varkappa_i]}-\frac{[1+\delta_i(1-\lambda_i)\varepsilon_{it}]}{1-\delta_i^2\lambda_i(1-\lambda_i)}\frac{[\varepsilon_{2t}-\mu^*_2(\boldsymbol{\varrho}_i)]}{\sigma^{*2}_2(\boldsymbol{\varrho}_i)}\right)\right\},\\
&\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial \varkappa_i}=\frac{\lambda_i(1-\lambda_i)}{2[\varkappa_i+(1-\lambda_i)\varkappa_i]}\\
&\times\left[\left\{-w_{1it}\left\{\frac{[\varepsilon_{it}-\mu^*_1(\boldsymbol{\varrho}_i)]^2}{\sigma^{*2}_1(\boldsymbol{\varrho}_i)}-1\right\}+\frac{w_{2it}}{[\varkappa_i+(1-\lambda_i)\varkappa_i]\varkappa_i}\right.\right.\\
&\left.\left.\left\{\frac{[\varepsilon_{it}-\mu^*_2(\boldsymbol{\varrho}_i)]^2}{\sigma^{*2}_2(\boldsymbol{\varrho}_i)}-1\right\}\right\}\right],\end{aligned}$$

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

and

$$
\begin{aligned}
\frac{\partial \ln f[\varepsilon_{it}^{*}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}{\partial \lambda_i} = & \, w_{1it}\left(1 + \frac{\lambda\{1-\varkappa+\delta^2[\lambda^2(\varkappa-1)+\varkappa-2\lambda\varkappa]\}}{2[1-\delta(1-\lambda)\lambda\}[\lambda(1-\varkappa)+\varkappa]}\right) \\
& -w_{2it}\left(1 - \frac{(1-\lambda)\{1-\varkappa+\delta^2[\lambda^2(\varkappa-1)+\varkappa-2\lambda\varkappa]\}}{2[1-\delta^2(1-\lambda)\lambda][\lambda(1-\varkappa)+\varkappa]}\right) \\
& +w_{1it}\frac{[\varepsilon_{it}-\mu_1^{*}(\boldsymbol{\varrho}_i)]\lambda}{2[1-\delta^2(1-\lambda)\lambda]^2} \times \{\delta[1+3\lambda(-1+\varkappa)-3\varkappa] \\
& -\delta^3(\lambda-1)[\lambda(\varkappa-1)-\varkappa]+\varepsilon_{it}(\varkappa-1)+\varepsilon_{it}\delta^2[\lambda^2(1-\varkappa) \\
& -\varkappa+2\lambda\varkappa]\} \\
& +w_{2it}\frac{[\varepsilon_{it}-\mu_2^{*}(\boldsymbol{\varrho}_i)](1-\lambda)}{2[1-\delta^2(1-\lambda)\lambda]^2\varkappa}\{\varepsilon_{it}(\varkappa-1+\delta^2[\lambda^2-\varkappa \\
& +2\lambda\varkappa-\lambda^2\varkappa)] \\
& +(\delta[2\delta^2\lambda^2(1-\varkappa)+\delta^2\lambda^3(\varkappa-1)-2\varkappa+\lambda(3+\delta^2)\varkappa-3\lambda]\}.
\end{aligned}
$$

The second derivatives of the log-density with respect to the shape parameters can be derived using the chain rule for second derivatives from the expressions in Boldea and Magnus (2009), who obtain them in terms of λ, $\mu_k^{*}(\boldsymbol{\varrho}_i)$ and $\sigma_k^{*2}(\boldsymbol{\varrho}_i)$ $(k = 1, 2)$. The precise expressions are available on request.

C Monte Carlo results for a trivariate static model

In this appendix, we report finite sample results for a trivariate extension of our benchmark DGP 1, which we denote by DGP 1T. Specifically, we generate samples of size T from

$$
\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^{*} \\ \varepsilon_{2t}^{*} \\ \varepsilon_{3t}^{*} \end{pmatrix} \tag{C2}
$$

As for the shocks, we choose $\varepsilon_{1t}^{*} \sim N(0,1)$, $\varepsilon_{2t}^{*} \sim DMN(-.859, .386, 1/5)$ and $\varepsilon_{2t}^{*} \sim DMN(.859, .386, 1/5)$, so that ε_{2t}^{*} and ε_{3t}^{*} follow discrete mixtures of two normals with kurtosis coefficients 4 and skewness coefficients equal to $-.5$ and $.5$, respectively.

Table 9 reports Monte Carlo rejection rates of the normality tests proposed in Sect. 4.1 for samples of size $T = 250$ (top panel) and $T = 1000$ (bottom panel). The first three columns of those panels report rejection rates using asymptotic critical values, while the last three columns show the bootstrap-based ones for $T = 250$ and the warp-speed bootstrap-based ones for $T = 1000$. Once again, the normality tests tend to be oversized at the usual nominal levels, especially for samples of length 250, while the standard bootstrap version of our tests is much more reliable for both the third and fourth moment tests. More importantly, the null of normality is correctly rejected a large number of times when it does not hold, even in samples of length

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 9 Monte Carlo size and power of normality tests: trivariate static model

Nominal size	$T = 250$					
	Asymptotic critical values			Bootstrap (399 samples) critical values		
	10%	5%	1%	10%	5%	1%
	Size (ε^*_{1t} normal)					
$H_3(\varepsilon^*_{1t})$	18.32	11.47	4.32	8.52	3.97	0.68
$H_4(\varepsilon^*_{1t})$	17.58	10.30	4.50	8.67	4.22	1.02
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	19.25	12.48	6.21	8.36	4.00	0.96
	Power (ε^*_{2t} DMN with negative skewness)					
$H_3(\varepsilon^*_{2t})$	81.73	76.37	63.77	73.58	65.53	45.71
$H_4(\varepsilon^*_{2t})$	71.22	64.85	52.56	62.86	53.88	30.68
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{3t})$	85.61	81.26	71.70	77.09	68.14	42.89
	Power (ε^*_{3t} DMN with positive skewness)					
$H_3(\varepsilon^*_{3t})$	82.25	77.25	64.50	73.94	65.78	45.16
$H_4(\varepsilon^*_{3t})$	71.33	64.97	53.06	63.22	53.85	29.73
$H_3(\varepsilon^*_{3t})$ & $H_4(\varepsilon^*_{3t})$	86.00	81.67	71.81	76.97	67.89	41.66
	$T = 1000$					
	Asymptotic critical values			Warp-speed bootstrap critical values		
Nominal size	10%	5%	1%	10%	5%	1%
	Size (ε^*_{1t} normal)					
$H_3(\varepsilon^*_{1t})$	12.32	6.61	1.61	9.69	4.76	0.77
$H_4(\varepsilon^*_{1t})$	12.22	6.56	1.84	9.71	4.71	0.93
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	12.73	6.91	2.10	9.38	4.83	0.81
	Power (ε^*_{2t} DMN with negative skewness)					
$H_3(\varepsilon^*_{2t})$	99.84	99.79	99.50	99.80	99.67	98.84
$H_4(\varepsilon^*_{2t})$	99.32	98.84	97.06	98.75	97.80	92.56
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{3t})$	99.95	99.91	99.83	99.89	99.83	99.39
	Power (ε^*_{3t} DMN with positive skewness)					
$H_3(\varepsilon^*_{3t})$	99.91	99.86	99.53	99.87	99.75	98.90
$H_4(\varepsilon^*_{3t})$	99.25	98.69	96.77	98.63	97.64	92.98
$H_3(\varepsilon^*_{3t})$ & $H_4(\varepsilon^*_{3t})$	99.98	99.95	99.86	99.94	99.89	99.42

Monte Carlo empirical rejection rates of normality tests; 20,000 replications. DGP 1T—Shocks: ε^*_{1t} normal, and ε^*_{2t} and ε^*_{2t} discrete mixture of two normals. See "Appendix C" for details on the data generating process. Asymptotic critical values: $H_3(\cdot) \sim \chi^2_1$, $H_4(\cdot) \sim \chi^2_1$ and $H_3(\cdot)$ & $H_4(\cdot) \sim \chi^2_2$. We describe the sampling procedure we use to implement the bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 10 Monte Carlo size of independence moment tests: trivariate static model

Nominal size	$T = 250$						$T = 1000$					
	Asymptotic critical values			Bootstrap (399 samples) critical values			Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	6.87	3.06	0.45	9.04	4.38	0.90	9.12	4.55	0.93	9.91	5.09	1.04
$E(\varepsilon^*_{1t}\varepsilon^*_{3t})$	7.07	3.10	0.55	9.28	4.53	0.84	9.42	4.65	0.90	10.35	5.09	1.10
$E(\varepsilon^*_{2t}\varepsilon^*_{3t})$	7.17	3.19	0.56	9.27	4.31	0.79	9.65	4.84	0.92	9.87	5.14	1.06
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	8.07	3.81	0.64	9.28	4.63	0.91	9.53	4.54	0.86	9.66	4.73	0.99
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{3t})$	7.54	3.52	0.60	8.77	4.26	0.81	9.44	4.54	0.90	9.45	4.76	0.97
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	10.49	5.21	1.07	11.35	5.84	1.28	10.28	5.29	0.99	10.18	5.12	0.89
$E(\varepsilon^*_{1t}\varepsilon^*_{2t}\varepsilon^*_{3t})$	9.46	4.66	0.95	10.37	5.12	1.12	9.95	4.85	1.02	10.53	5.05	1.05
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{3t})$	10.43	5.18	0.95	11.30	5.70	1.22	10.29	5.10	1.15	10.12	5.08	1.22
$E(\varepsilon^{*2}_{2t}\varepsilon^*_{3t})$	9.43	4.81	1.11	10.43	5.34	1.30	9.54	4.76	1.03	9.88	4.81	0.93
$E(\varepsilon^*_{2t}\varepsilon^{*2}_{3t})$	9.51	4.75	1.03	10.39	5.13	1.07	9.64	4.76	1.03	10.48	4.90	0.99
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	6.75	3.04	0.55	8.68	4.16	0.77	8.78	4.32	0.87	9.63	4.75	1.02
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{3t})$	6.30	2.85	0.61	8.48	3.88	0.81	9.09	4.50	0.89	9.86	4.98	1.10
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	7.01	3.19	0.73	9.55	4.82	1.01	9.32	4.45	0.88	10.06	4.82	0.88
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t}\varepsilon^*_{3t})$	6.79	3.31	0.82	9.10	4.47	0.97	9.36	4.54	0.92	10.54	5.34	0.89

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Table 10 continued

Nominal size	$T = 250$						$T = 1000$					
	Asymptotic critical values			Bootstrap (399 samples) critical values			Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{3t})$	7.04	3.13	0.77	9.65	4.71	0.91	8.80	4.13	0.84	9.68	4.63	0.76
$E(\varepsilon^{*}_{1t}\varepsilon^{*3}_{2t})$	8.27	3.90	0.68	9.96	5.05	1.01	10.19	4.98	0.99	10.61	5.11	0.93
$E(\varepsilon^{*}_{1t}\varepsilon^{*2}_{2t}\varepsilon^{*}_{3t})$	7.89	4.19	1.06	9.91	5.09	1.08	8.99	4.53	1.04	9.82	4.83	1.04
$E(\varepsilon^{*}_{1t}\varepsilon^{*}_{2t}\varepsilon^{*2}_{3t})$	7.24	3.61	0.91	9.15	4.56	1.00	9.50	4.91	1.11	10.04	5.02	0.97
$E(\varepsilon^{*}_{1t}\varepsilon^{*3}_{3t})$	8.55	4.16	0.83	10.26	5.18	1.19	9.45	4.75	0.94	9.87	4.86	1.01
$E(\varepsilon^{*3}_{2t}\varepsilon^{*}_{3t})$	7.64	3.85	0.86	9.66	4.81	1.01	9.45	5.09	1.34	10.04	5.29	1.09
$E(\varepsilon^{*2}_{2t}\varepsilon^{*2}_{2t})$	7.32	3.51	1.29	9.91	5.03	1.20	8.87	4.38	1.10	9.95	4.98	1.00
$E(\varepsilon^{*}_{2t}\varepsilon^{*3}_{3t})$	7.38	3.66	0.88	9.42	4.54	0.91	9.35	4.83	1.14	9.88	5.14	0.99
Covariance	5.89	2.42	0.33	8.77	4.25	0.68	9.39	4.68	0.94	10.47	5.20	1.14
Co-skewness	8.74	4.48	1.15	10.57	5.45	1.30	9.46	4.77	0.91	9.78	4.79	0.86
Co-kurtosis	6.93	4.13	1.53	9.33	4.64	1.01	9.10	5.41	1.85	9.73	4.83	0.93
Joint test	6.62	3.84	1.51	8.71	4.30	1.05	9.14	5.28	1.67	9.84	4.67	0.84

Monte Carlo empirical rejection rates of independence tests; 20,000 replications. DGP 1T—Shocks: ε^*_{1t} normal, and ε^*_{2t} and ε^*_{2t} discrete mixture of two normals. See "Appendix C" for details on the data generating process. We present the asymptotic distribution of the test statistics in Sect. 5.2.2 and describe the sampling procedure we use to implement the bootstrap in Sect. 5.1.3

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

250. Nevertheless, there is a moderate loss of power relative to Table 2, which may reflect the need to estimate almost twice as many parameters as in the bivariate case. In larger dimensions, one might expect a similar pattern, although in general, the main determinants of the power of our normality test will be the non-normality of the structural shock under consideration and how precisely identified it is.

Finally, in Table 10 we report the Monte Carlo rejection rates of the tests we have proposed in Sect. 4.2 under the null of independence for samples of size $T = 250$ (left panel) and $T = 1000$ (right panel). As in Table 9, the first (last) three columns of those panels report rejection rates using asymptotic (bootstrapped) critical values. As in the bivariate case (cf. Table 4), we can see some small to moderate finite sample size distortion when $T = 250$, although in almost all cases they are corrected by the bootstrap. Finite sample sizes improve considerably for samples of length 1000, as expected.

References

Almuzara M, Amengual D, Sentana E (2019) Normality tests for latent variables. Quant Econ 10:981–1017

Amengual D, Fiorentini G, Sentana E (2021a) Multivariate Hermite polynomials and information matrix tests. CEMFI Working Paper 2103

Amengual D, Fiorentini G, Sentana E (2021b) Specification tests for non-Gaussian structural vector autoregressions. Work in progress

Beran R (1988) Prepivoting test statistics: a bootstrap view of asymptotic refinements. J Am Stat Assoc 83:687–697

Blanchard OJ, Quah D (1989) The dynamic effects of aggregate demand and supply disturbances. Am Econ Rev 79:655–673

Boldea O, Magnus JR (2009) Maximum likelihood estimation of the multivariate normal mixture model. J Am Stat Assoc 104:1539–1549

Bontemps C, Meddahi N (2005) Testing normality: a GMM approach. J Econom 124:149–186

Comon P (1994) Independent component analysis, a new concept? Signal Process 36:287–314

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Dolado JJ, Motyovszki G, Pappa E (2020) Monetary policy and inequality under labor market frictions and capital-skill complementarity. Am Econ J Macroecon 13:292–332

Faust J (1998) The robustness of identified Var conclusions about money. Carnegie-Rochester Conf Ser Public Policy 49:207–244

Fiorentini G, Sentana E (2019) Consistent non-Gaussian pseudo maximum likelihood estimators. J Econom 213:321–358

Fiorentini G, Sentana E (2020) Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. CEMFI Working Paper 2023

Fiorentini G, Sentana E (2021) Specification tests for non-Gaussian maximum likelihood estimators. Quant Econ 12:683–742

Giacomini R, Politis DN, White H (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Economet Theor 29:567–589

Golub GH, van Loan CF (2013) Matrix computations, 4th edn. Johns Hopkins, Baltimore

Gouriéroux C, Monfort A, Renne J-P (2017) Statistical inference for independent component analysis. J Econom 196:111–126

Herwartz H (2018) Hodges-Lehmann detection of structural shocks—an analysis of macroeconomic dynamics in the euro area. Oxford Bull Econ Stat 80:736–754

Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Netw 10:626–634

Hyvärinen A (2013) Independent component analysis: recent advances. Philos Trans R Soc A 371:20110534

SPI Journal: 13209 Article No.: 0247 TYPESET DISK LE CP Disp.:2021/10/30 Pages: 46 Layout: Small-Ex

Ilmonen P, Paindaveine D (2011) Semiparametrically efficient inference based on signed ranks in symmetric independent component models. Ann Stat 39:2448–2476

Jarque CM, Bera AK (1980) Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals. Econ Lett 6:255–259

Lanne M, Luoto J (2021) GMM estimation of non-Gaussian structural vector autoregressions. J Bus Econ Stat 39:69–81

Lanne M, Lütkepohl H (2010) Structural vector autoregressions with nonnormal residuals. J Bus Econ Stat 28:159–168

Lanne M, Meitz M, Saikkonen P (2017) Identification and estimation of non-Gaussian structural vector autoregressions. J Econom 196:288–304

Magnus JR, Neudecker H (2019) Matrix differential calculus with applications in Statistics and Econometrics, 3rd edn. Wiley, New York

Magnus JR, Sentana E (2020) Zero-diagonality as a linear structure. Econ Lett 196:109513

Magnus JR, Pijls HGJ, Sentana E (2021) The Jacobian of the exponential function. J Econ Dyn Control 127:104122

Meijer E (2005) Matrix algebra for higher order moments. Linear Algebra Appl 410:112–134

Mencía J, Sentana E (2012) Distributional tests in multivariate dynamic models with Normal and Student *t* innovations. Rev Econ Stat 94:133–152

Mertens K, Ravn MO (2012) The dynamic effects of personal and corporate income tax changes in the United States. Am Econ Rev 103:1212–1247

Moneta A, Pallante G (2020) Identification of structural VAR models via Independent Component Analysis: a performance evaluation study. LEM Working Paper 2020/44, Scuola Superiore Sant'Anna

Newey WK (1985) Maximum likelihood specification testing and conditional moment tests. Econometrica 53:1047–70

Newey WK, McFadden DL (1994) Large sample estimation and hypothesis testing. In: Engle RF, McFadden DL (eds) Handbook of econometrics, vol IV. Elsevier, New York, pp 2111–2245

Phillips PCB, Durlauf SN (1986) Multiple time series regression with integrated processes. Rev Econ Stud 53:473–495

Sentana E, Fiorentini G (2001) Identification, estimation and testing of conditionally heteroskedastic factor models. J Econom 102:143–164

Sims CA (1980) Macroeconomics and reality. Econometrica 48:1–48

Stock JH, Watson MW (2018) Identification and estimation of dynamic causal effects in macroeconomics using external instruments. Econ J 28:917–948

Tauchen G (1985) Diagnostic testing and evaluation of maximum likelihood models. J Econom 30:415–443

Uhlig H (2005) What are the effects of monetary policy on output? Results from an agnostic identification procedure. J Monet Econ 52:381–419

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal: 13209 Article: 247

Author Proof

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check and confirm the corresponding author is correctly identified.	
2.	Please check the layout of Tables 1–10, and correct if necessary.	
3.	Please specify the significance of the symbol [italics] reflected inside Table [3] by providing a description in the form of a table footnote. Otherwise, kindly amend if deemed necessary.	