working paper

2102

Moment tests of
independent components

Dante Amengual
Gabriele Fiorentini
Enrique Sentana

February 2021

cemf

Moment tests of independent components

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 extensive simulation exercises and discuss some bootstrap procedures. We also show that our tests have non-negligible power against a variety of empirically plausible alternatives.

Dante Amengual
CEMFI
amengual@cemfi.es

Gabriele Fiorentini
Università di Firenze and RCEA
gabriele.fiorentini@unifi.it

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

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) and Blanchard and Quah (1989)), sign restrictions (see, e.g., Faust (1998) and 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, Motyosvski and Pappa (2020)). Recently, identification through independent non-Gaussian shocks has become increasingly popular after Lanne, Meitz and Saikkonen (2017) and Gouriéroux, Monfort and Renne (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 $\mathbf{y}$ – 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 $\mathbf{I}_N$ without loss of generality, namely

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

then the matrix $\mathbf{C}$ 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 $\mathbf{y}$ 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) $\mathbf{C}$ 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 $V(\mathbf{y}) = \mathbf{C}\mathbf{C}'$ but not $\mathbf{C}$ 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 $\boldsymbol{\varepsilon}^{**} = \mathbf{Q}\boldsymbol{\varepsilon}^*$, where $\mathbf{Q}$ 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 $\mathbf{C}$ is due to the fact that $\boldsymbol{\varepsilon}^*$ and $\boldsymbol{\varepsilon}^{**}$ share not only their mean vector ($\mathbf{0}$) and covariance matrix ($\mathbf{I}$), but also the same non-linear 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

---

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

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 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, Amengual and Sentana (2019), though, we focus on the latent shocks rather than the observed variables in view of the fact that the identifying Assumption 1 is written in terms of $\boldsymbol{\varepsilon}^*$ rather than $\mathbf{y}$.

If we knew the true values of $\boldsymbol{\mu}$ and $\mathbf{C}$, $\boldsymbol{\mu}_0$ and $\mathbf{C}_0$ say, with $rank(\mathbf{C}_0) = N$, 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 $\mathbf{C}$ 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 implicitly based correspond to the pseudo log-likelihood function scores, 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 for not only for the 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 section 3, and particularise them to assess normality and independence in section 4. Next, section 5 contains the results of our Monte Carlo experiments. We present our conclusions and suggestions for further research in section 6, and relegate some technical material to several appendices.

2 Structural vector autoregressions

2.1 Model specification

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

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

where I_{t-1} is the information set, $\mathbf{C}$ 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 $\boldsymbol{\varepsilon}_t = \mathbf{C}\boldsymbol{\varepsilon}_t^*$ denote the reduced form innovations, so that $\boldsymbol{\varepsilon}_t|I_{t-1} \sim i.i.d. \ (\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$. As we mentioned in the 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 $\mathbf{C}$ are only identified up to an orthogonal transformation. Specifically, we can use the so-called LQ matrix decomposition[2] to relate the matrix $\mathbf{C}$ to the Cholesky decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_L \boldsymbol{\Sigma}_L'$ as

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

where $\mathbf{Q}$ 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 $|\mathbf{Q}| = 1$.[3] Notice that if $|\mathbf{Q}| = -1$ instead, we can change the sign of the ith structural shock and its impact multipliers in the ith column of the matrix $\mathbf{C}$ 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 non-zero odd moments.

In this context, Lanne, Meitz and Saikkonen (2017) show that statistical identification of both the structural shocks and $\mathbf{C}$ (up to column permutations and sign changes) is possible under

---

[2]The LQ decomposition is intimately related to the QR decomposition. Specifically, $\mathbf{Q}'\boldsymbol{\Sigma}_L'$ provides the QR decomposition of the matrix $\mathbf{C}'$, 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 9 of Magnus, Pijls and Sentana (2020) 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.

the 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 $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\mathbf{A}_1), \ldots, vec'(\mathbf{A}_p), vec'(\mathbf{C})]' = (\boldsymbol{\tau}', \mathbf{a}_1', \ldots, \mathbf{a}_p', \mathbf{c}') = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$ denote the structural parameters characterising the first two conditional moments of $\mathbf{y}_t$. 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. 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 $q = \sum_i q_i$, 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 t $(t = 1, \ldots, T)$ will be given by

$$l_t(\mathbf{y}_t; \boldsymbol{\phi}) = -\ln|\mathbf{C}| + \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}] = -\ln|\mathbf{C}| + \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N], \quad (4)$$

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

2.2.2 The score vector

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

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

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

where

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

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

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

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

and

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

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 the SVAR model (2) generates a covariance stationary process.[4] Consider the reparametrisation $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix whose elements contain the scale of the structural shocks, while the columns of $\mathbf{J}$, 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 $\mathbf{a}_i = vec(\mathbf{A}_i)$ and $\mathbf{j} = veco(\mathbf{J})$ 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. $\mathbf{a}_{i\infty} = \mathbf{a}_{i0}$ and $\mathbf{j}_\infty = \mathbf{j}_0$. 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 $\boldsymbol{\psi}$

---

[4]If the autoregressive polynomial $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)$ had some unit roots, $\mathbf{y}_t$ 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.

which simultaneously estimate $\boldsymbol{\varrho}$ will be consistent too when the univariate distributions used for estimation purposes are discrete mixtures of normals, in which 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 $\boldsymbol{\theta} = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$.

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

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

and

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

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

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

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, Fiorentini and Sentana (2021).

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 $\mathbf{C}$ that appear in the SVAR (2) critically hinges on the validity of the 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 $\mathbf{C}$ 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 equiv-

---

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

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

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

where $\mathbf{j} = \{j_1, ..., j_N\}$, with $j_i \in \mathbb{Z}_{0+}$, denotes the index vector characterising a specific product moment. While the influence function in (14) will generally require the estimation of $E[\varepsilon_{it}^{*j_i}(\boldsymbol{\theta}_0)]$ for some of the shocks, the constant term $\prod_{i=1}^N E[\varepsilon_{it}^{*j_i}(\boldsymbol{\theta}_0)]$ 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, Fiorentini and Sentana (2021) discuss in detail how to deal with the estimation of the required $E[\varepsilon_{it}^{*j_i}(\boldsymbol{\theta}_0)]$ in the general case.

Although we have motivated (14) as the basis for our tests of independence, by setting all the elements of $\mathbf{j}$ 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 $j_i = 3$ 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 $\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})]$ denote a $K \times 1$ vector containing a collection of influence functions $m_{\mathbf{j}^k}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ of the form (14) for different index vectors $\mathbf{j}^1, \ldots, \mathbf{j}^k, , \ldots, \mathbf{j}^K$. 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 $\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})]$ 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*

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^T \mathbf{m}[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)] \to N[0, \mathcal{W}(\phi_\infty; \varphi_0)],$$

where

$$\begin{aligned} \mathcal{W}(\phi_\infty; \varphi_0) &= \mathcal{V}(\phi_\infty; \varphi_0) + \mathcal{J}(\phi_\infty; \varphi_0)\mathcal{A}^{-1}(\phi_\infty; \varphi_0)\mathcal{B}(\phi_\infty; \varphi_0)\mathcal{A}^{-1}(\phi_\infty; \varphi_0)\mathcal{J}'(\phi_\infty; \varphi_0) \\ &\quad + \mathcal{F}(\phi_\infty; \varphi_0)\mathcal{A}^{-1}(\phi_\infty, \boldsymbol{v}_0)\mathcal{J}'(\phi_\infty; \varphi_0) + \mathcal{J}(\phi_\infty; \varphi_0)\mathcal{A}^{-1}(\phi_\infty; \varphi_0)\mathcal{F}'(\phi_\infty; \varphi_0), \end{aligned}$$

$$\mathcal{V}(\phi; \varphi) = V\left\{ \mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})] \middle| \varphi \right\},$$

$$\mathcal{J}(\phi; \varphi) = E\left\{ \frac{\partial \mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})]}{\partial \phi'} \middle| \varphi \right\},$$

$$\mathcal{F}(\phi; \varphi) = cov\left\{ \frac{\partial \mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})]}{\partial \phi'}, \mathbf{s}_{\phi t}(\phi) \middle| \varphi \right\}$$

and $\mathcal{A}(\phi_\infty; \varphi_0)$ and $\mathcal{B}(\phi_\infty; \varphi_0)$ are defined in (12) and (13), respectively.

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

3.2.1 Covariance across influence functions

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

$$cov\{m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta})], m_{\mathbf{j}'}[\varepsilon_t^*(\boldsymbol{\theta})]|\varphi\} = E\{m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta})]m_{\mathbf{j}'}[\varepsilon_t^*(\boldsymbol{\theta})]|\varphi\} - E\{m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta})]|\varphi\}E\{m_{\mathbf{j}'}[\varepsilon_t^*(\boldsymbol{\theta})]|\varphi\}.$$

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

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

we obtain

$$cov\{m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)], m_{\mathbf{j}'}[\varepsilon_t^*(\boldsymbol{\theta}_0)]|\varphi_0\} = \prod_{i=1}^N E\left[\varepsilon_{it}^{*(j_i+j_i')}\right] - \prod_{i=1}^N E(\varepsilon_{it}^{*j_i})E(\varepsilon_{it}^{*j_i'}). \qquad (15)$$

3.2.2 The expected Jacobian

Straightforward application of the chain rule implies that

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

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

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

$$\mathrm{J}_{\mathbf{j}\boldsymbol{\tau}}(\boldsymbol{\varrho}_{i\infty}, \varphi_0) = E\left[\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\frac{\partial \varepsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\tau}'}\bigg|\varphi_0\right] = -E\left[\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\bigg|\varphi_0\right]\mathbf{C}_0^{-1},$$

$$\mathrm{J}_{\mathbf{j}\mathbf{a}_i}(\boldsymbol{\varrho}_{i\infty}, \varphi_0) = E\left[\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\frac{\partial \varepsilon_t^*(\boldsymbol{\theta}_0)}{\partial \mathbf{a}_i'}\bigg|\varphi_0\right] = -E\left[\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\bigg|\varphi_0\right]\left[E(\mathbf{y}_{t-i}'|\varphi_0)\otimes\mathbf{C}_0^{-1}\right]$$

and

$$\mathrm{J}_{\mathbf{j}\mathbf{c}}(\boldsymbol{\varrho}_{i\infty}, \varphi_0) = E\left[\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\frac{\partial \varepsilon_t^*(\boldsymbol{\theta}_0)}{\partial \mathbf{c}'}\bigg|\varphi_0\right] = -E\left\{\frac{\partial m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial \varepsilon^{*\prime}}\left[\varepsilon_t(\boldsymbol{\theta}_0)\otimes\mathbf{C}_0^{-1}\right]\bigg|\varphi_0\right\}.$$

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

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

where $\mathbf{D}_N$, $\mathbf{T}_N$ and $\mathbf{Q}_N$ are the duplication, triplication and quadruplication matrices, respectively (see Meijer (2005) for details), the following lemma provides an easy way to compute all those relevant quantities recursively:

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

$$d\{[\varepsilon_t^*(\boldsymbol{\theta})]^{\otimes k}\} = \{\mathbf{I}_N \otimes [\varepsilon_t^*(\boldsymbol{\theta})]^{\otimes k-1}\} d\varepsilon_t^*(\boldsymbol{\theta}) + [\varepsilon_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] d\{[\varepsilon_t^*(\boldsymbol{\theta})]^{\otimes k-1}\}.$$

A trivial – but useful – consequence of Lemma 1 that we will extensively use in our subsequent analysis is:

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

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

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

and

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

3.2.3   The covariance with the score

Let ℓ_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 $m_{\mathbf{j}}(\cdot)$ and the pseudo log-likelihood scores evaluated at the (pseudo) true values is given by*

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

where

$$\mathcal{F}_{\mathbf{j}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = \left[\begin{array}{c} \mathcal{F}_{\mathbf{j}l}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \\ \mathcal{F}_{\mathbf{j}s}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \\ \mathcal{F}_{\mathbf{j}r}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \end{array} \right] \left[\begin{array}{ccc} \mathbf{Z}'_{lt}(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{Z}'_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{array} \right],$$

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

$$\mathrm{F}_{\mathbf{j}l(i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -cov\left\{ m_{\mathbf{j}}[\varepsilon_t^*(\boldsymbol{\theta}_0)], \frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_0); \boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*} \bigg| \boldsymbol{\varphi}_0 \right\}$$

9

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

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

and zero otherwise, and finally

$$\mathcal{F}_{\mathbf{j}r}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = \mathrm{F}_{\mathbf{j}r}'(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) \ell_N,$$

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

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

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.

It turns out, though, 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 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 the log-likelihood function (4) for any given values of $\mathbf{a}$ and $\mathbf{j}$ 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.

4.2 Testing independence

At first sight, the arguments at the end of 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 $\mathbf{C}$ 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

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

In other words, we are simply assessing if the sample correlation between the ith 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:

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

and

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

depending on whether they involve two or three shocks.

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

pair of shocks, namely

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

and

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

three shocks

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

and even four shocks

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

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 (2019), 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 non-zero 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 non-zero elements of the covariance matrix of $\mathbf{m}^{cv}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$ are

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

In turn, in the case of $\mathbf{m}^{cs}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$ and $\mathbf{m}^{ck}[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})]$, the non-zero 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

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

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^*\varepsilon_{i''t}^*, \varepsilon_{i't}^{*2}\varepsilon_{it}^*\varepsilon_{i''t}^*) = E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),$$

$$cov(\varepsilon_{it}^{*3}\varepsilon_{i't}^*, \varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) = E(\varepsilon_{it}^{*5})E(\varepsilon_{i't}^{*3}),$$

$$cov(\varepsilon_{it}^{*3}\varepsilon_{i't}^*, \varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) = E(\varepsilon_{it}^{*5})E(\varepsilon_{i't}^{*3}),$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^*\varepsilon_{i''t}^*, \varepsilon_{i't}^{*2}\varepsilon_{it}^*\varepsilon_{i''t}^*) = E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}, \varepsilon_{it}^{*2}\varepsilon_{i''t}^{*2}) = E(\varepsilon_{it}^{*4}) - 1,$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i''t}^*, \varepsilon_{i't}^{*2}\varepsilon_{i''t}^*) = 1,$$

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

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

$$cov(\varepsilon_{it}^*\varepsilon_{i't}^*, \varepsilon_{it}^{*2}\varepsilon_{i't}^*) = E(\varepsilon_{it}^{*3}),$$

$$cov(\varepsilon_{it}^*\varepsilon_{i't}^*, \varepsilon_{it}^{*3}\varepsilon_{i't}^*) = E(\varepsilon_{it}^{*4}),$$

$$cov(\varepsilon_{it}^*\varepsilon_{i't}^*, \varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) = E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*3}),$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^*, \varepsilon_{it}^{*3}\varepsilon_{i't}^*) = E(\varepsilon_{it}^{*5}),$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^*, \varepsilon_{i't}^{*3}\varepsilon_{it}^*) = E(\varepsilon_{it}^{*3})E(\varepsilon_{i't}^{*4}), \text{ and}$$

$$cov(\varepsilon_{it}^{*2}\varepsilon_{i't}^*, \varepsilon_{it}^{*2}\varepsilon_{i't}^{*2}) = E(\varepsilon_{it}^{*4})E(\varepsilon_{it}^{*3}).$$

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

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

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

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

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

while for those in (21) we get

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

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

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

13

while for (23) we get

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

and

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

Similarly, the expected Jacobian of (24) involves

$$\mathrm{J_{\mathbf{j}\tau}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = \mathbf{0}, \ \mathrm{J_{\mathbf{ja}_k}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = \mathbf{0} \text{ and } \mathrm{J_{\mathbf{jc}}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = -(\mathbf{e}_{i'}' \otimes \mathbf{c}_0^{i''\cdot}) - (\mathbf{e}_{i''}' \otimes \mathbf{c}_0^{i'\cdot}).$$

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

$$\mathrm{J_{\mathbf{j}\tau}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = \mathbf{0}, \ \mathrm{J_{\mathbf{ja}_k}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = \mathbf{0} \text{ and } \mathrm{J_{\mathbf{jc}}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0) = \mathbf{0}.$$

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 $\mathbf{j}$ is such that $j_i = 1$, we will have

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

and

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

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

As a result, (17) will be zero for the second moments $E(\varepsilon_{it}^* \varepsilon_{i't}^*)$, except for $\mathrm{F_{\mathbf{j}s(i,i')}}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\varphi}_0)$, 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 non-zero covariance elements for the co-skewness influence functions $E(\varepsilon_{it}^{*2} \varepsilon_{i't}^*)$ will be

$$\mathrm{F_{\mathbf{j}l(i')}}(\boldsymbol{\varrho}_{\infty}, \boldsymbol{\varphi}_0) = 1, \ \mathrm{F_{\mathbf{j}s(i,i')}}(\boldsymbol{\varrho}_{\infty}, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \varepsilon_{it}^{*2}\right], \ \mathrm{F_{\mathbf{j}s(i',i)}}(\boldsymbol{\varrho}_{\infty}, \boldsymbol{\varphi}_0) = E(\varepsilon_{it}^{*3}),$$

$$\mathrm{F_{\mathbf{j}s(i',i')}}(\boldsymbol{\varrho}_{\infty}, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{i't}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_{i'}^*} \varepsilon_{i't}^{*2}\right] \text{ and } \mathrm{F_{\mathbf{j}r(i')}}(\boldsymbol{\varrho}_{\infty}, \boldsymbol{v}_0) = E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \varepsilon_{it}^*\right],$$

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 $E(\varepsilon_{it}^{*2} \varepsilon_{i't}^{*2})$, the only

non-zero terms are

$$\mathrm{F}_{\mathbf{j}l(i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \varepsilon_{it}^{*2}\right], \ \mathrm{F}_{\mathbf{j}s(i,i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \varepsilon_{it}^{*3}\right],$$

$$\mathrm{F}_{\mathbf{j}s(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \varepsilon_{it}^{*2}\right] \ \text{and} \ \mathrm{F}_{\mathbf{j}r(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \varepsilon_{it}^*\right].$$

In turn, we end up with

$$\mathrm{F}_{\mathbf{j}l(i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = E(\varepsilon_{it}^{*3}), \ \mathrm{F}_{\mathbf{j}s(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \varepsilon_{it}^{*3}\right],$$

$$\mathrm{F}_{\mathbf{j}s(i',i)}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0 = E(\varepsilon_{it}^{*3}), \ \mathrm{F}_{\mathbf{j}s(i',i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -E[\varepsilon_{it}^{*3}]E\left[\frac{\partial \ln f(\varepsilon_{i't}^*; \boldsymbol{\varrho}_{i'\infty})}{\partial \varepsilon_{i'}^*} \varepsilon_{i't}^{*2}\right]$$

and

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

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

In contrast, the only non-zero covariance of the co-kurtosis influence functions $E(\varepsilon_{it}^*\varepsilon_{i't}^*\varepsilon_{i''t}^{*2})$ with the scores will be $\mathrm{F}_{\mathbf{j}s(i,i')}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = 1$ 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.

5 Monte Carlo analysis

In this section, we assess the finite sample size and power of the normality and independence tests discussed in section 4 by means of an extensive Monte Carlo simulation exercise. 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, but provide relevant insights that can be extrapolated to general mulvariate designs. Specifically, we generate samples of size T from the following bivariate static process

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

with $\tau_1 = 1$, $\tau_2 = -1$, $c_{11} = 1$, $c_{12} = .5$, $c_{21} = 0$ and $c_{22} = 2$. Nevertheless, our PML estimation procedure does not exploit the restriction that the loading matrix of the shocks is upper triangular. We consider both $T = 250$, which is realistic in most macro applications with monthly or quarterly data, and $T = 1,000$, which is representative of financial applications with

daily data.

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

$$\varepsilon_{it}^* = \begin{cases} N[\mu_1^*(\boldsymbol{\varrho}_i), \sigma_1^{*2}(\boldsymbol{\varrho}_i)] \text{ with probability } \lambda_i \\ N[\mu_2^*(\boldsymbol{\varrho}_i), \sigma_2^{*2}(\boldsymbol{\varrho}_i)] \text{ with probability } 1 - \lambda_i \end{cases} \tag{26}$$

with

$$\begin{aligned} \mu_1^*(\boldsymbol{\varrho}_i) &= \delta_i(1 - \lambda_i), \\ \mu_2^*(\boldsymbol{\varrho}_i) &= -\delta_i\lambda_i, \\ \sigma_1^{*2}(\boldsymbol{\varrho}_i) &= \frac{1 - \lambda_i(1 - \lambda_i)\delta_i^2}{\lambda_i + (1 - \lambda_i)\varkappa_i}, \\ \sigma_2^{*2}(\boldsymbol{\varrho}_i) &= \varkappa_i\sigma_1^{*2}(\boldsymbol{\varrho}_i), \end{aligned}$$

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.[7]

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

$$l[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i] = \ln\{\lambda_i \cdot \phi[\varepsilon_{it}^*(\boldsymbol{\theta}); \mu_1^*(\boldsymbol{\varrho}_i), \sigma_1^{*2}(\boldsymbol{\varrho}_i)] + (1 - \lambda_i) \cdot \phi[\varepsilon_{it}^*(\boldsymbol{\theta}); \mu_2^*(\boldsymbol{\varrho}_i), \sigma_2^{*2}(\boldsymbol{\varrho}_i)]\},$$

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 $\mathbf{C}$ 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 $\sigma_1^{*2}(\boldsymbol{\varrho}_i)$. 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 require $\lambda_i \in (0, 1)$ to avoid degenerate mixtures.[8]

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 $\mathbf{C}$, $\overline{\boldsymbol{\tau}}_{FICA}$ and $\overline{\mathbf{C}}_{FICA}$ say, using the *FastICA* algorithm of Gävert,

---

[7]We can trivially extend this procedure to three or more components if we replace the normal random variable in the first branch of (26) by a k-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\varrho})$ and $\sigma_1^{*2}(\boldsymbol{\varrho})$, respectively, so that the resulting random variable will be a $(k + 1)$-component Gaussian mixture with zero mean and unit variance.

[8]Effectively, we impose $\varkappa_i \in [\underline{\varkappa}, 1]$ with $\underline{\varkappa} = .0001$, and $\lambda_i \in [\underline{\lambda}, \overline{\lambda}]$ with $\underline{\lambda} = .001$ and $\overline{\lambda} = .999$.

Hurri, Särelä, and Hyvärinen.[9] In addition, we obtain initial values of the shape parameters of each shock by performing 20 iterations[10] of the expectation maximisation (EM) algorithm in Dempster, Laird and Rubin (1977) on each of the elements of $\overline{\boldsymbol{\varepsilon}}^*_{t,FICA} = \overline{\mathbf{C}}^{-1}_{FICA}(\mathbf{y}_t - \overline{\boldsymbol{\tau}}_{FICA})$.

As we mentioned in section 2.2, Assumption 1 only guarantees the identification of $\mathbf{C}$ 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 $\mathbf{C}$ using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne, Meitz and Saikkonen (2017). In addition, we impose that $diag(\mathbf{C})$ 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.

5.1.2 DGPs under the null and the alternative

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

DGP 1: 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 2: A normal distribution and the first discrete mixture of normals in DGP 1, i.e. $\varepsilon^*_{1t} \sim N(0,1)$ 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 4.10 and $-.54$, respectively, so that $\beta = -1.354$ and $\nu = 18.718$ in the notation in Mencía and Sentana (2012).

The left panels of Figures 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:

---

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

[10]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 achieves the right balance between CPU time and convergence of the parameters.

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

DGP 5: Bivariate asymmetric t with skewness vector $\boldsymbol{\beta} = -5\ell_2$ 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 Figures 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 under the alternative. Additionally, we evaluate the small sample size and power of the normality tests presented in section 4.1 using the results from the simulation designs DGP 2 (null), and DGP 1 and DGP 3 (alternative).

5.1.3 Bootstrap procedures

To improve the finite sample reliability of the testing procedures we have described in sections 4.1 and 4.2, we consider simple bootstrap procedures.

In the case 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.[11] First, we generate NT draws R_{is} from a discrete uniform distribution between 1 and T, which we then use to construct

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

---

[11]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.

where $\tilde{\varepsilon}_{is}^* = \hat{\varepsilon}_{iR_{is}}^*$ and $\hat{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T) = \hat{\mathbf{C}}_T^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\tau}}_T)$ 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 ith shock $\tilde{\varepsilon}_{is}^*$ are independently drawn from a $N(0,1)$ distribution while the draws for the remaining shocks $\tilde{\varepsilon}_{js}^*$ ($j \neq i$) 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. For this reason, we often rely on the warp-speed method of Giacomini, Politis and White (2013). Nevertheless, we explicitly evaluate the performance of the standard bootstrap in a couple of designs to assess whether the performance of the warp-speed approach is representative.

5.2 Simulation results

5.2.1 Testing normality

As mentioned earlier, failure of the second condition in Assumption 1 results in an under-identified model. For that reason, it is of the utmost importance to test the normality of each structural shock, in the hope that one only fails to reject the null at most once.

Table 1 reports Monte Carlo rejection rates of the tests proposed in section 4.1 for DGP 1, 2 and 3. As can be seen, the null of normality is correctly rejected most of the time 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 panel DGP 2, which is the only one with a Gaussian shock, show that the normality tests tend to be oversized at the usual nominal levels, especially for samples of length 250. For that reason, we have generated another $10,000$ samples of length $T = 250$ under DGP 2 in which we generate $N_{Boot} = 399$ bootstrap samples at each Monte Carlo replication, as described in section 5.1.3. In addition, we store the first of these 399 samples to check the reliability of the warp-speed bootstrap procedure.

Table 2 shows that the standard bootstrap version of our tests is very accurate for both the third and fourth moment tests. The warp-speed version performs comparably well, although it seems to be somewhat less accurate.[12]

---

[12] Given 10,000 Monte Carlo replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels, respectively.

5.2.2 Testing independence

In Tables 3 ($T = 250$) and 4 ($T = 1,000$) we report the Monte Carlo rejection rates of the tests we have proposed in section 4.2 under the null of independence. Specifically, we look at the second, third and fourth moment individual tests in $\mathbf{m}^{cv}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$, $\mathbf{m}^{cs}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ and $\mathbf{m}^{ck}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$, 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 warp-speed bootstrap-based ones.[13]

We can see in Table 3 some small to moderate finite sample size distortion when $T = 250$, although in most cases they are corrected by the warp-speed bootstrap. The only exception seems to be DGP 2, which displays a small sample size distortion even with this procedure. Given that in this design there is only one non-Gaussian shock, a plausible explanation is that the identification of $\mathbf{C}$ may be weaker, a conjecture we will revisit in the next section. For this reason, we have performed additional simulation exercises with the standard bootstrap procedure (with 399 samples) for both DGP 2 and DGP 3. The results in Table 5 clearly show that the usual bootstrap version of the tests, which is the relevant one in empirical applications, has much better size properties.

As expected, finite sample sizes improve considerably for $T = 1,000$, when even the empirical rejection rates based on asymptotic critical values become generally very close to the nominal ones. Indeed, the bootstrap versions of the tests seem unnecessary for this sample size.

Next, we assess the power of the independence tests for $T = 250$ and $T = 1,000$ 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.[14] On the other hand, power is mostly coming from the co-skewness component (20) in the case of the joint asymmetric t. Still, the test based on the covariance of the 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 the second moment (19) also has non-negligible power in this case.

In summary, although the rejection rates naturally depend on the type of departure from the

---

[13]In all our *i.i.d.* desings, 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.

[14]In Amengual, Fiorentini and Sentana (2021), we propose moment tests based on non-linear transformations of the shocks which may have more power in this context.

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 the 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 $\mathbf{C}$ 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 $\mathbf{C}$ in the 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 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 extensive 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 the introduction, there are many estimators for the parameters of the

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, Monfort and Renne (2017), which first estimates the reduced form parameters $\boldsymbol{\tau}$, $\mathbf{a}$ 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 $\mathbf{Q}$ 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 $\hat{\mathbf{a}}_T$ and $\hat{\mathbf{j}}_T$.

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 (2019), 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 combine two or more of those 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 also study the 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, Fiorentini and Sentana (2021) we study moment tests that look at the product of non-linear transformations of the shocks, such as $I(q_{\alpha i} \leq \varepsilon_{it} \leq q_{\omega i})$, where $q_{\alpha i}$ and $q_{\omega i}$ the α and ω quantiles of the marginal distribution of the ith shock (with $0 \leq \alpha < \omega \leq 1$), $I(k_{li} \leq \varepsilon_{it} \leq k_{ui})$, where $k_{li} < k_{ui}$ and some fixed values, or indeed $\varepsilon_{it} I(k_{li} \leq \varepsilon_{it} \leq k_{ui})$. Extending this approach in such a way that it leads to a consistent test of independence constitutes another promising research avenue.

References

Almuzara, M., Amengual, D. and Sentana, E. (2019): "Normality tests for latent variables", *Quantitative Economics* 10, 981–1017.

Amengual, D., Fiorentini, G. and Sentana, E. (2021): "Specification tests for non-Gaussian structural vector autoregressions", work in progress.

Blanchard, O.J. and Quah, D. (1989): "The dynamic effects of aggregate demand and supply disturbances", *American Economic Review* 79, 655–673.

Boldea, O. and Magnus, J.R. (2009): "Maximum likelihood estimation of the multivariate normal mixture model", *Journal of the American Statistical Association* 104, 1539–1549.

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

Comon, P. (1994), "Independent component analysis, a new concept?", *Signal Processing* 36, 287–314.

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

Dolado, J.J., Motyovszki, G. and Pappa, E. (2020): "Monetary policy and inequality under labor market frictions and capital-skill complementarity", forthcoming in the *American Economic Journal: Macroeconomics*.

Faust, J. (1998): "The robustness of identified VAR conclusions about money", *Carnegie-Rochester Conference Series in Public Policy* 49, 207–244.

Fiorentini, G. and Sentana, E. (2019): "Consistent non-Gaussian pseudo maximum likelihood estimators", *Journal of Econometrics* 213, 321–358.

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

Fiorentini, G. and Sentana, E. (2021): "Specification tests for non-Gaussian maximum likelihood estimators", forthcoming in *Quantitative Economics*.

Giacomini, R., Politis D.N. and White, H. (2013): "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators", *Econometric Theory* 29, 567–589.

Golub, G.H. and van Loan, C.F. (2013): *Matrix computations*, 4th edition, Johns Hopkins.

Gouriéroux, C., Monfort, A. and Renne, J.-P. (2017), "Statistical inference for independent component analysis", *Journal of Econometrics* 196, 111–126.

Herwartz, H. (2018): "Hodges-Lehmann detection of structural shocks - an analysis of macroeconomic dynamics in the euro area", *Oxford Bulletin of Economics and Statistics* 80, 736–754.

Hyvärinen, A. (1999): "Fast and robust fixed-point algorithms for independent component

analysis", *IEEE transactions on Neural Networks* 10, 626–634.

Hyvärinen, A. (2013): "Independent component analysis: recent advances", *Philosophical Transactions of The Royal Society A* 371, 20110534.

Ilmonen, P. and Paindaveine, D. (2011): "Semiparametrically efficient inference based on signed ranks in symmetric independent component models", *Annals of Statistics* 39, 2448–2476.

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

Lanne, M. and Luoto, J. (2019): "GMM estimation of non-Gaussian structural vector autoregressions", forthcoming in the *Journal of Business and Economic Statistics*.

Lanne, M. and Lütkepohl, H. (2010): "Structural vector autoregressions with nonnormal residuals", *Journal of Business and Economic Statistics* 28, 159–168.

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

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in Statistics and Econometrics*, 3rd edition, Wiley.

Magnus, J.R., Pijls, H.G.J. and Sentana, E. (2020): "The Jacobian of the exponential function", CEMFI Working Paper 2005.

Magnus, J.R. and Sentana, E. (2020): "Zero-diagonality as a linear structure", *Economics Letters* 196, 109513.

Meijer, E. (2005): "Matrix algebra for higher order moments", *Linear Algebra and its Applications* 410, 112–134.

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

Mertens, K. and Ravn, M.O. (2012): "The dynamic effects of personal and corporate income tax changes in the United States", *American Economic Review* 103, 1212–1247.

Moneta, A. and 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, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047–70.

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Phillips, P.C.B. and Durlauf, S.N. (1986): "Multiple time series regression with integrated processes", *Review of Economic Studies* 53, 473-495.

Sentana, E. and Fiorentini, G. (2001): "Identification, estimation and testing of conditionally heteroskedastic factor models", *Journal of Econometrics* 102, 143-164.

Sims, C.A. (1980): "Macroeconomics and reality", *Econometrica* 48, 1–48.

Stock, J. H. and Watson, M. W. (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments", *Economic Journal* 28, 917–948.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415–443.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure", *Journal of Monetary Economics* 52, 381–419.

Appendices

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

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\mathbf{m}[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)] = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta}_0)] + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta}_0)]}{\partial\boldsymbol{\theta}}\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) + o_p(1)$$

$$= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta}_0)] + \mathcal{J}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) + o_p(1).$$

But since

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

we can combine both expressions to write

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

whence the asymptotic distribution in the proposition follows. $\square$

Proposition 2

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

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

which in our case reduces to

$$\frac{\partial\varepsilon_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = -\mathbf{C}^{-1}\left(\begin{array}{ccccc}\mathbf{I}_N & \mathbf{y}_{t-1}'\otimes\mathbf{I}_N & \cdots & \mathbf{y}_{t-p}'\otimes\mathbf{I}_N & \mathbf{0}_{N\times N^2}\end{array}\right)$$

$$-[\varepsilon_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N](\mathbf{I}_N\otimes\mathbf{C}^{-1})\left(\begin{array}{ccccc}\mathbf{0}_{N^2\times N} & \mathbf{0}_{N^2\times N^2} & \cdots & \mathbf{0}_{N^2\times N^2} & \mathbf{I}_{N^2}\end{array}\right)$$

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

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

where

$$\mathbf{C}^{-1} = \left(\begin{array}{c}\mathbf{c}^{1\cdot}\\ \vdots\\ \mathbf{c}^{i\cdot}\\ \vdots\\ \mathbf{c}^{N\cdot}\end{array}\right).$$

Similarly,

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

Finally,

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

If we combine these expressions with the fact that

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

we obtain the desired results. $\qquad\qquad\square$

Lemma 1

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

$$
\begin{aligned}
vec(\mathbf{A}_{m \times n} \otimes \mathbf{B}_{p \times q}) &= (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)[vec(\mathbf{A}_{m \times n}) \otimes vec(\mathbf{B}_{p \times q})] \\
&= \{\mathbf{I}_n \otimes [(\mathbf{K}_{qm} \otimes \mathbf{I}_p)[\mathbf{I}_m \otimes vec(\mathbf{B}_{p \times q})]]\}vec(\mathbf{A}_{m \times n}) \\
&= \{[(\mathbf{I}_n \otimes \mathbf{K}_{qm})[vec(\mathbf{A}_{m \times n}) \otimes \mathbf{I}_q] \otimes \mathbf{I}_p\}vec(\mathbf{B}_{p \times q}), \qquad (A1)
\end{aligned}
$$

(see section 3.7 in Magnus and Neudecker (2019)). $\qquad\qquad\square$

Corollary 1

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

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

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

Expression (A1) then yields

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

$$(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) = \{(\mathbf{K}_{1N^2} \otimes \mathbf{I}_N)[\mathbf{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]vec(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*)$$
$$= [\mathbf{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]vec(\boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*)$$
$$= [\mathbf{I}_{N^2} \otimes vec(\boldsymbol{\varepsilon}_t^*)]\{(1 \otimes \mathbf{K}_{1N})[vec(\boldsymbol{\varepsilon}_t^*) \otimes 1]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*)$$
$$= [(\mathbf{I}_{N^2} \otimes \boldsymbol{\varepsilon}_t^*)(\boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_N)]\mathrm{d}\boldsymbol{\varepsilon}_t^*$$

and

$$(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) = \{(1 \otimes \mathbf{K}_{1N^2})[vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes 1] \otimes \mathbf{I}_N\}vec(d\boldsymbol{\varepsilon}_t^*)$$
$$= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_N)d\boldsymbol{\varepsilon}_t^*$$

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

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

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

Once again, expression (A1) yields

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

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

$$(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) = [\{(1 \otimes \mathbf{K}_{1N^2})[vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes 1]\} \otimes \mathbf{I}_N^2]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*)$$
$$= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_{N^2})[1 \otimes \{(\mathbf{K}_{1N} \otimes \mathbf{I}_N)[\mathbf{I}_N \otimes vec(\boldsymbol{\varepsilon}_t^*)]\}]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*)$$
$$= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_{N^2})(\mathbf{I}_N \otimes \boldsymbol{\varepsilon}_t^*)\mathrm{d}\boldsymbol{\varepsilon}_t^*$$

and

$$(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathrm{d}\boldsymbol{\varepsilon}_t^*) = [\{(1 \otimes K_{1N^3})[vec(\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^*) \otimes 1]\} \otimes \mathbf{I}_N]vec(\mathrm{d}\boldsymbol{\varepsilon}_t^*)$$
$$= (\boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_N)\mathrm{d}\boldsymbol{\varepsilon}_t^*,$$

as desired. $\qquad\square$

Proposition 3

The 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 $m_{\mathbf{j}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)]$, $\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)$, $\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$ 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

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

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

The same argument applies to the remaining blocks. $\qquad\square$

B   Univariate discrete mixtures of normals

B.1   Moments

The parameters δ, $\varkappa$ and λ of the two-component Gaussian mixture we consider in section 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,

$$E(\varepsilon_t^*|s_t = k; \boldsymbol{\varrho}) = \mu_k^*(\boldsymbol{\varrho}),$$

$$E(\varepsilon_t^{*2}|s_t = k; \boldsymbol{\varrho}) = \mu_k^{*2}(\boldsymbol{\varrho}) + \sigma_k^{*2}(\boldsymbol{\varrho}),$$

$$E(\varepsilon_t^{*3}|s_t = k; \boldsymbol{\varrho}) = \mu_k^{*3}(\boldsymbol{\varrho}) + 3\mu_k^*(\boldsymbol{\varrho})\sigma_k^{*2}(\boldsymbol{\varrho}),$$

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

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

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

B.2 Score with respect to $\boldsymbol{\varrho}$

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

$$\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}{\partial \varepsilon_{it}^*} = -\frac{1}{f[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}\left\{\lambda_i \frac{\phi_{1it}[\varepsilon_{it}^*(\boldsymbol{\theta}) - \mu_1^*(\boldsymbol{\varrho}_i)]}{\sigma_1^{*2}(\boldsymbol{\varrho}_i)} + (1-\lambda_i)\frac{\phi_{2it}[\varepsilon_{it}^*(\boldsymbol{\theta}) - \mu_2^*(\boldsymbol{\varrho}_i)]}{\sigma_2^{*2}(\boldsymbol{\varrho}_i)}\right\}$$
$$= -\left\{\lambda_i w_{1it}\frac{[\varepsilon_{it}^*(\boldsymbol{\theta}) - \mu_1^*(\boldsymbol{\varrho}_i)]}{\sigma_1^{*2}(\boldsymbol{\varrho}_i)} + (1-\lambda_i)w_{2it}\frac{[\varepsilon_{it}^*(\boldsymbol{\theta}) - \mu_2^*(\boldsymbol{\varrho}_i)]}{\sigma_2^{*2}(\boldsymbol{\varrho}_i)}\right\},$$

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

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

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

with

$$\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}{\partial \delta_i} = \lambda_i(1-\lambda_i)$$
$$\times \left\{w_{1it}\left(\frac{\delta_i\lambda_i}{\sigma_1^{*2}(\boldsymbol{\varrho}_i)[\varkappa_i + (1-\lambda_i)\varkappa_i]} - \frac{[1 + \delta_i(1-\lambda_i)\varepsilon_{it}]}{1 - \delta_i^2\lambda_i(1-\lambda_i)}\frac{[\varepsilon_{it} - \mu_1^*(\boldsymbol{\varrho}_i)]}{\sigma_1^{*2}(\boldsymbol{\varrho}_i)}\right)\right.$$
$$\left. + w_{2it}\left(\frac{\delta_i(1-\lambda_i)\varkappa_i}{\sigma_2^{*2}(\boldsymbol{\varrho}_i)[\varkappa_i + (1-\lambda_i)\varkappa_i]} - \frac{[1 + \delta_i(1-\lambda_i)\varepsilon_{it}]}{1 - \delta_i^2\lambda_i(1-\lambda_i)}\frac{[\varepsilon_{2t} - \mu_2^*(\boldsymbol{\varrho}_i)]}{\sigma_2^{*2}(\boldsymbol{\varrho}_i)}\right)\right\},$$

$$\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}{\partial \varkappa_i} = \frac{\lambda_i(1-\lambda_i)}{2[\varkappa_i + (1-\lambda_i)\varkappa_i]}$$
$$\times \left[\left\{-w_{1it}\left\{\frac{[\varepsilon_{it} - \mu_1^*(\boldsymbol{\varrho}_i)]^2}{\sigma_1^{*2}(\boldsymbol{\varrho}_i)} - 1\right\} + \frac{w_{2it}}{[\varkappa_i + (1-\lambda_i)\varkappa_i]\varkappa_i}\left\{\frac{[\varepsilon_{it} - \mu_2^*(\boldsymbol{\varrho}_i)]^2}{\sigma_2^{*2}(\boldsymbol{\varrho}_i)} - 1\right\}\right\}\right],$$

and

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

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, $\mu_j^*(\boldsymbol{\varrho}_i)$ and $\sigma_j^{*2}(\boldsymbol{\varrho}_i)$ $(j = 1, 2)$. The precise expressions are available on request.

Table 1: Monte Carlo size and power of normality tests.

| Nominal size | Asymptotic critical values | | | | | |
| | Sample size $T = 250$ | | | Sample size $T = 1{,}000$ | | |
	10%	5%	1%	10%	5%	1%
	DGP 1 – Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN					
$H_3(\varepsilon^*_{1t})$	84.36	78.73	64.33	99.94	99.85	99.55
$H_4(\varepsilon^*_{1t})$	70.53	64.07	51.13	99.28	98.69	96.72
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	86.54	81.92	71.58	99.97	99.94	99.88
$H_3(\varepsilon^*_{2t})$	85.14	79.63	65.82	99.91	99.86	99.60
$H_4(\varepsilon^*_{2t})$	70.86	64.31	51.46	99.39	98.87	96.84
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	87.34	82.88	72.26	99.98	99.97	99.88
	DGP 2 – Shocks: ε^*_{1t} normal & ε^*_{2t} DMN					
$H_3(\varepsilon^*_{1t})$	14.06	8.11	2.59	11.03	5.96	1.32
$H_4(\varepsilon^*_{1t})$	12.34	6.93	2.95	10.38	5.32	1.38
$H_3(\varepsilon^*_{1t})$ & $H_4(\varepsilon^*_{1t})$	13.51	8.38	3.96	10.56	5.76	1.67
$H_3(\varepsilon^*_{2t})$	82.86	77.34	63.90	99.93	99.88	99.50
$H_4(\varepsilon^*_{2t})$	70.34	63.87	51.09	99.26	98.79	96.80
$H_3(\varepsilon^*_{2t})$ & $H_4(\varepsilon^*_{2t})$	85.58	81.05	70.91	99.95	99.94	99.90
	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

Notes: 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, $\varepsilon^*_{1t} \sim DMN(-.859, .386, 1/5)$ 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). Asymptotic critical values: $H_3(\cdot) \sim \chi^2_1$, $H_4(\cdot) \sim \chi^2_1$ and $H_3(\cdot)$ & $H_4(\cdot) \sim \chi^2_2$.

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

DGP 2 – Shocks: ε^*_{1t} normal & ε^*_{2t} DMN

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values			Bootstrap (399 samples) critical values		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Size (ε^*_{1t} normal)									
$H_3(\varepsilon^*_{1t})$	13.73	7.71	2.44	9.07	4.17	0.30	9.42	4.58	1.16
$H_4(\varepsilon^*_{1t})$	12.44	6.99	2.82	9.49	4.21	0.47	9.96	4.91	1.17
Power (ε^*_{2t} DMN)									
$H_3(\varepsilon^*_{2t})$	83.31	77.86	64.61	80.45	74.69	55.50	79.89	73.44	55.53
$H_4(\varepsilon^*_{2t})$	71.02	64.90	52.17	67.92	60.32	38.49	68.24	60.72	38.05

Notes: Monte Carlo empirical rejection rates of normality tests; 10,000 replications. DMN denotes discrete mixture of two normals. Data generated according to DGP 2, that is, $\varepsilon^*_{1t} \sim N(0,1)$ 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: $H_3(\cdot) \sim \chi^2_1$ and $H_4(\cdot) \sim \chi^2_1$. We present the asymptotic distribution of the test statistics in section 5.2.2 and describe the sampling procedure we use to implement both, the bootstrap based in 399 samples and Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

Table 3: Monte Carlo size of independence moment tests: Sample size $T = 250$.

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%
DGP 1 – Shocks: ε_{1t}^* DMN & ε_{2t}^* DMN						
$E(\varepsilon_{1t}^* \varepsilon_{2t}^*)$	7.11	3.28	0.46	9.88	5.13	0.85
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^*)$	9.62	4.71	0.97	9.86	4.82	0.60
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*2})$	8.49	4.03	0.69	9.12	4.40	0.72
$E(\varepsilon_{1t}^{*3} \varepsilon_{2t}^*)$	7.42	3.74	0.90	9.51	4.88	1.06
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*3})$	7.22	3.49	0.88	9.28	4.50	0.99
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^{*2})$	6.80	3.26	1.14	9.63	4.51	1.01
Co-skewness	8.93	4.60	1.04	9.41	4.58	0.64
Co-kurtosis	6.48	3.69	1.44	9.06	4.61	1.09
Joint test	6.84	4.00	1.47	8.79	4.25	0.92
DGP 2 – Shocks: ε_{1t}^* normal & ε_{2t}^* DMN						
$E(\varepsilon_{1t}^* \varepsilon_{2t}^*)$	6.76	2.92	0.36	9.43	4.29	0.56
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^*)$	7.10	3.12	0.43	7.40	3.15	0.39
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*2})$	9.98	5.04	0.92	9.50	4.53	0.60
$E(\varepsilon_{1t}^{*3} \varepsilon_{2t}^*)$	6.41	2.77	0.42	8.40	3.84	0.56
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*3})$	8.29	3.98	0.69	10.06	5.15	1.00
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^{*2})$	7.15	3.04	0.64	10.17	5.01	0.84
Co-skewness	8.06	3.78	0.58	7.75	3.31	0.36
Co-kurtosis	5.90	2.93	0.80	8.70	4.21	0.78
Joint test	5.40	2.81	0.89	7.34	3.33	0.64
DGP 3 – Shocks: ε_{1t}^* asymmetric t & ε_{2t}^* Student t						
$E(\varepsilon_{1t}^* \varepsilon_{2t}^*)$	6.86	3.19	0.60	9.38	4.66	0.90
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^*)$	10.24	5.51	1.31	9.98	4.88	0.78
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*2})$	8.53	4.06	0.78	8.39	3.90	0.60
$E(\varepsilon_{1t}^{*3} \varepsilon_{2t}^*)$	7.22	3.54	0.87	10.03	4.99	1.09
$E(\varepsilon_{1t}^* \varepsilon_{2t}^{*3})$	7.13	3.54	0.78	9.39	4.59	0.96
$E(\varepsilon_{1t}^{*2} \varepsilon_{2t}^{*2})$	7.03	3.67	1.15	9.86	5.13	0.92
Co-skewness	9.22	4.87	1.17	8.71	4.12	0.57
Co-kurtosis	6.58	3.89	1.43	9.80	4.80	1.00
Joint test	7.34	4.31	1.60	8.67	4.30	0.95

Notes: 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 section 5.2.2 and describe the sampling procedure we use to implement Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

Table 4: Monte Carlo size of independence moment tests: Sample size $T = 1,000$.

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%
DGP 1 – Shocks: ε_{1t}^* DMN & ε_{2t}^* DMN						
$E(\varepsilon_{1t}^*\varepsilon_{2t}^*)$	9.53	4.67	1.10	10.24	5.10	1.20
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	9.92	4.96	1.05	10.57	5.64	1.28
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	10.17	5.00	1.12	10.42	5.38	1.22
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	9.78	4.96	1.21	10.42	5.14	1.20
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	9.22	4.88	1.31	10.14	5.17	1.24
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	8.98	4.48	1.13	9.74	5.01	0.91
Co-skewness	9.76	5.08	1.29	10.58	5.59	1.44
Co-kurtosis	9.00	4.98	1.78	10.43	5.26	1.05
Joint test	9.50	5.29	1.81	10.77	5.58	1.27
DGP 2 – Shocks: ε_{1t}^* normal & ε_{2t}^* DMN						
$E(\varepsilon_{1t}^*\varepsilon_{2t}^*)$	9.26	4.47	0.90	10.76	5.25	0.81
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	9.77	4.79	0.91	9.98	4.84	0.89
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	10.23	5.47	1.11	10.37	5.46	0.95
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	9.03	4.55	0.81	9.92	4.93	0.87
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	9.71	4.92	0.99	10.66	5.29	1.00
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	9.24	4.50	0.90	10.03	5.21	1.08
Co-skewness	9.79	5.08	1.03	10.09	5.10	0.83
Co-kurtosis	8.45	4.38	1.15	10.34	5.04	0.83
Joint test	8.52	4.37	1.15	10.35	4.97	0.87
DGP 3 – Shocks: ε_{1t}^* asymmetric t & ε_{2t}^* Student t						
$E(\varepsilon_{1t}^*\varepsilon_{2t}^*)$	9.34	4.64	1.06	11.16	5.73	1.34
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	10.15	5.23	1.33	10.87	5.46	1.24
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	9.08	4.32	0.90	9.41	4.67	0.94
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	9.14	4.84	1.26	10.37	5.31	1.33
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	9.29	4.75	1.23	10.73	5.34	1.19
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	8.65	4.30	1.29	10.02	5.10	1.11
Co-skewness	9.30	4.92	1.21	9.87	5.15	1.05
Co-kurtosis	8.69	5.05	1.88	10.61	5.68	1.47
Joint test	8.98	5.27	1.93	10.85	5.79	1.31

Notes: 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 section 5.2.2 and describe the sampling procedure we use to implement Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

Table 5: Monte Carlo size of independence moment tests with bootstrap: Sample size $T = 250$.

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values			Bootstrap (399 samples) critical values		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP 2 – Shocks: ε^*_{1t} normal & ε^*_{2t} DMN									
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	7.12	3.08	0.38	9.98	4.29	0.76	10.10	4.98	0.85
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	7.82	3.30	0.54	8.29	3.45	0.46	8.34	3.54	0.71
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	10.22	4.94	0.82	9.60	4.59	0.66	10.31	4.95	1.03
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	6.30	2.62	0.46	8.03	3.24	0.64	8.85	3.83	0.68
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	8.88	4.12	0.50	10.42	4.90	0.77	10.58	5.50	0.93
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	7.01	2.84	0.64	9.93	4.54	0.77	9.93	4.67	0.84
Co-skewness	8.01	3.73	0.72	8.01	3.57	0.51	8.32	4.07	0.83
Co-kurtosis	5.79	2.84	0.82	8.75	3.95	0.70	8.59	4.16	0.90
Joint test	5.50	3.06	0.91	6.77	3.46	0.67	7.54	3.65	0.96
DGP 3 – Shocks: ε^*_{1t} asymmetric t & ε^*_{2t} Student t									
$E(\varepsilon^*_{1t}\varepsilon^*_{2t})$	6.18	2.67	0.51	8.71	4.08	0.74	9.12	4.32	0.72
$E(\varepsilon^{*2}_{1t}\varepsilon^*_{2t})$	10.47	5.33	1.29	9.79	4.65	0.93	10.43	5.23	1.15
$E(\varepsilon^*_{1t}\varepsilon^{*2}_{2t})$	18.18	3.92	0.81	8.57	3.94	0.58	8.34	4.00	0.70
$E(\varepsilon^{*3}_{1t}\varepsilon^*_{2t})$	6.82	3.15	0.81	9.36	4.24	1.09	9.61	4.53	0.95
$E(\varepsilon^*_{1t}\varepsilon^{*3}_{2t})$	6.63	3.36	0.69	8.43	4.70	1.00	8.49	4.37	0.85
$E(\varepsilon^{*2}_{1t}\varepsilon^{*2}_{2t})$	7.32	3.45	1.26	9.97	5.01	1.15	9.80	4.79	1.30
Co-skewness	8.85	4.78	1.08	8.66	4.29	0.64	8.77	4.50	0.96
Co-kurtosis	6.44	3.93	1.51	9.23	5.01	1.27	9.08	4.83	1.28
Joint test	6.97	4.07	1.55	8.51	4.47	1.17	8.37	4.33	1.19

Notes: 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 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 section 5.2.2 and describe the sampling procedure we use to implement both, the bootstrap based in 399 samples and Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

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

Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%
DGP 4 – Joint Student t						
$E(\varepsilon_{1t}^*\varepsilon_{2t}^*)$	6.90	3.32	0.68	10.80	5.36	1.28
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	9.80	5.10	1.10	11.42	6.16	1.22
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	10.02	5.12	1.04	10.94	5.88	1.12
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	8.50	4.84	1.40	11.86	6.00	1.50
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	8.92	5.18	1.70	11.80	6.66	1.84
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	25.38	16.34	6.54	25.12	16.06	4.56
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	19.64	12.54	4.58	20.54	12.80	4.56
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	14.46	9.68	3.52	16.94	11.02	3.56
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	14.14	9.02	3.52	17.90	11.44	4.88
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^*)$	25.26	17.34	7.80	26.44	18.16	6.50
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*2})$	28.00	20.26	9.50	29.44	20.22	7.54
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^*)$	28.48	21.00	10.92	30.90	20.48	7.46
$E(\varepsilon_{1t}^*\varepsilon_{2t}^{*3})$	34.60	26.26	15.26	36.22	25.14	9.14
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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

Notes: Monte Carlo empirical rejection rates of independence tests; 5,000 replications. Details on the data generating processes: DGP 4, joint Student t: $(\varepsilon_{1t}^*, \varepsilon_{2t}^*) \sim t(\mathbf{0}, \mathbf{I}_2, 6)$; DGP 5, $(\varepsilon_{1t}^*, \varepsilon_{2t}^*) \sim At(\mathbf{0}, \mathbf{I}_2, -5\ell_2, 16)$ (see Mencía and Sentana (2012) for details); and DGP 6, $(\varepsilon_{1t}^*, \varepsilon_{2t}^*) \sim DMN_{LL}(.1, .2, 1/5)$ (see section 5.1.2 for details). We present the asymptotic distribution of the test statistics in section 5.2.2 and describe the sampling procedure we use to implement Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

Table 7: Monte Carlo power of independence moment tests: Sample size $T = 1,000$.

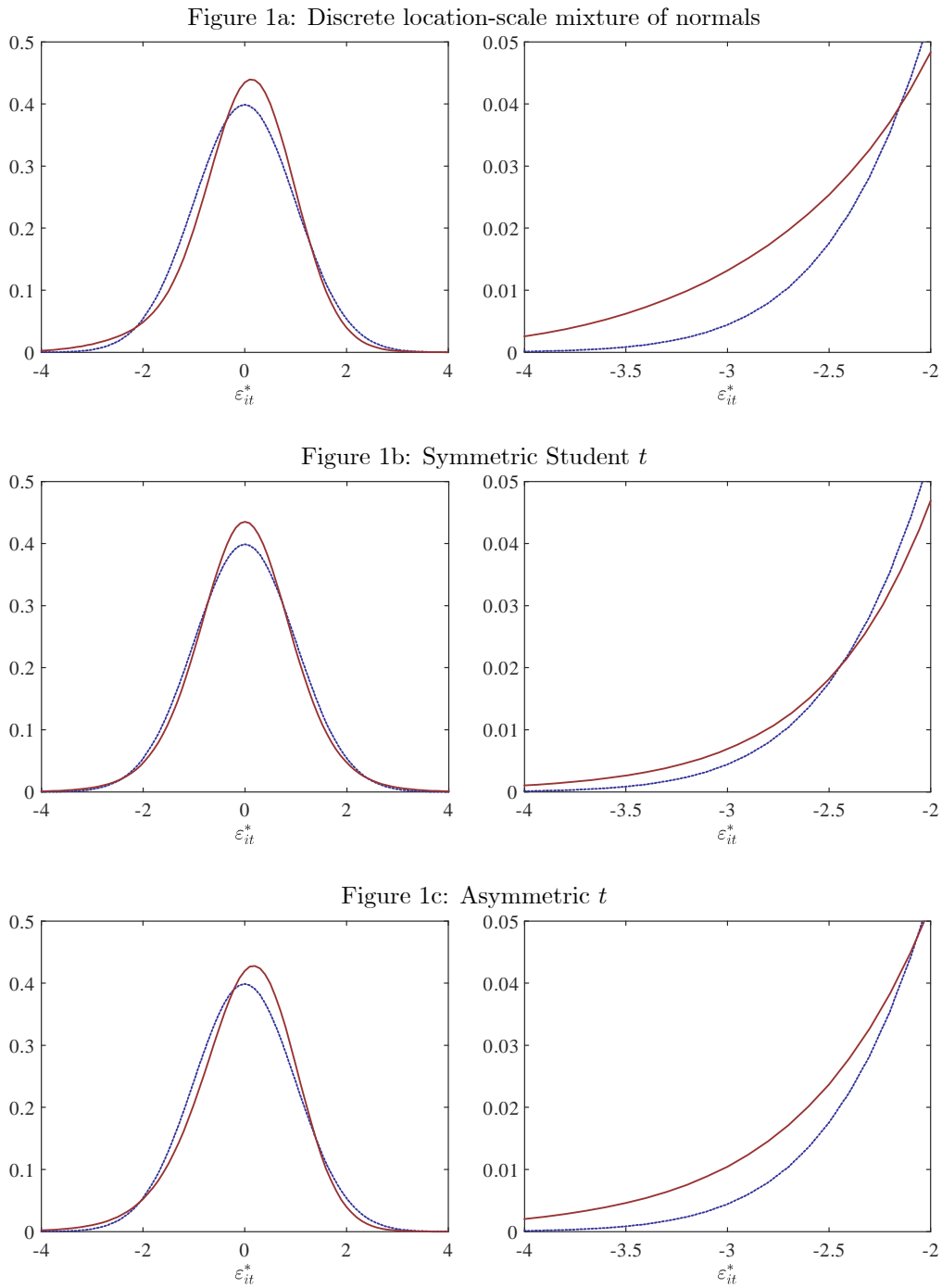| Nominal size	Asymptotic critical values			Warp-speed bootstrap critical values		
	10%	5%	1%	10%	5%	1%
DGP 4 – Joint Student t						
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*})$	15.72	10.04	2.82	17.36	11.26	3.30
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*})$	16.02	9.10	2.86	16.32	9.82	2.86
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*2})$	15.74	9.44	2.90	15.98	9.74	3.18
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^{*})$	18.68	12.44	5.42	20.94	13.02	4.96
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*3})$	19.30	12.42	4.94	20.14	12.78	4.48
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*})$	69.28	64.76	56.38	69.78	65.38	55.58
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*2})$	98.72	98.28	96.98	98.72	98.24	96.62
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^{*})$	56.36	50.28	40.08	57.54	50.08	39.96
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*3})$	65.62	59.52	48.36	66.02	59.62	45.64
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*})$	29.00	20.16	9.72	29.88	20.50	9.12
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*2})$	33.28	24.64	12.68	32.74	23.92	12.02
$E(\varepsilon_{1t}^{*3}\varepsilon_{2t}^{*})$	46.70	38.44	26.34	47.42	37.76	23.24
$E(\varepsilon_{1t}^{*}\varepsilon_{2t}^{*3})$	55.76	48.12	34.64	57.80	48.02	28.78
$E(\varepsilon_{1t}^{*2}\varepsilon_{2t}^{*2})$	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

Notes: Monte Carlo empirical rejection rates of independence tests; 5,000 replications. Details on the data generating processes: DGP 4, joint (standardised) Student t: $(\varepsilon_{1t}^{*}, \varepsilon_{2t}^{*}) \sim t(\mathbf{0}, \mathbf{I}_2, 6)$; DGP 5, $(\varepsilon_{1t}^{*}, \varepsilon_{2t}^{*}) \sim At(\mathbf{0}, \mathbf{I}_2, -5\ell_2, 16)$ (see Mencía and Sentana (2012) for details); and DGP 6, $(\varepsilon_{1t}^{*}, \varepsilon_{2t}^{*}) \sim DMN_{LL}(.1, .2, 1/5)$ (see section 5.1.2 for details). We present the asymptotic distribution of the test statistics in section 5.2.2 and describe the sampling procedure we use to implement Giacomini, Politis and White (2013)'s warp-speed bootstrap in section 5.1.3.

Table 8: Monte Carlo distribution of parameter estimators.

Parameter θ (θ_0)	$\tau_1(1)$	$\tau_2(-1)$	$c_{11}(1)$	$c_{21}(0)$	$c_{12}(.5)$	$c_{22}(2)$	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} DMN & ε^*_{2t} DMN							DGP 4 – Joint Student t					
Mean	1.001	-1.000	0.983	0.029	0.472	1.951	1.000	-1.001	0.975	0.417	0.226	1.741
Std.dev.	0.071	0.128	0.101	0.384	0.206	0.149	0.071	0.127	0.163	0.831	0.465	0.290
DGP 2 – Shocks: ε^*_{1t} normal & ε^*_{2t} DMN							DGP 5 – Joint asymmetric t					
Mean	1.000	-1.000	0.974	0.087	0.432	1.906	1.001	-1.000	0.981	0.490	0.200	1.727
Std.dev.	0.071	0.126	0.120	0.542	0.310	0.201	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
Panel B: Sample size $T=1,000$												
Under the null of independence							*Under the alternative*					
DGP 1 – Shocks: ε^*_{1t} DMN & ε^*_{2t} DMN							DGP 4 – Joint Student t					
Mean	1.000	-1.000	0.997	-0.000	0.498	1.993	1.000	-1.001	0.982	0.434	0.219	1.748
Std.dev.	0.036	0.063	0.045	0.142	0.072	0.056	0.036	0.063	0.147	0.824	0.462	0.269
DGP 2 – Shocks: ε^*_{1t} normal & ε^*_{2t} DMN							DGP 5 – Joint asymmetric t					
Mean	1.000	-1.000	0.993	0.002	0.496	1.985	0.999	-1.000	0.994	0.788	-0.020	1.591
Std.dev.	0.035	0.063	0.060	0.223	0.120	0.064	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

Notes: 20,000 (5,000) replications under the null (alternative). 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 $\varepsilon^*_{2t} \sim DMN(-.859,.386,1/5)$; 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: $(\varepsilon^*_{1t},\varepsilon^*_{2t}) \sim t(\mathbf{0},\mathbf{I}_2,6)$; DGP 5, $(\varepsilon^*_{1t},\varepsilon^*_{2t}) \sim At(\mathbf{0},\mathbf{I}_2,-5\ell_2,16)$ (see Mencía and Sentana (2012) for details); and DGP 6, $(\varepsilon^*_{1t},\varepsilon^*_{2t}) \sim DMN_{LL}(.1,.2,1/5)$ (see section 5.1.2 for details).

40

Figure 1: Univariate densities of the independent shocks

Figure 1a: Discrete location-scale mixture of normals

Figure 1b: Symmetric Student t

Figure 1c: Asymmetric t

Notes: Dashed lines represent the standard normal distribution. Figure 1a 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 $\lambda = 1/5$); Figure 1b plots a standardised symmetric Student t with the same kurtosis (i.e. 10 degrees of freedom); while Figure 1c plots a standardised asymmetric t with skewness and kurtosis as the one in Figure 1a (i.e. with $\beta = -1.354$ and $\nu = 18.718$, see Mencía and Sentana (2012) for details).

Figure 2: Densities and contours of the bivariate distributions under the alternative hypotheses

Figure 2a: Standardised Student t density
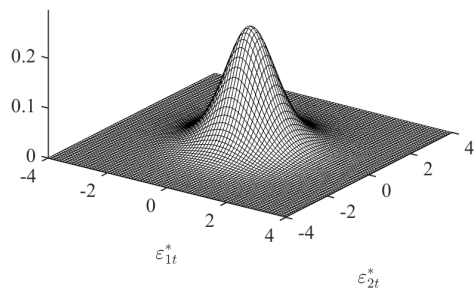
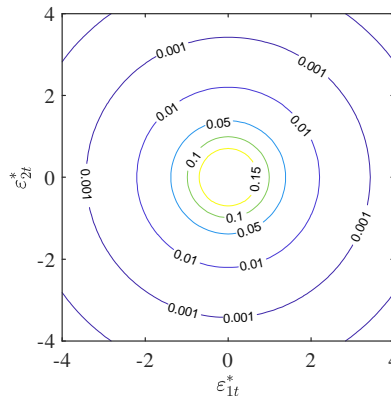Figure 2b: Contours of a standardised Student t density

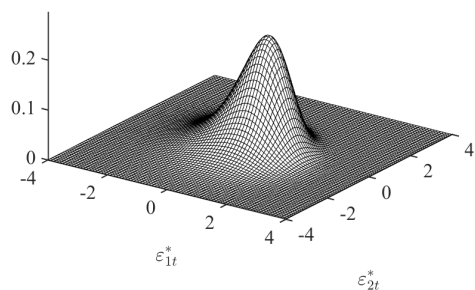Figure 2c: Standardised asymmetric t density

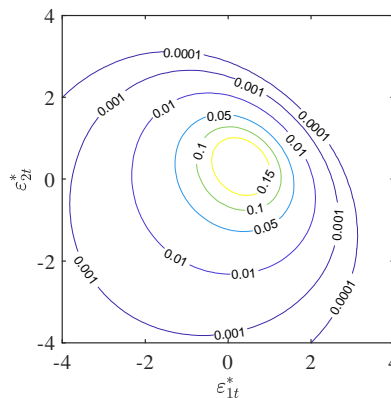

Figure 2d: Contours of a standardised asymmetric t density

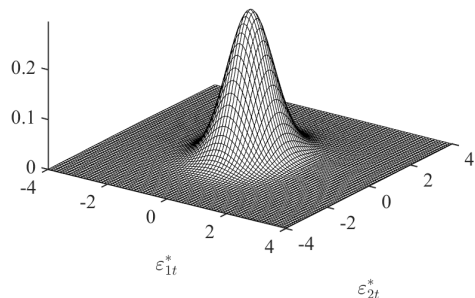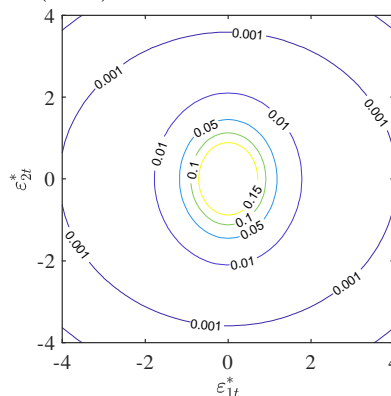Figure 2e: Standardised Lanne and Lütkepohl (2010)'s mixture of normals density

Figure 2f: Contours of a standardised Lanne and Lütkepohl (2010)'s mixture of normals density

Notes: Figures 2a–b plot a bivariate Student t with 6 degrees of freedom; Figures 2c–d a standardised bivariate asymmetric t with $\boldsymbol{\beta} = -5\ell_N$ and $\nu = 16$ (see Mencía and Sentana (2012) for details); while Figures 1e–f plot 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 section 5.1.2 and Lanne and Lütkepohl (2010) for details).