Specification tests for non-Gaussian structural vector autoregressions*

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

Gabriele Fiorentini
Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy
<gabriele.fiorentini@unifi.it>

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

December 2022
Revised: March 2024

Abstract

We propose specification tests for independent component analysis and structural vector autoregressions that assess the cross-sectional independence of non-Gaussian shocks by comparing their joint cumulative distribution with the product of their marginals at both discrete and continuous grids of argument values, the latter yielding a consistent test. We explicitly consider the sampling variability from computing the shocks using consistent estimators. We study the finite sample size of resampled versions of our tests in simulation exercises and show their non-negligible power against a variety of empirically plausible alternatives. Finally, we apply them to a dynamic model for three popular volatility indices.

Keywords: Consistest tests, Copulas, Finite normal mixtures, Independence tests, Pseudo maximum likelihood estimators, Volatility indices.

JEL: C32, C52, C58

---

*We would like to thank participants at the 2022 Time Series Econometrics Workshop (Zaragoza), Seoul National University, the 2022 NBER - NSF Time Series Conference (BU), CEMFI's Fall Econometrics Conference (November 2022), Florence Center for Data Science, New York Fed, USC, the conference in honour of Stock & Watson (May 2023), the 2023 ICEEE (Cagliari), the BSE Summer Forum (June 2023), ESEM 2023 (Barcelona), 34^{th} $(EC)^2$ Conference (Manchester, December 2023), XII Encuentro de la SEU (Montevideo, December 2023), Alicante and CUNEF for useful comments and suggestions. We are grateful to Kenneth Lei and Jianxuan Re too for their assistance in developing and documenting user-friendly Matlab code to compute our proposed tests, which is available on request. We also thank the editor, an associate editor, and two anonymous referees for valuable feedback. Of course, the usual caveat applies. The first and third authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00, while the second one is thankful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic".

1 Introduction

There are several popular identification schemes for structural vector autoregressions (SVAR), including 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) or Stock and Watson (2018)). Recently, identification through independent non-Gaussian shocks has become increasingly popular after Lanne, Meitz and Saikkonen (2017) and Gouriéroux, Monfort and Renne (2017).[1] The signal processing literature on Independent Component Analysis (ICA) popularised by Comon (1994) shares the same identification scheme.[2] Specifically, consider a static model the $N \times 1$ observed, square-integrable 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

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

In this context, the matrix $\mathbf{C}$ of loadings of the observed variables on the latent ones in (1) can be identified (up to column permutations and sign changes)[3] from an *i.i.d.* sample of observations on $\mathbf{y}$ provided the following assumption holds:[4]

Assumption 1 ***: ICA 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. In particular, suppose that $\boldsymbol{\varepsilon}^*$ follows a non-Gaussian spherically symmetric distribution, such as the standardised multivariate Student t, so that the marginal distribution of each shock is also a standardised Student t but there is tail dependence among them. The problem is that any rotation of the structural shocks generates another set of N shocks $\boldsymbol{\varepsilon}^{**} = \mathbf{Q}\boldsymbol{\varepsilon}^*$, where $\mathbf{Q}$ is a special orthogonal matrix, which share not only their mean vector ($\mathbf{0}$), covariance matrix ($\mathbf{I}_N$) and margins, but also the same non-linear dependence structure, rendering $\mathbf{C}$ underidentified.

In Amengual, Fiorentini and Sentana (2022), we proposed simple to implement and interpret specification tests that check potential cross-sectional dependence among several shocks

[1] See Fiorentini and Sentana (2023) for a selected list of recent SVAR papers that exploit the non-Gaussian features of the structural shocks.

[2] See Hyvärinen et al (2010) and Moneta et al (2013) for earlier links between ICA and SVARs.

[3] The elements of $\boldsymbol{\varepsilon}^*$ will be stochastically independent if and only if the elements of $\mathbf{I}_N^{1/2}\boldsymbol{\varepsilon}^*$ are too, where $\mathbf{I}_N^{1/2}$ denotes any orthogonal square root of the identity matrix of order N, so the choice of $\mathbf{I}_N^{1/2}$ is irrelevant.

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

by comparing some integer product moments of those shocks in the sample with their population counterparts. Specifically, we assessed 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, thereby nesting the integer moment tests for independence in Hyvärinen (2013) and Lanne and Luoto (2021). Although our Monte Carlo simulation results indicated that the tests we proposed have non-negligible power against a variety of empirical plausible forms of dependence among the shocks, tests based on a fixed number of cross-moments are not consistent because it is possible to create examples of tail-dependent shocks for which all those cross-moments are 0. Moreover, tests based on higher-order moments are quite sensitive to outliers, which renders asymptotic theory unreliable for their finite sample distributions.

The purpose of this paper is to provide alternative moment tests of independence consistent against any alternative to the null hypothesis under the maintained assumptions that at least $N-1$ shocks are non-Gaussian and $\mathbf{C}$ is invertible. Effectively, our proposed procedures check that the joint cumulative distribution function (cdf) of the shocks is the product of their marginal cdfs. For pedagogical reasons, we first develop our tests for a finite grid of values of the arguments of the cdfs, but then we explain how to extend them to the entire range of values by exploiting a generalisation of the continuum of moments inference procedures put forward by Carrasco and Florens (2000), which results in a consistent test. Interestingly, we can directly relate our discrete grid test to the classical Pearson's independence test statistic for categorical variables in contingency tables. We also explicitly compare our continuous grid procedures to the consistent independence tests of Hoeffding (1948), who considered a Cramér-von Misses type-test based on the integral of the square differences between the joint cdf and the product of the marginal cdfs, and Blum, Kiefer and Rosenblat (1961), who proposed Kolmogorov-Smirnov-type tests based on the maximum absolute discrepancy.

Importantly, we focus on the latent shocks rather than the observed variables because 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 $rank(\mathbf{C}_0) = N$, we could trivially recover the latent shocks from the observed signals without error. In practice, though, both $\boldsymbol{\mu}$ and $\mathbf{C}$ are unknown, and the same is true of the autoregressive coefficients in the SVAR case, so we need to estimate them before conducting our tests and take into account their sampling variability in computing the asymptotic covariance matrix of the influence functions in the discrete grid case, or its operator counterpart in the continuous one.

Although many estimation procedures for those parameters have been proposed in the literature (see, e.g., Moneta and Pallante (2022) and the references therein), in this paper we consider the discrete mixtures of normals-based pseudo maximum likelihood estimators (PMLEs) in Fiorentini and Sentana (2023) for three main reasons. First, they are consistent for all the model parameters under standard regularity conditions provided that Assumption 1 holds regardless of the true marginal distributions of the shocks. Second, they seem to be rather efficient, the rationale being that finite normal mixtures can provide good approximations to many univariate distributions. And third, the influence functions on which they are based are the scores of the pseudo log-likelihood, which we can easily compute in closed-form. As is well known, these influence functions play a crucial role in capturing 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 and operators of the sample moments underlying our tests under the null adjusted for parameter uncertainty. Importantly, we do so not only for the static ICA model (1) but also for a SVAR, which is far more relevant for economic and financial time series data. In both cases, though, we show that only the sampling variability in the off-diagonal elements of **C** matters.

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 and discuss resampling procedures that seem to improve their reliability. Finally, we explicitly compare our tests to several existing procedures, including not only our earlier cross-moments tests but also the aforementioned Blum, Kiefer and Rosenblat (1961) procedure and the consistent Matteson and Tsay (2017) distance covariance statistic, showing that ours have non-negligible power against a variety of empirically plausible alternatives in which the cross-sectional independence of the shocks does not hold.

Finally, we apply our proposed tests to a SVAR identified through independent non-Gaussian shocks for market-based volatility indices representative of three of the most actively traded asset classes: stocks, exchange rates and commodities. Specifically, we analyse the VIX, which captures the one-month ahead volatility of the S&P500 stock market index; the EVZ, which computes the 30-day volatility of the $US/Euro exchange rate, and the GVZ, which measures the market's expectation of 30-day volatility of gold prices. Interestingly, the tests that we propose fail to reject the null hypothesis of independence, and the same is true of other popular independence statistics. In contrast, moments tests based on co-skewness and co-kurtosis are very sensitive to the treatment of the unusual values for those indices observed at the onset of the COVID-19 pandemic.

The rest of the paper is organised as follows. Section 2 discusses the model and the estimation procedure. Then, we present our moment tests for independence for a finite number of grid points in section 3, and a continuum of points in section 4. Next, section 5 contains the results of our Monte Carlo experiments, while we describe our empirical application to the aforementioned volatility indices in section 6. Finally, we present our conclusions and suggestions for further research in section 7, relegating the main proofs to the appendix, and computational details and several auxiliary results to the supplementary material.

2 Structural vector autoregressions

2.1 Model specification

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

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

where I_{t-1} is the information set, $\mathbf{C}$ the matrix of impact multipliers and $\boldsymbol{\varepsilon}_t^*$ the structural shocks, which we normalise to have zero means, unit variances and zero covariances under our maintained assumption that they are square-integrable.

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 is well known, 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 QR matrix decomposition of $\mathbf{C}'$ to relate this matrix to the Cholesky decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_L \boldsymbol{\Sigma}_L'$ as $\mathbf{C} = \boldsymbol{\Sigma}_L \mathbf{Q}$, 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 $|\mathbf{Q}| = 1$ (see e.g. Golub and van Loan (2013)). While $\boldsymbol{\Sigma}_L$ is identified from the Gaussian log-likelihood, $\boldsymbol{\omega}$ is not. In fact, the underidentification of $\boldsymbol{\omega}$ would persist even if we assumed for estimation purposes that $\boldsymbol{\varepsilon}_t^*$ followed an elliptical distribution or a location-scale mixture of normals.

Nevertheless, Lanne et al. (2017) show that statistical identification of both the structural shocks and $\mathbf{C}$ (up to column permutations and sign changes) is possible under the ICA identification Assumption 1, which we maintain henceforth. Popular choices of univariate non-normal distributions are the Student t (see Brunnermeier et al. (2021)), the generalised error (or Gaussian) distribution, which includes normal, Laplace and uniform as special cases, and finite Gaussian mixtures. Henceforth, we assume that the researcher chooses the order of the columns of $\mathbf{C}$ as well as their signs following for example the scheme proposed by Ilmonen and Paindaveine (2011) and explained in detail in section 3.3 of Lanne et al (2017), thereby transforming the local identification into a global one.

2.2 Going beyond integer moments

The Lanne et al. (2017) identification result, though, critically hinges on the validity of Assumption 1. As a consequence, it would be desirable that empirical researchers who rely on it reported specification tests that would check this assumption. In this paper, we focus on testing that the structural shocks are indeed independent of each other.

As is well known, stochastic independence between the elements of a random vector is equivalent to the joint cdf 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 influence functions of the form

$$c_{\mathbf{h}}(\boldsymbol{\varepsilon}_t^*) = \prod_{i=1}^{N} \varepsilon_{it}^{*h_i} - \prod_{i=1}^{N} E(\varepsilon_{it}^{*h_i}), \tag{3}$$

where $\mathbf{h} = \{h_1, ..., h_N\}$, with $h_i \in \mathbb{Z}_{0+}$, denotes the index vector characterising a specific product moment. This is precisely the approach that we followed in Amengual, Fiorentini and Sentana (2022), where we paid particular attention to third and fourth cross-moments. As we mentioned in the introduction, though, this type of moment test suffers from two problems. First, standard asymptotic theory provides poor finite sample approximations for tests based on higher-order moments, whose estimates are quite sensitive to outliers. Second, for any choice of $\mathbf{h}$, one can find joint distributions of the shocks for which (3) is zero on average even though the shocks are cross-sectionally dependent. For example, Figure 1a displays the contours of the copula corresponding to a spherically symmetric fourth-order Hermite expansion of the bivariate normal such that all second, third and fourth cross-moments satisfy this condition when the margins come from the same distribution even though the shocks are not stochastically independent.

To avoid these criticisms, in what follows we propose to assess the potential cross-sectional dependence among two or more shocks by directly comparing their joint empirical cdf to the product of the marginal empirical cdfs. We do so not only for a discrete grid of values of the arguments of the joint cdf, which provides the intuition for our approach, but also for a continuous grid of values using an extension of the continuum of moments inference procedures in Carrasco and Florens (2000), which provides a consistent test.

2.3 Consistent parameter estimation

Let $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\mathbf{A}_1), \dots, vec'(\mathbf{A}_p), vec'(\mathbf{C})]' = (\boldsymbol{\tau}', \mathbf{a}_1', \dots, \mathbf{a}_p', \mathbf{c}') = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$ denote the structural parameters characterising the first two conditional moments of $\mathbf{y}_t$. If we knew the true values of $\boldsymbol{\theta}_0$, we could easily recover the true shocks from the observed variables using the expression

$$\varepsilon_t^*(\boldsymbol{\theta}) = \mathbf{C}^{-1}(\mathbf{y}_t - \boldsymbol{\tau} - \textstyle\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j}). \tag{4}$$

In practice, though, all those mean and variance parameters are unknown, so we need to both estimate them before computing our tests and take into account that we will be working with estimated shocks in deriving the asymptotic covariance matrices of the average influence functions underlying them.

Maximum likelihood estimation (MLE) and inference in SVAR models with independent non-Gaussian shocks is conceptually simple: the joint log-likelihood function is the sum of N univariate log-likelihoods plus the Jacobian term $|\mathbf{C}|$. As is well known, MLE leads to efficient estimators of all the structural parameters if the assumed univariate distributions are correctly specified. Unfortunately, while Gaussian pseudo maximum likelihood estimators (PMLE) remain consistent when the true shocks are not Gaussian, the same is not generally true for other distributions (see e.g. Newey and Steigerwald (1997)). In this context, though, we cannot use a Gaussian PMLE because we lose identification.

Fiorentini and Sentana (2023) showed that if the univariate log-likelihoods are based on an unrestricted finite Gaussian mixture, then all conditional mean and variance parameters will be consistently estimated under standard regularity conditions when Assumption 1 holds and the shape parameters of the mixtures are simultaneously obtained.[5] Let $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1', \ldots, \boldsymbol{\varrho}_N')'$ denote those shape parameters, so that $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho}')'$. Similarly, let $\mathbf{s}_{\phi t}(\boldsymbol{\phi})$ denote the score vector used for simultaneously estimating $\boldsymbol{\theta}$ and the finite mixture shape parameters $\boldsymbol{\varrho}$. Finally, let $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_0', \boldsymbol{\upsilon}_0')'$ denote the true values of the parameters characterising the true DGP, with $\boldsymbol{\upsilon}$ containing the potentially infinite-dimensional shape parameters of the true distributions of the shocks, and $\boldsymbol{\phi}_\infty = (\boldsymbol{\theta}_0', \boldsymbol{\varrho}_\infty')'$ the pseudo-true values of the estimated parameters. Supplemental Appendix C provides detailed expressions not only for the relevant pseudo log-likelihood function, but also for its score and Hessian, as well as the conditional variance of the former and the conditional expected value of the latter, on the basis of which we can obtain closed-form expressions for the asymptotic variances of the PMLEs of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$:

$$\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \quad \text{where}$$

$$\mathcal{A}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\partial \mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)/\partial \boldsymbol{\phi}' | \boldsymbol{\varphi}_0] \quad \text{and} \tag{5}$$

$$\mathcal{B}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V[\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty) | \boldsymbol{\varphi}_0] \tag{6}$$

denote the (-) expected value of the non-Gaussian log-likelihood Hessian and the variance of its

[5] The rationale is that the discrete normal mixture-based PMLEs of the unconditional mean vector and covariance matrix of a random vector coincide with the corresponding sample moments, just like in the Gaussian case, as shown by Fiorentini and Sentana (2022).

score, respectively. Supplemental Appendix A explains how to consistently estimate these matrices in practice, while Supplemental Appendix C.2 lists the regularity conditions that guarantee their consistency.

3 Discrete grid tests

Our starting point is a test based on the joint probability of events that involve two or more elements of $\boldsymbol{\varepsilon}_t^*$, which should coincide with the product of the marginal probabilities under the null of independence. Specifically, we begin by defining H points, $k_1 < \cdots < k_h < \cdots < k_H$, so that we can then form a partition of the support of ε_{it}^* into $H+1$ segments, namely $k_{h-1} \leq \varepsilon_{it}^* \leq k_h$ for $h = 1, ..., H+1$ after suitably defining $k_0 = -\infty$ and $k_{H+1} = \infty$.[6] We then collect the indices of the shocks involved in the set $M = \{i^1, ..., i^{\text{M}}\}$, where M denotes the cardinality of the set M, so that we can test for pairwise independence (M=2), joint independence of the entire vector of structural innovations (M=N), and any other intermediate situation.

Next, we define the dummy variables $P_{ht}^{\diamond i} = 1_{(k_{h-1},k_h)}(\varepsilon_{it}^*)$, where $1_A(x)$ denotes the usual indicator function for $x \in A \subset \mathbb{R}$. Moreover, let $\upsilon^{\diamond}(\mathbf{k}, \mathbf{u}^{\diamond})$ denote the difference between the joint probability of the event defined by the vector $\mathbf{k} = (k_{h_{i^1}}, \ldots, k_{h_{i^{\text{M}}}})$, $\Pr(\bigcap_{i \in M}\{P_{h_i t}^{\diamond i} = 1\})$, and the product of the marginal probabilities, $u_h^{\diamond i} = \Pr(P_{ht}^{\diamond i} = 1)$, with $\mathbf{u}^{\diamond} = (u_{i^1}, \ldots, u_{i^{\text{M}}})$, so that $\upsilon^{\diamond}(\mathbf{k}, \mathbf{u}^{\diamond}) = 0$ under independence. Using this notation, we could in principle test the null on the basis of influence functions of the form:

$$p^{\diamond}(\boldsymbol{\varepsilon}_t^*; \mathbf{k}, \mathbf{u}^{\diamond}) = \prod_{i \in M} P_{h_i t}^{\diamond i} - \prod_{i \in M} u_{h_i}^{\diamond i} - \upsilon^{\diamond}(\mathbf{k}, \mathbf{u}^{\diamond}). \tag{7}$$

However, (7) is not computable unless one knows the marginal probabilities in $\mathbf{u}^{\diamond}$, like in Fisher's (1922) famous tea cup classification example. Therefore, in practice those probabilities will in turn be estimated from the exactly identified moment conditions

$$\begin{aligned} E[p_{h_{i^1}}^{\diamond}(\varepsilon_{i^1 t}^*)] &= 0, ..., E[p_{h_{i^{\text{M}}}}^{\diamond}(\varepsilon_{i^{\text{M}} t}^*)] = 0, \\ p_h^{\diamond}(\varepsilon_{it}^*) &= P_{ht}^{\diamond i} - u_h^{\diamond i}, \text{ for } i \in M,\ h = 1, ..., H, \end{aligned} \tag{8}$$

which results in the analogue estimator $\hat{u}_h^{\diamond i} = \frac{1}{T}\sum_{t=1}^{T} P_{ht}^{\diamond i}$, a fact that we need to take into account in computing the asymptotic covariance matrix of the feasible version of (7) that adequately reflects the sampling uncertainty in $\hat{u}_h^{\diamond i}$ for all intervals and shocks.

If the true shocks were observed and we considered all possible cells from an N-dimensional

[6] For notational simplicity, we maintain the assumption that the number of intervals and their limits are common across shocks. Although this assumption is plausible when a researcher has no prior views on the marginal distributions of the different standardised shocks, it would be straightforward to relax it.

contingency table whose elements are the Cartesian product of the different marginal partitions,[7] then we would end up with a GMM version of Pearson's joint (or multi-way) independence test, which would in fact be numerically identical to Pearson's original test statistic (see Sentana (2022)).[8]

Let us now replace the partition of the support of ε^*_{it} into the $H+1$ segments discussed above by the sequence of overlapping increments $\varepsilon^*_{it} \leq k_h$ for $h = 1, ..., H+1$. For practical purposes, let us define $P^i_{ht} = 1_{(-\infty,k_h)}(\varepsilon^*_{it})$ and

$$p_h(\varepsilon^*_{it}) = P^i_{ht} - u^i_h, \tag{9}$$

with $u^i_h = E(P^i_{ht}) = \Pr(\varepsilon^*_{it} \leq k_h) \equiv F_i(k_h)$, as the new dummy variables and marginal influence functions, respectively, which trivially give rise to the empirical cdf estimator

$$\hat{u}^i_h = \frac{1}{T}\sum_{t=1}^{T} P^i_{ht}. \tag{10}$$

Let us also define the joint influence function

$$p(\boldsymbol{\varepsilon}^*_t; \mathbf{k}, \mathbf{u}) = \prod_{i\in M} P^i_{h_i t} - \prod_{i\in M} u^i_{h_i} - \upsilon(\mathbf{k}, \mathbf{u}), \tag{11}$$

which is such that $\upsilon(\mathbf{k}, \mathbf{u}) = 0$ under the independence null.

Importantly, the fact that the estimating moment conditions (9) exactly identify the relevant u^i_h's implies that there is no efficiency loss in sequentially estimating the $\upsilon(\mathbf{k}, \mathbf{u})$'s from (11) by replacing the marginal cdfs by their sample counterparts relative to estimating them jointly from (9) and (11), which in turn implies that the non-centrality parameters of corresponding moment tests that impose $\upsilon(\mathbf{k}, \mathbf{u}) = 0$ will coincide.

Then, we can show the following:

Proposition 1 *Let $\mathbf{p}^\diamond(\boldsymbol{\varepsilon}^*_t)$ and $\mathbf{p}(\boldsymbol{\varepsilon}^*_t)$ denote the $H^{\mathrm{M}} \times 1$ vectors containing two mutually compatible full sets of non-redundant influence functions (7) and (11), respectively. Independence tests based on them are numerically identical after taking into consideration the estimation of the marginal probabilities from (8) and (9), respectively.*

This convenient re-interpretation of the usual Pearson test for independence will allow us to extend our tests to a continuous grid in section 4.

[7] The adding up restrictions of the elements of the contingency table by rows and columns imply that the information in some of the cells is redundant, so we can avoid using generalised inverses in computing the test statistic by getting rid of them. We would suggest excluding all the cells involving a specific category for each of the M shocks, but the choice of excluded category for each shock is arbitrary.

[8] In principle, one could deviate from Pearson's test by not including all non-redundant cells in the contingency table, but unless one has a priori knowledge of which specific subset of intervals is likely to capture larger departures from the null, it is not clear that the consequent reduction in degrees of freedom will translate into power gains.

The choice of the k's, though, will crucially affect power even though it does not affect the (first-order) asymptotic distribution of the test under the null. For that reason, it would be useful to adapt the grid to the marginal distribution of the shocks. With this in mind, we recommend the following simple way of choosing the partition which achieves precisely that goal: instead of fixing arbitrarily the grid points at which we evaluate the cdfs of each of the ε_i^*'s, we chose them so that they correspond to specific quantiles of the marginal distributions.

Specifically, let $k_h^i = \varkappa_i(u_h)$ for each $i \in M$ be the u_h-quantile of ε_{it}^* for $h = 1, ..., H$, with $0 = u_0 < u_1 < \cdots < u_H < u_{H+1} = 1$, $\varkappa_i(0) = -\infty$ and $\varkappa_i(1) = \infty$. We can compute an alternative independence test for the shocks using the same influence function $p(\boldsymbol{\varepsilon}_t^*)$ in (11), but now estimating the marginal quantiles k_h^i for given u_h^i from the exactly identified moment conditions (9) rather than each marginal cdf u_h^i for fixed k_h^i. Intuitively, a moment test based on a collection of such influence functions will effectively assess that the copula linking the different marginal distributions is flat, which corresponds to the independent one.

An obvious question at this stage is whether practitioners should rely on the event-based approach that treats the k's as fixed or the copula-flavoured test, which instead treats the u's as fixed. A priori, it might seem that the former should dominate the latter because the asymptotic variance of the estimators of the probabilities of an interval only depend on the probability of said interval, while the asymptotic variance of the estimators of the quantiles depend not only on the quantile probability (directly), but also on the value of the density at said quantile (inversely). Somewhat surprisingly, though, it turns out that both tests are asymptotically equivalent if we chose the limits of the intervals k_h^i's so that they exactly match the theoretical quantiles $\varkappa_i(u_h)$'s, as we show in the next proposition:

Proposition 2 *Regardless of whether* $\mathbf{k}$ *is fixed and* $\mathbf{u}$ *estimated, or* $\mathbf{u}$ *fixed and* $\mathbf{k}$ *estimated:*
a) The testing influence function linearised to consider the estimation of the relevant marginal quantities is given by

$$m_t(\mathbf{u}) = \left[\prod_{i\in M} 1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*) - \prod_{i\in M} u_i\right] - \sum_{i\in M}\left[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in M, i'\neq i} u_{i'}. \quad (12)$$

b) If the shocks in M *are stochastically independent, then the asymptotic covariance of the influence functions* $m_t(\mathbf{u})$ *and* $m_t(\mathbf{u}')$ *will be given by*

$$\prod_{i\in M}\min(u_{k_i}^i, u'^i_{k'_i}) + (\mathrm{M}-1)\prod_{i\in M} u_{k_i}^i u'^i_{k'_i} - \sum_{i\in M}\min(u_{k_i}^i, u'^i_{k'_i})\left(\prod_{i'\in M, i'\neq i} u_{k_{i'}}^{i'}\right)\left(\prod_{i'\in M, i'\neq i} u'^{i'}_{k'_{i'}}\right). \quad (13)$$

For empirical considerations, in what follows we focus on the copula version, which naturally adapts the grid to the unknown marginal distribution of each shock.

The linearised influence functions (12) are particularly useful in practice because $\boldsymbol{\theta}_0$ is unknown and the quantiles are computed on the basis of estimated $\boldsymbol{\varepsilon}_t^*$'s that replace those true values with their PMLE $\hat{\boldsymbol{\theta}}$. Specifically, we can apply the theory of moment tests in Newey (1985) and Tauchen (1985) to them to derive the following result:

Proposition 3 *Let $\mathbf{m}_t(\boldsymbol{\theta})$ denote an $H^{\mathrm{M}} \times 1$ vector containing a full set of non-redundant linearised influence functions (12) and $\hat{\boldsymbol{\theta}}$ the consistent PMLE estimators of $\boldsymbol{\theta}$. Under standard regularity conditions*

$$T\bar{\mathbf{m}}_T'(\hat{\boldsymbol{\theta}})\mathcal{W}_{\mathbf{p}}^{-1}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}) \to \chi^2_{H^{\mathrm{M}}}, \tag{14}$$

under the independence null, where $\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ is the sample mean of $\mathbf{m}_t(\boldsymbol{\theta})$,

$$\mathcal{W}_{\mathbf{m}} = \mathcal{V}_{\mathbf{m}} + \mathcal{J}_{\mathbf{m}}\mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1}\mathcal{J}_{\mathbf{m}}' + \mathcal{F}_{\mathbf{m}}\mathcal{A}^{-1}\mathcal{J}_{\mathbf{m}}' + \mathcal{J}_{\mathbf{m}}\mathcal{A}^{-1}\mathcal{F}_{\mathbf{m}}', \tag{15}$$

$\mathcal{V}_{\mathbf{m}} = \mathcal{V}_{\mathbf{m}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = V[\mathbf{m}_t(\boldsymbol{\theta}_0)|\,\boldsymbol{\varphi}_0]$, whose entries are given by (13);
$\mathcal{J}_{\mathbf{m}} = \mathcal{J}_{\mathbf{m}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = E[\partial\mathbf{m}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\phi}'|\,\boldsymbol{\varphi}_0]$, whose only non-zero elements are

$$\mathrm{J}_{m_{\mathbf{u}}c_{ii'}}(\boldsymbol{\varrho}_\infty,\boldsymbol{\varphi}_0) = -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f_i(k_i), \quad for \;\; i\neq i', \tag{16}$$

with $\eta_i = E_0[\varepsilon_{it}^* 1_{(-\infty,k_i)}(\varepsilon_{it}^*)]$ and $f_i(.)$ denoting the true marginal density of shock $i \in M$;
$\mathcal{F}_{\mathbf{m}} = \mathcal{F}_{\mathbf{m}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = cov[\mathbf{m}_t(\boldsymbol{\theta}_0), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\,\boldsymbol{\varphi}_0]$, whose typical element is $E[\mathcal{K}_{m_{\mathbf{u}}t}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)]$,

$$\mathcal{K}_{m_{\mathbf{u}}t}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0) = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathcal{K}_{p_{\mathbf{k}}}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) \\ \mathbf{0} \end{bmatrix},$$

$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = [(1, \mathbf{y}_{t-1}', ..., \mathbf{y}_{t-p}', \mathbf{0}_{1\times N^2})' \otimes \mathbf{I}_N]\mathbf{C}^{-1\prime}$, $\mathbf{Z}_{st}(\boldsymbol{\theta}) = (\mathbf{0}_{N^2\times N(N+1)}, \mathbf{I}_{N^2})(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})$, and $\mathcal{K}_{m_{\mathbf{u}}}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)$ an $N^2 \times 1$ vector whose entries $s = N(i-1)+i'$ for $i, i' = 1, ..., N$ are

$$\mathrm{K}_{m,s}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) = -\sum_{i\in M}\sum_{i'\in M}\left(\prod_{i''\in M, i''\neq i, i''\neq i'} u_{i''}\right)\eta_{i'} E\left\{ 1_{(\varepsilon_{it}^*\leq k_{h_i})} \cdot \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\},$$

for $i \neq i'$, and zero otherwise; and $\mathcal{A} = \mathcal{A}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{B} = \mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ given by (5) and (6).

Supplemental Appendix A explains in detail how to transform the infeasible statistic in (14) into a feasible one by replacing $\mathcal{W}_{\mathbf{m}}$ with a consistent estimator without altering its asymptotic distribution.

A key implication of this result is that the sampling variability in estimating the mean parameters or the diagonal elements of the matrix $\mathbf{C}$ is asymptotically irrelevant. In fact, the variability in the intercepts τ_i' and scales c_{ii}'s of the shocks does not matter even in finite samples because a contingency table based on quantiles is numerically invariant to affine linear transformations of each shock as the new quantiles are the same affine transformation of the original ones. And even though changes in the elements in $\mathbf{A}_j$ will affect shocks differently in different periods, it turns out that the corresponding expected Jacobian is zero. Therefore, the only parameters whose sampling variability matter are the off-diagonal elements of $\mathbf{C}$.

Importantly, our tests will also be numerically invariant to alternative ways of selecting a particular column permutation of $\mathbf{C}$ for identification purposes as long as we use the same quantile grid for all shocks because those permutations only affect the labelling of the shocks. Similarly, a change in the sign of one shock and the corresponding column is also numerically irrelevant as long as we adjust its quantiles accordingly. In fact, given that we recommend using the same equally spaced quantiles in the discrete grid case (say terciles, quartiles, quintiles, etc.), we do not even need to consider such an adjustment.

In contrast, the choice of H is crucial for both small sample performance and power considerations even though the asymptotic distribution under the null is always a χ^2 with H^{M} degrees of freedom. Intuitively, too fine a partition relative to the sample size may introduce size distortions because the joint probability of some individual cells will be poorly estimated. Even in large samples, a fine partition will generate substantial correlation between the influence functions, potentially causing numerical instability. Finally, there is also a power trade-off between the size of the non-centrality parameter and the number of degrees of freedom of the limiting distribution. Partly for these reasons, next we discuss tests which do not depend on H.

4 A continuous grid

A more fundamental problem with the tests discussed in the previous section is that they are not consistent for any specific finite partition of the domain of the shocks because one could always find joint distributions such that the probability of each joint interval is exactly the product of the marginal probabilities even though the shocks are stochastically dependent. In fact, any spherically symmetric bivariate distribution for the shocks, like the one in Figure 1a, will provide an example of such a situation with only two equally likely intervals for each shock. More interestingly, Figure 1b relies on another spherically symmetric Hermite expansion of the bivariate normal to illustrate the same issue if we considered three equally likely intervals per shock. To address this shortcoming, we now extend our procedures to a continuous grid.

Consistent tests of independence based on comparing the joint cdf to the product of the marginal cdfs for all possible values of the arguments go back at least to Hoeffding (1948), who considered a Cramér-von Misses type-test based on the integral of the square differences between the joint cdf and the product of the marginal cdfs, and Blum, Kiefer and Rosenblat (1961), who considered Kolmogorov-Smirnov-type tests based on the maximum absolute discrepancy.[9] However, those tests rely on specific functionals of the difference, while the discrete grid tests

[9]See Kheifets (2015) for an application of these procedures to the probability integral transforms of the conditionally standardised residuals of a fully parametric univariate time series model for the purposes of testing its correct specification taking into account the estimated character of those residuals, and Scaillet (2005) for a

that we studied in the previous section also take into account not only the asymptotic variance of the influence functions for each value of the arguments, like an Anderson-Darling (1952) test would do, but more importantly, the covariance between those influence functions for different values of the arguments.

In principle, we could try to find the limiting distribution of our discrete grid tests in a double asymptotic framework in which the partitions get finer and finer as the sample size increases. However, this is really unnecessary because the influence functions indexed with respect to the arguments of the joint cdf over $\mathbb{R}^{\mathrm{M}}$ give rise to a continuum of moments in an L^2 space. As a result, we can readily extend Carrasco and Florens (2000) and directly construct a Hansen (1982) overidentifying restrictions-type test based on the same influence functions as in the discrete grid case, but with a covariance operator playing the role of the usual covariance matrix.[10]

Specifically, by transforming ε_{it}^* into its empirical uniform rank

$$\epsilon_{it}^* = \frac{1}{T}\sum\nolimits_{s=1}^{T} 1_{(-\infty,\varepsilon_{it}^*)}(\varepsilon_{is}^*), \tag{17}$$

we can define the marginal and joint influence functions

$$q_{it}(u_i) = 1_{(0,u_i)}(\epsilon_{it}^*) - u_i, \quad \text{and} \tag{18}$$

$$q_t(\mathbf{u}) = \prod_{i\in M} 1_{(0,u_i)}(\epsilon_{it}^*) - \prod_{i\in M} u_i, \tag{19}$$

which are numerically identical to (9) and (11) in the previous section, respectively.

Next, let ϖ be a probability density function with support the unit hypercube. Then, the function $q_t(\mathbf{u})$ may be regarded as a random element of $L^2(\varpi)$, the space of real-valued functions which are square integrable with respect to the density ϖ. For any functions f and g in $L^2(\varpi)$, the inner product on this Hilbert space is defined as $\langle f, g\rangle = \int_{[0,1]^{\mathrm{M}}} f(\mathbf{u})\, g(\mathbf{u})\varpi(\mathbf{u})\, d\mathbf{u}$. By the central limit theorem for *iid* random elements of a separable Hilbert space (see e.g. proof of Theorem 9 in Rackauskas and Suquet (2006)), we have that under independence, $\sqrt{T}\bar{q}_T(\mathbf{u}) \Rightarrow \mathcal{N}(0, K)$ in $L^2(\varpi)$ as T goes to infinity, where $\bar{q}_T(\mathbf{u})$ denotes the sample average of (19) and $\mathcal{N}(0, K)$ a Gaussian process of $L^2(\varpi)$ fully characterised by its covariance operator K, which is an integral operator from $L^2(\varpi)$ to $L^2(\varpi)$ such that

$$(Kf)(\mathbf{u}) = \int_{[0,1]^{\mathrm{M}}} k(\mathbf{u},\mathbf{v}) f(\mathbf{v})\varpi(\mathbf{v})\, d\mathbf{v}, \tag{20}$$

whose kernel $k(\mathbf{u},\mathbf{v}) = E[q_t(\mathbf{u})q_t(\mathbf{v})]$ is given by (13).

related test against positive quadrant dependence.

[10] A straightforward extension of Proposition 2 implies that the continuum of moments test that looks at (11) over $\mathbb{R}^{\mathrm{I}}$ will be numerically equivalent to the one that looks at the difference between the empirical copula and the unit hyperplane over the unit hypercube. For that reason, in what follows we simply focus on the copula-based version of the moment tests for overidentifying restrictions.

As we mentioned before, we are interested in applying an overidentifying restrictions test to our continuum of moments, but replacing the usual covariance matrix by the aforementioned covariance operator K, which has a countable infinite number of positive eigenvalues λ_{jk} and associated eigenfunctions μ_{jk}. Specifically, Blum, Kiefer and Rosenblat (1961) proved that in the bivariate case with $\boldsymbol{\theta}_0$ known, the eigenvalues λ_{jk} and the complete set of orthonormal eigenfunctions $\boldsymbol{\mu}_{jk}(\mathbf{u})$ of K, which are the solutions to the functional equation

$$K\boldsymbol{\mu}_{jk}(\mathbf{u}) = \int_0^1 \int_0^1 K(\mathbf{u}, \mathbf{v})\mu_{jk}(\mathbf{v})d\mathbf{v} = \lambda_{jk}\mu_{jk}(\mathbf{u}),$$

are given by $1/(\pi^4 j^2 k^2)$ and $2(\sin \pi j u_1)(\sin \pi k u_2)$ for $j, k = 1, 2, \ldots$.[11] This covariance operator is compact, meaning that its inverse is not bounded. Consequently, its smallest eigenvalues will converge to zero as j or k go to infinity, as can be clearly seen in the bivariate case we have just discussed, so taking the inverse of K is problematic. In terms of the spectral decomposition of K, the direct analogue to the J test statistic would be written as

$$\left\langle \sqrt{T}\bar{q}_T, K^{-1}\bar{q}_T \right\rangle = \sum_j \sum_k \frac{1}{\lambda_{kj}} \left| \left\langle \sqrt{T}\bar{q}_T, \mu_{jk} \right\rangle \right|^2. \tag{21}$$

Unfortunately, this expression will blow up because of the division by the small eigenvalues. This is related to the problem of solving an integral equation $Kf = g$ where g is known and f is the object of interest. This problem is said to be ill-posed because f is not continuous in g. Indeed, a small perturbation in g will result in a large change in f. To stabilise the solution, one needs to use some regularisation scheme (see Kress (1999) and Carrasco, Florens, and Renault (2007) for various possibilities). As in Carrasco and Florens (2000), we use Tikhonov regularisation, which consists in replacing $K^{-1}g$ by the regularised solution $\left(K^2 + \alpha I\right)^{-1} Kg$ where $\alpha \geq 0$ is a regularisation parameter. In what follows, we use the notation $(K^\alpha)^{-1}$ for $\left(K^2 + \alpha I\right)^{-1} K$, which is the operator with eigenvalues $\lambda_{jk}(\lambda_{jk}^2 + \alpha)^{-1}$ and corresponding eigenfunctions μ_{jk}, and $(K^\alpha)^{-1/2}$ for the operator with eigenvalues $\lambda_{jk}^{1/2}(\lambda_{jk}^2 + \alpha)^{-1/2}$ and the same eigenfunctions. Thus, the regularised version of the J-type test will be

$$\left\| (K^\alpha)^{-1/2} \sqrt{T}\bar{q}_T \right\|^2 = \sum_j \sum_k \frac{\lambda_{jk}}{\lambda_{jk}^2 + \alpha} \left| \left\langle \sqrt{T}\bar{q}_T, \mu_{jk} \right\rangle \right|^2. \tag{22}$$

Comparing the expressions (21) and (22), it is easy to see that we have effectively replaced λ_{jk}^{-1} with $\lambda_{jk}(\lambda_{jk}^2 + \alpha)^{-1}$, which is bounded.

For computational reasons, it is convenient to rewrite the test statistic (22), which uses as eigenvalues and eigenfunctions those of K, in terms of certain matrices and vectors (see Carrasco

[11] It is not worth extending their results for I> 2 because they apply to observed variables rather than estimated shocks.

et al. (2007) for analogous expressions for K under time series dependence). Specifically, we use the following computationally convenient expression for (22):

$$\mathbf{W}'\{\alpha\mathbf{I}_T + [(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\mathcal{D}^2(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)]^2\}^{-1}\mathbf{W}, \tag{23}$$

where $\mathbf{W}$ is a $T \times 1$ vector whose t^{th} element is $w_t = \int q_t(\mathbf{u})\,\bar{q}_T(\mathbf{u})\,\varpi(\mathbf{u})\,d\mathbf{u}$, $\mathcal{D}$ is a $T \times T$ matrix whose $(t,s)^{th}$ element is $d_{ts} = \langle q_t, q_s \rangle / T$, and $\boldsymbol{\ell}_T$ is a $T \times 1$ vector of ones. In practice, only $\mathcal{D}$ is needed to compute the test statistic because (23) is equivalent to

$$\boldsymbol{\ell}_T'\mathcal{D}(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\{\alpha\mathbf{I}_T + [(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\mathcal{D}^2(\mathbf{I}_T - \ell_T\ell_T'/T)]^2\}^{-1}(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\mathcal{D}\boldsymbol{\ell}_T, \tag{24}$$

with the analytical expression for the $(t,s)^{th}$ element of the matrix $\mathcal{D}$ provided in the following result, which generalises expression (13) to a continuous grid:

Proposition 4 *If the* M *shocks in M are stochastically independent, then*

$$d_{ts} = \frac{1}{T}\left\{\prod_{i\in M}[1-\max[\epsilon_{it}^*, \epsilon_{is}^*)] - \left(\frac{1}{2}\right)^{\text{M}}\prod_{i\in M}(1-\epsilon_{it}^{*2}) - \left(\frac{1}{2}\right)^{\text{M}}\prod_{i\in M}(1-\epsilon_{is}^{*2}) + \left(\frac{1}{3}\right)^{\text{M}}\right\}, \tag{25}$$

for $t, s = 1, ..., T$.

In addition to the effects of estimating the marginal cdfs of the shocks on the covariance operator, we must again take into account the sampling variability in estimating $\boldsymbol{\theta}$. Fortunately, the only difference with the discrete grid case is that the expected Jacobian will now be a function of the values of the arguments of the cdf, and the same will be true of the covariance between the influence functions and the score of the Gaussian PMLE. With this trivial re-interpretation, all the expressions in Proposition 3 continue to be valid. In effect, the only thing we need to do is to apply the Carrasco and Florens (2000) procedure to the residuals from projecting the influence function (19) on the linear span generated by the influence functions defining the marginal cdfs and the scores of the pseudo log-likelihood function for each value of $\mathbf{u}$ (see Khmaladze (1981) for an analogous transformation). As we explained in the previous section, the only parameters whose sampling variability matter are the off-diagonal elements of $\mathbf{C}$.

In this context, we can obtain the adjusted covariance operator by combining the expressions in Proposition 4 with Lemma 3 in Supplemental Appendix B to obtain:

Proposition 5 *Let $\hat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$. Under standard regularity conditions, the overidentifying test statistic will be given*

$$\boldsymbol{\ell}_T'\mathcal{E}(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\{\alpha\mathbf{I}_T + [(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\mathcal{E}^2(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)]^2\}^{-1}(\mathbf{I}_T - \boldsymbol{\ell}_T\boldsymbol{\ell}_T'/T)\mathcal{E}\boldsymbol{\ell}_T, \tag{26}$$

with $\mathcal{E} = \mathcal{D} + \mathcal{C}$, where the elements of $\mathcal{D}$ are given by (25),

$$\mathcal{C} = \boldsymbol{\ell}_T \boldsymbol{\ell}_T' \cdot \int_{[0,1]^M}' \mathbf{j}(\mathbf{u}_M)' T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{j}(\mathbf{u}_M) d\mathbf{u}_M, \tag{27}$$

and $\mathbf{j}(\mathbf{u}_M)$ is a vector whose only non-zero entries are the ones corresponding to $c_{ii'}$ (for $i \neq i'$), whose expression appears in (16).

For a fixed value of α, the results in Carrasco and Florens (2000) indicate that the test statistic in (26) will converge under the null to a weighted sum of χ^2's so that one could obtain the corresponding p-values using the approach in Imhof (1961). In this respect, Supplemental Appendix A explains in detail how to transform the infeasible test statistic in (26) into a feasible one by replacing $\mathcal{C}$ and $\mathcal{D}$ with consistent estimators without altering its asymptotic distribution, as originally shown in section 3 of Carrasco and Florens (2000) (see also Proposition 6 in Amengual, Carrasco and Sentana (2020)). In principle, one could also study the rates at which α should go to 0 with the sample size for the asymptotic distribution of (26) suitably centred and scaled to converge to a standard normal distribution, as in Proposition 10 in Amengual, Carrasco and Sentana (2020). Nevertheless, we recommend the resampling procedures described in section 5.1.3 to obtain more reliable p-values in small samples.

5 Monte Carlo analysis

In this section, we evaluate the finite sample behaviour of the independence tests discussed in the previous sections by means of several Monte Carlo simulation exercises. We also compare our proposed procedures not only to the co-skewness/co-kurtosis tests in Amengual, Fiorentini and Sentana (2022), but also to the Kolmogorov-Smirnov test that imposes the independent copula under the null, like in Blum, Kiefer and Rosenblat (1961), as well as to the normalized marginal ranks version of the Matteson and Tsay (2017) distance covariance statistic, which is also consistent against all types of dependence.

5.1 Design and computational details

To keep CPU time within bounds, we focus on bivariate and trivariate DGPs with VAR(1) dynamics. Specifically, we generate samples of size T from the following processes:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1/2 & 1/4 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 & 1/2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \end{pmatrix} \tag{28}$$

and

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1/2 & 1/4 & 1/8 \\ 0 & 1/3 & 1/9 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \\ \varepsilon_{3t}^* \end{pmatrix}. \tag{29}$$

Our PML estimation procedure, though, assumes that the drift vector, the matrix of autoregressive coefficients and the matrix of the impact multipliers are fully unconstrained, so it does not exploit the upper triangularity of the last two. Given that our tests are asymptotically pivotal with respect to the parameters of the companion matrix of the VAR, their true values are largely irrelevant in our simulations. More importantly, our results below do not depend on the true values of $\boldsymbol{\tau}$ or $\mathbf{C}$ either as long as $rank(\mathbf{C}) = N$ because we can show that the estimated shocks are numerically invariant to full-rank multivariate affine transformations of the y's, and the same is true of the different test statistics.

We consider both $T = 250$, which is realistic in most macro applications, and $T = 1,000$, which is representative of empirical finance ones. In the next subsection, we describe in detail our estimation method. Next, in section 5.1.2, we characterise the precise DGPs we consider for the shocks. Finally, we outline the resampling procedures that we use in section 5.1.3.

5.1.1 Estimation details

To estimate the model parameters by non-Gaussian PMLE, we assume that each shock ε^*_{it} is serially and cross-sectionally identically and independently distributed as a standardised discrete mixture of two normals, or $\varepsilon^*_{it} \sim DMN(\delta_i, \varkappa_i, \lambda_i)$ for short, so that

$$\varepsilon^*_{it} = \begin{cases} N[\mu^*_1(\boldsymbol{\varrho}_i), \sigma^{*2}_1(\boldsymbol{\varrho}_i)] \text{ with probability } \lambda_i \\ N[\mu^*_2(\boldsymbol{\varrho}_i), \sigma^{*2}_2(\boldsymbol{\varrho}_i)] \text{ with probability } 1-\lambda_i \end{cases} \tag{30}$$

where $\boldsymbol{\varrho}_i = (\delta_i, \varkappa_i, \lambda_i)'$,

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

Thus, 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.[12]

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

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

where $\phi(\varepsilon; \mu, \sigma^2)$ denotes the pdf of a Gaussian random variable with mean μ and variance σ^2 evaluated at ε. We jointly maximise the log-likelihood with respect to the N elements of $\boldsymbol{\tau}$,

[12] We can trivially extend this procedure to three or more components if we replace the normal random variable in the first branch of (30) by a k-component normal mixture with mean and variance given by $\mu^*_1(\boldsymbol{\varrho})$ and $\sigma^{*2}_1(\boldsymbol{\varrho})$, respectively, so that the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

the N^2 elements of $\mathbf{A}$, the N^2 elements of $\mathbf{C}$, and the $3N$ shape parameters. Without loss of generality, we also restrict $\varkappa_i \in (0, \infty)$ which in turn ensures the strict positivity of $\sigma_2^{*2}(\boldsymbol{\varrho}_i)$. Finally, we impose $\lambda_i \in (0, 1)$ to avoid degenerate mixtures.[13]

We maximise the log-likelihood subject to these constraints on the shape parameters using a derivative-based quasi-Newton algorithm, which converges quadratically in the neighbourhood of the optimum.[14] To exploit this property, we start the iterations by obtaining consistent initial estimators as follows. First, we compute OLS estimates of the VAR parameters $\boldsymbol{\tau}$ and $\mathbf{a}$, $\bar{\boldsymbol{\tau}}_{OLS}$ and $\bar{\mathbf{a}}_{OLS}$. We then apply the *FastICA*[15] algorithm of Gävert, Hurri, Särelä, and Hyvärinen to the residuals $\mathbf{y}_t - \bar{\boldsymbol{\tau}}_{OLS} - \bar{\mathbf{A}}_{OLS}\mathbf{y}_{t-1}$, obtaining $\overline{\mathbf{C}}_{FICA}$. Finally, we obtain initial values of the shape parameters of each shock by performing 20 iterations of the expectation maximisation (EM) algorithm in Dempster, Laird and Rubin (1977) on each of the elements of $\bar{\boldsymbol{\varepsilon}}^*_{t,FICA} = \overline{\mathbf{C}}^{-1}_{FICA}\left(\mathbf{y}_t - \bar{\boldsymbol{\tau}}_{OLS} - \bar{\mathbf{A}}_{OLS}\mathbf{y}_{t-1}\right)$.[16]

As we mentioned before, Assumption 1 only guarantees the identification of $\mathbf{C}$ up to sign changes and column permutations. 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 et al. (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 necessary changes to the shape parameters estimates, and in particular to the sign of δ_i. In any event, our tests are not numerically affected by these choices.

5.1.2 DGPs under the null and the alternative

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

DGP 0: In the bivariate case, ε^*_{1t} follows a Student t with 10 degrees of freedom (and kurtosis coefficient equal to 4), and ε^*_{2t} is generated as an asymmetric t with kurtosis and skewness coefficients equal to 4 and $-.5$, respectively, so that $\beta = -1.354$ and $\nu = 18.718$ in the notation of Mencía and Sentana (2012). These first two shocks share the same distributions in the trivariate case, while ε^*_{3t} follows an asymmetric t with the same kurtosis but opposite skewness coefficient as ε^*_{2t}.

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

[14] This maximization can be made effectively unconstrained by a suitable reparametrisation. In particular, we consider $\lambda = 2/T + (1 - 4/T)(1 + e^{-h_1\lambda^*})^{-1}$ and $\varkappa = \underline{\varkappa} + e^{-h_2\varkappa^*}$ where h_1 and h_2 are arbitrary constants that control the slope of the functions, which we set to 1.

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

[16] 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.

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

DGP 1: Standardised scale mixture of two zero mean normals with scalar covariance matrices in which the higher variance component has probability $\lambda = 0.8$ and the ratio of the two variances is $\varkappa = 0.05$.

DGP 2: Multivariate discrete mixture of two normals with parameters

$$\boldsymbol{\delta}_2 = \begin{pmatrix} 0.5 \\ \text{-}0.5 \end{pmatrix} \text{ and } \aleph_2 = \begin{pmatrix} 0.2 & 0 \\ 0.2 & 0.2 \end{pmatrix}, \text{ or } \boldsymbol{\delta}_3 = \begin{pmatrix} 0.5 \\ \text{-}0.5 \\ 0 \end{pmatrix} \text{ and } \aleph_3 = \begin{pmatrix} 0.2 & 0 & 0 \\ 0.2 & 0.2 & 0 \\ 0.2 & 0.2 & 0.2 \end{pmatrix}$$

for the bivariate and trivariate cases, respectively. In both cases, the mixing probability is set to $\lambda = 0.7$ (see Appendix D in Amengual, Fiorentini and Sentana (2023) for details).

DGP 3: Asymmetric Student t with skewness vector $\boldsymbol{\beta} = -10\boldsymbol{\ell}_N$ and degrees of freedom parameter $\nu = 12$.

Panels A–D of Figure 2 display the contours of the copula densities associated to DGP 0–3 in the bivariate case.

5.1.3 Resampling procedures

We follow Matteson and Tsay (2017) and Davis and Ng (2023) in reshuffling the estimated standardised residuals as follows. For each Monte Carlo sample, we generate another B samples of size T that impose the null by generating NT draws R_{is} from random permutations of the vector $(1, \ldots, T)$ independently drawn for each shock, which we then use to construct

$$\tilde{\mathbf{y}}_s = \hat{\boldsymbol{\tau}}_T + \hat{\mathbf{A}}_T \tilde{\mathbf{y}}_{s-1} + \hat{\mathbf{C}}_T \tilde{\boldsymbol{\varepsilon}}_s^*,$$

where $\tilde{\varepsilon}_{is}^* = \hat{\varepsilon}_{iR_{is}}^*$ and $\hat{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T) = \hat{\mathbf{C}}_T^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\tau}}_T - \hat{\mathbf{A}}_T \mathbf{y}_{t-1})$ are the estimated residuals in said Monte Carlo sample.[17]

5.1.4 Simulation results

To gauge the finite sample size and power of our proposed independence tests, we generate $10,000$ samples for the designs under the null and $2,500$ for those under the alternative. For each sample, we also compute $B = 99$ random permutation samples that impose the null to obtain resampling critical values, as explained in the previous subsection.

[17] Two implications of this approach is that the marginal empirical cdfs do not include jumps of size bigger than 1/T and that the tails of the shocks are the same in the actual and simulated data (see e.g. Camponovo, Scaillet and Trojani (2012) for the effects that influential observations may have on the reliability of resampling procedures).

In Table 1 we report the results on the finite sample size of the independence tests proposed in sections 3 and 4 for $T = 250$ and $T = 1,000$ in the bivariate case, and $T = 250$ in the trivariate one. As can be observed, overall, the size of the tests is quite accurate and the resampling procedures tend to adjust the slight size distortions of the discrete grid test when $T = 250$. In particular, the Monte Carlo rejection rates are not significantly different from the nominal ones in any of the cases, although the continuous grid test with $N = 3$ and $T = 250$ is moderately undersized. Interestingly, the size of the quantile-based discrete grid test does not deteriorate when the dimension of the partition becomes larger. For example, when $N = 3$ and $H = 5$, the size of the test is acceptable when $T = 250$ even though it is effectively based on $H^3 = 125$ moment conditions. Remarkably, the continuous grid test is not very sensitive to the choice of the regularisation parameter α either, with stable results over the interval of values we have experimented with, namely $\alpha \in [1e^{-5}, 1e^{-8}]$.

In turn, Tables 2, 3 and 4 display the simulation results on finite sample power for the three other DGPs in section 5.1.2. For comparison, we have also included the power of the integer moment tests based on the influence functions (3) in Amengual, Fiorentini and Sentana (2022), the Kolmogorov-Smirnov (KS) test that imposes the independent copula under the null, like in Blum, Kiefer and Rosenblat (1961), and the normalised marginal ranks version of the Matteson and Tsay (2017) (MT) test.

Under DGP 1 (scale mixtures of normals), our contingency table tests with estimated quantiles have substantially more power than the tests based on integer cross-moments of third- and fourth-order. When $T = 250$ and $N = 2$, the discrete grid test is better than the continuous one when $H \leq 3$, but it becomes worse for larger values of H. The discrete grid test is also the best for $T = 250$ and $N = 3$ or $T = 1,000$ and $N = 2$. In contrast, the tests based on integer cross moments largely fails to detect the dependence among the structural components when $T = 250$, and only displays limited power for $T = 1,000$. The MT and KS tests have little power when the sample size is small ($T = 250$) but their performance improves substantially when $T = 1,000$ even though they both underperform our proposed tests.

When the true distribution is a mixture of two multivariate Gaussian components (DGP 2), the power of the continuous grid test is very close to 1 in all cases. Still, the discrete grid test performs very well, especially when $H \geq 3$. The integer moment test is again the worst, as it only has an acceptable power when $T = 1,000$. The MT test also performs very well in all cases, but it is slightly worse than the continuous grid test. In turn, the power of the KS test deteriorates for $N = 3$.

Under DGP 3 (asymmetric Student t), the integer moment and MT tests are the most powerful, with most of the power of the former coming from the co-skewness component. This is perhaps not surprising given that the influence functions that this test uses coincide with the ones underlying the LM tests for a Gaussian copula with arbitrary correlation, which includes the independent one, versus an asymmetric Student t copula in Amengual and Sentana (2020). Nevertheless, the continuous grid test performs reasonably well and it is better than the discrete grid version, whose performance is similar to the one of the KS test. When $T = 1,000$, all tests have power close to one.

Finally, notice that the power of the tests is larger in the trivariate case than in the bivariate one except for KS, a fact that is most evident for the continuous grid test, which is the best overall. From the computational point of view, though, the cross-moment tests and the finite grid tests are the fastest, especially for large T.

6 Empirical application to volatility indices

We consider the same three series of market-based implied volatilities that Fiorentini and Sentana (2023) used, namely the VIX index, the EVZ EuroCurrency ETF volatility index and the GVZ Gold ETF volatility index. They represent three of the most actively traded asset classes, namely stocks, exchange rates and commodities, and since their inception have become incredibly popular among academics, financial market practitioners and commentators. Our sample spans from January 7th, 2009 to June 21st, 2023, a total of 753 weekly observations.[18]

Let $\mathbf{x}_t = (x_{GVZ,t}, x_{EVZ,t}, x_{VIX,t})'$ denote the log-transformation of these volatility indexes, which we depict in Figure 3a. As can be seen, they show mean reversion over the long run but persistent deviations from the mean during extended periods. We can also identify specific events, such as various phases of the European Sovereign Debt crises, the February 2018 scare or the onset of the COVID-19 pandemic.

To study the dynamic linkages between them, we estimate the following trivariate SVAR(2) model

$$\mathbf{y}_t = \begin{pmatrix} \underset{(.031)}{.068} \\ \underset{(.032)}{.135} \\ \underset{(.045)}{.146} \end{pmatrix} + \begin{pmatrix} \underset{(.037)}{.906} & \underset{(.035)}{-.002} & \underset{(.024)}{-.006} \\ \underset{(.039)}{.137} & \underset{(.036)}{.664} & \underset{(.024)}{.043} \\ \underset{(.055)}{.155} & \underset{(.051)}{-.045} & \underset{(.035)}{.663} \end{pmatrix} \mathbf{y}_{t-1} + \begin{pmatrix} \underset{(.037)}{.068} & \underset{(.035)}{.014} & \underset{(.024)}{-.011} \\ \underset{(.038)}{-.122} & \underset{(.036)}{.266} & \underset{(.025)}{-.033} \\ \underset{(.054)}{-.126} & \underset{(.051)}{.082} & \underset{(.035)}{.227} \end{pmatrix} \mathbf{y}_{t-2} + \begin{pmatrix} \underset{(.004)}{.090} & \underset{(.004)}{.011} & \underset{(.004)}{.024} \\ \underset{(.004)}{.030} & \underset{(.005)}{.090} & \underset{(.004)}{.032} \\ \underset{(.006)}{.023} & \underset{(.006)}{.016} & \underset{(.006)}{.141} \end{pmatrix} \boldsymbol{\varepsilon}_t^*$$

whose lag length we selected by looking at the Akaike information criterion and the likelihood

[18] The series, which are compiled by the Chicago Board of Options Exchange (CBOE), can be freely downloaded from the St. Louis FRED site. To minimise the effect of holidays, we focus on Wednesday observations, filling forward the previous day data when they are not available.

ratio test for lack of residual serial correlation.

As in the Monte Carlo section, we consistently estimate $(\boldsymbol{\tau}, \mathbf{a}, \mathbf{c}, \boldsymbol{\varrho})$ by PML assuming that the structural shocks follow a two-component normal mixture, or $\varepsilon^*_{it} \sim i.i.d.\ DMN(\delta_i, \kappa_i, \lambda_i)$ for short. For initial values, once again we run OLS regressions for each of the three variables ($\boldsymbol{\tau}$ and $\mathbf{a}$), apply the fastICA routine to the OLS residuals ($\mathbf{c}$), and finally, employ the EM algorithm for the mixture parameters $\boldsymbol{\varrho}$.

The estimated structural shocks are shown in Figure 3b. Reassuringly, they appear to be serially *i.i.d.* but highly non-normal. Specifically, their skewness and excess kurtosis coefficients are $(0.095, 1.134, 1.248)$ and $(1.200, 3.901, 4.027)$, respectively, which are highly significantly different from 0.

If we use the "unmixing" matrix $\mathbf{C}^{-1}$ to interpret the shocks as "long/short portfolios" of the one-period ahead prediction errors, we find that each of them "invests" approximately between 133 and 187% on one of the reduced form shocks and between -13 and -54% on each of the other two.[19] We can get a more standard interpretation by looking at the IRFs and FEVDs up to a year ahead depicted in Figure 4. The strong persistence implied by the SVAR(2) parameter estimates implies that all the IRFs decay rather slowly. As can be seen, each series reacts mostly to one shock. Nevertheless, they also react significantly to the other ones, especially in the case of the Gold volatility index and to some extent the VIX.

The main objective of our empirical exercise, though, is to assess whether the structural shocks are stochastically independent. In this respect, the different test procedures that we have considered fail to reject the null hypothesis, with the exception of the co-skewness and co-kurtosis tests in Amengual, Fiorentini and Sentana (2022). Nevertheless, this rejection seems to be closely associated to the unusual behaviour of the three series at the onset of the COVID-19 pandemic (see Ng (2021) for a discussion of pandemic shocks in macroeconometric models). Specifically, if we remove two additive outliers from the observations on the three series for March 11 and 18, 2020, using the procedures in Chen and Liu (1993), the integer moment tests no longer reject. In contrast, the quantile-based independence tests that we have proposed in this paper, and indeed the MT and KS tests, seem far more robust to the presence of these unusual observations.

7 Conclusions and directions for further research

Identification of SVAR models through independent non-Gaussian shocks is a very powerful tool. At the same time, it is not without concerns, as forcefully argued by Montiel-Olea,

[19] Specifically, the (normalised) rows of $\mathbf{C}^{-1}$ are (1.33, -0.13, -0.20), (-0.53, 1.87, -0.33) and (-0.31, -0.23, 1.55).

Plagborg-Møller and Qian (2022). In particular, given that the parametric identification of the structural shocks and their impact coefficients $\mathbf{C}$ in the SVAR model in (2) critically hinges on the validity of the identifying restrictions in Assumption 1, as we illustrated in section 5.3 of Amengual, Fiorentini and Sentana (2022), 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. The specification tests that we propose in this paper can be very useful in this respect.

Our tests effectively check that the joint distribution function of some or all of the structural shocks is the product of their marginal distribution functions. We do so first for a finite grid of values for the arguments of the distribution functions, explicitly relating our proposed test to Pearson's test for independence in contingency tables. But then we extend them to a continuum of values, which results in consistent tests different from the existing ones in the literature. Importantly, we explicitly consider the sampling variability resulting from using shocks computed with consistent parameter estimators. We study the finite sample size of our tests in several simulation exercises and discuss some resampling procedures. We also show that our tests have non-negligible power against a variety of empirically plausible alternatives.

Finally, we fail to reject independence of the shocks in an application to three volatility indices, while simultaneously highlighting the lack of robustness of tests based on third- and fourth-order cross-moments to the unusual observations right at the beginning of the COVID-19 pandemic.

Most of the existing empirical applications that rely on cross-sectionally independent shocks make use of estimators for the parameters of the static ICA model (1) or the dynamic SVAR (2) which differ from the ones we have considered in this paper. In this respect, the fact that the only parameters whose sampling variability matter for our discrete or continuous grid copula-based tests are the off-diagonal elements of $\mathbf{C}$, combined with Proposition 1 in Fiorentini and Sentana (2023), which says that if we reparametrise $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, with $\boldsymbol{\Psi}$ a diagonal matrix whose elements contain the scale of the structural shocks, then $\mathbf{a}$ and the off-diagonal elements of $\mathbf{J}$ will be consistently estimated even if we do not rely on finite Gaussian mixtures, implies that our tests would continue to be valid for models estimated with alternative marginal distributions of the shocks. Obviously, the asymptotic covariance matrices that take into account their sampling variability will depend on the specific estimation method used.

The numerical invariance of our tests to the estimators of $\boldsymbol{\tau}$ and the diagonal of $\boldsymbol{\Psi}$ also suggests that our approach may be robust (in the statistical sense of the word) to the presence of outliers in shocks with fat tails, which will affect mostly the estimation of the mean parameters and the scale of the shocks rather than their quantiles. Studying this issue in more detail along the lines of Davis and Ng (2023) constitutes an interesting topic for further research.

The moment conditions that we consider for testing independence could also form the basis of a GMM estimation procedure for the model parameters $\boldsymbol{\theta}$ along the lines of Lanne and Luoto (2021), although with either a much larger but finite set of moments or a continuum of them. The overidentification restrictions tests obtained as a by product of such procedures could be used as a specification test of the assumed cross-sectional independence assumption.

Similarly, we could consider related tests of independence that exploit the fact that the joint characteristic function is the product of the marginal characteristic functions under the independence null, along the lines of Csörgó (1985), but using an overidentification test for a continuum of moment conditions, as in Amengual, Carrasco and Sentana (2020), rather than the Cramér-von Mises and Kolmogorov-Smirnov functionals that he used.

The behaviour of our proposed tests in the other extreme case in which the true joint distribution of the shocks is Gaussian is also of interest. If the parameters in $\boldsymbol{\theta}$ were known, our independence test will continue to work without any problem, as the assumption of mutually independent shocks will be automatically guaranteed by the combination of multivariate normality with the orthogonality of the shocks. However, the parameters in $\mathbf{C}$ will no longer be identified, which will affect the distribution of their estimators, as Hoesch, Lee and Mesters (2022) have recently shown. The extent to which this will also affect the independence tests remains unknown.

Finally, it should also be of interest to apply our independence tests to the shocks of SVAR models identified using some of the more traditional methods mentioned in the introduction, even when they have been estimated by Gaussian PMLE. The reason is that most of the theoretical macroeconomic models that justify those identifying strategies implicitly assume the independence of the underlying economic shocks. We are currently pursuing some of these research avenues.

References

Amengual, A., Carrasco, M. and Sentana, E. (2020): "Testing distributional assumptions using a continuum of moments", *Journal of Econometrics* 218, 655-689.

Amengual, D., Fiorentini, F. and Sentana, E. (2013): "Sequential estimators of shape parameters in multivariate dynamic models", *Journal of Econometrics* 177, 233-249.

Amengual, D., Fiorentini, F. and Sentana, E. (2022): "Moment tests for independent components", *SERIEs* 13, 429-474.

Amengual, D., Fiorentini, F. and Sentana, E. (2023): "PML vs minimum χ^2: the comeback", *SERIEs* 14, 253-300.

Amengual, D. and Sentana, E. (2020): "Is a normal copula the right copula?", *Journal of Business and Economic Statistics* 38, 350-366.

Anderson, T. W. and Darling, D. A. (1952): "Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes", *Annals of Mathematical Statistics* 23, 193–212.

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

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961): "Distribution free tests of independence based on the sample distribution function", *Annals of Mathematical Statistics* 32, 485-498

Brunnermeier, M., Palia, D., Sastry, K.A. and Sims, C.A. (2021): "Feedbacks: financial markets and economic activity", *American Economic Review* 111, 1845-1879.

Camponovo, L., Scaillet, O. and Trojani, F. (2012): "Robust subsampling", *Journal of Econometrics* 167, 197-210.

Carrasco, M., Florens, J.-P. (2000): "Generalization of GMM to a continuum of moment conditions", *Econometric Theory* 16, 797–834.

Carrasco, M., Florens, J.-P. and Renault, E. (2007): "Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization", in J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics* vol. 6B, 5633–5751, Elsevier.

Cheng, C. and Liu, L.-M. (1993): "Joint estimation of model parameters and outlier effects in time series", *Journal of the American Statistical Association* 88, 284-297.

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

Csörgó, S. (1985): "Testing for independence by the empirical characteristic function", *Journal of Multivariate Analysis* 16, 290-299.

Davis, R. and Ng, S. (2023): "Time series estimation of the dynamic effects of disaster shocks", *Journal of Econometrics* 235, 180-201.

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.

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

Fisher, R. A (1922): "On the interpretation of χ^2 from contingency tables, and the calculation of p", *Journal of the Royal Statistical Society* 85, 87-94.

Fiorentini, G. and Sentana, E. (2022): "Consistent estimation with finite mixtures", mimeo, CEMFI.

Fiorentini, G. and Sentana, E. (2023): "Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions", *Journal of Econometrics* 235, 643-665.

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

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

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029–1054.

Hoeffding, W. (1948): "A nonparametric test for independence", *Annals of Mathematical Statististics* 19, 546-557.

Hoesch, L., Lee, A. and Mesters, G. (2022): "Robust inference for non-Gaussian SVAR models", mimeo, UPF.

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.

Hyvärinen, A., Zhang, K., Shimizu, S. and Hoyer, P.O. (2010): "Estimation of a structural vector autoregression model using non-Gaussianity", *Journal of Machine Learning Research* 11, 1709-1731.

Imhof, J.P. (1961): "Computing the distribution of quadratic forms in normal variables", *Biometrika* 48, 419–426.

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

Kheifets, I.L. (2015): "Specification tests for nonlinear dynamic models", *Econometrics Journal* 18, 67–94.

Khmaladze, E.V. (1981): "Martingale approach in the theory of goodness-of-tests", *Theory* *of probability and its applications* 26, 240-257.

Kress, R. (1999): *Linear integral equations*, Springer-Verlag.

Lanne, M. and Luoto, J. (2021): "GMM estimation of non-Gaussian structural vector autoregressions", *Journal of Business and Economic Statistics* 39, 69–81.

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

Matteson, D. S. and Tsay, R. S. (2017): "Independent component analysis via distance covariance", *Journal of the American Statistical Association* 112, 623–637.

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., Entner, D., Hoyer, P.O. and Coad, A. (2013): "Causal inference by independent component analysis: theory and applications", *Oxford Bulletin of Economics and Statistics* 75, 705–730.

Moneta, A. and Pallante, G. (2022): "Identification of structural VAR models via Independent Component Analysis: a performance evaluation study", *Journal of Economic Dynamics and Control* 144, 104530.

Montiel-Olea, J.L., Plagborg-Møller, M. and Qian, E. (2022): "SVAR identification from higher moments: has the simultaneous causality problem been solved?", *AEA Papers and Proceedings* 112, 481-485.

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

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587–99.

Ng, S. (2021): "Modeling macroeconomic variation after COVID-19", mimeo, Columbia University.

Rackauskas, A. and Suquet, C. (2006): "Testing epidemic changes of infinite dimensional parameters", *Statistical Inference for Stochastic Processes* 9, 111-134.

Scaillet, O. (2005): "A Kolmogorov-Smirnov type test for positive quadrant dependence", *Canadian Journal of Statistics* 3, 415-427.

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

Sentana, J. (2022): "Tests for independence between categorical variables", *Economics Let-*

ters 220, 110850.

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.

Appendix

Proofs of Propositions

Preliminaries

Given that

$$\frac{\partial 1_{(-\infty,k)}(\varepsilon_{it}^*)}{\partial \varepsilon_{it}^*} = \frac{\partial \Delta_k(\varepsilon_{it}^*)}{\partial \varepsilon_{it}^*} = -\delta_{\{\varepsilon_{it}^* - k\}},$$

where $\delta_{\{.\}}$ denotes the Dirac delta function, we have

$$E\left[\frac{\partial 1_{(-\infty,k)}(\varepsilon_{it}^*)}{\partial \varepsilon_{it}^*}\right] = -f_i(k).$$

This is due to the fact that $1_{(-\infty,k)}(\varepsilon_{it}^*)$ is a shifted and flipped Heaviside step function, i.e. the indicator function of the one-dimensional positive half-line, whose distributional derivative is equal to the Dirac delta function. Specifically, since $1_{(0,\infty)}(x) = \delta(x)$ and

$$\int_{-\infty}^{\infty} \delta(x) f_i(x) dx = f_i(0),$$

then

$$E_0[\delta_{\{\varepsilon_{it}^* - k\}}] = f_i(k).$$

We will also exploit the related fact that

$$E_0[\varepsilon_{it}^* \delta_{\{\varepsilon_{it}^* - k\}}] = k f_i(k).$$

Proposition 1

Regardless of the independence between the shocks, we have that, first,

$$F_i(k_{h_i}) = \Pr(\varepsilon_{it}^* \leq k_{h_i}) = \sum_{j_i=1}^{h_i} \Pr(k_{j_i-1} \leq \varepsilon_{it}^* \leq k_{j_i}),$$

which implies that the $\hat{u}_h^{\diamond i}$ will be replaced by the values of the empirical cdf at the chosen grid points, say $\hat{u}_h^i$; and second,

$$F_{\mathbf{k}}(k_{h_i}, k_{h_{i'}}, ..., k_{h_{i_M}}) = \Pr\left[\bigcap_{i \in M} \{\varepsilon_{it}^* \leq k_{h_i}\}\right] = \sum_{i \in M} \sum_{j_i=1}^{h_i} \Pr\left[\bigcap_{i \in M} \{k_{j_i-1} \leq \varepsilon_{it}^* \leq k_{j_i}\}\right].$$

In addition, it is also easy to see that under the independence null

$$F_{\mathbf{k}}(k_{h_i}, k_{h_{i'}}, ..., k_{h_{i_M}}) = \Pr\left[\bigcap_{i \in M} \{\varepsilon_{it}^* \leq k_{h_i}\}\right] = \prod_{i \in M}\left[\sum_{j_i=1}^{h_i} \Pr(k_{j_i-1} \leq \varepsilon_{it}^* \leq k_{j_i})\right] = \prod_{i \in M} F_i(k_{h_i})$$

because

$$\Pr\left[\bigcap_{i\in M}\{k_{j_i-1}\le\varepsilon^*_{it}\le k_{j_i}\}\right]=\prod_{i\in M}\Pr(k_{j_i-1}\le\varepsilon^*_{it}\le k_{j_i})\quad\forall i\in M \text{ and } \forall k_{j_i}, j\in H,$$

whence the result follows. □

Proposition 2

To simplify the notation, in what follows let $u_i = u^i_{h_i}$ and $v_i = u^i_{h'_i}$. Regarding part a), linearising the influence function (10) when the k_{h_i}'s are fixed yields

$$\sqrt{T}\,[\hat{u}_i(k_{h_i})-u_i]=\frac{\sqrt{T}}{T}\sum_{t=1}^{T}1_{(-\infty,k_{h_i})}(\varepsilon^*_{it})-u_i+o_p(1),$$

while for the influence function (11), we have that

$$\frac{\partial}{\partial u_i}E[p(\boldsymbol{\varepsilon}^*_t;\mathbf{k},\mathbf{u})]=\frac{\partial}{\partial u_i}E\left[\prod_{i\in M}1_{(-\infty,k_{h_i})}(\varepsilon^*_{it})-\prod_{i\in M}u_i\right]=-\prod_{i'\in M,i'\neq i}u_{i'}.$$

Thus, it is clear that (12) is the linearised influence function that takes into account the estimation of u_i, for $i\in M$. Analogously, when u_i's are fixed, linearising the influence function (10) that defines $\hat{\varkappa}_i(u_i)$ yields

$$\sqrt{T}\,[\hat{\varkappa}_i(u_i)-\varkappa_i(u_i)]=-\frac{1}{f_i[\varkappa_i(u_i)]}\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it})-u_i\right]+o_p(1),$$

while

$$\begin{aligned}\frac{\partial}{\partial\varkappa_i}\left[\prod_{i\in M}1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it})-\prod_{i\in M}u_i\right] &= \left(\prod_{i'\in M,i'\neq i}u_{i'}\right)E\left[\frac{\partial 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it})}{\partial\varkappa_i}\right]\\ &= \left(\prod_{i'\in M,i'\neq i}u_{i'}\right)f_i[\varkappa_i(u_i)].\end{aligned}$$

for (11). Then, the result follows by choosing the limits of the intervals k^i_h's so that they exactly match the theoretical quantiles $\varkappa_i(u_h)$'s.

As for part b), we have to compute

$$
\begin{aligned}
E[m_t(\mathbf{u})m_t(\mathbf{u}')] &= E\left\{\left[\prod_{i\in M}1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-\prod_{i\in M}u_i\right]\left[\prod_{j\in M}1_{(-\infty,k_{h_j})}(\varepsilon_{jt}^*)-\prod_{j\in M}u_j'\right]\right\}\\
&\quad -E\left[\left[\prod_{i\in M}1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-\prod_{i\in M}u_i\right]\left\{\sum_{j\in M}\left[1_{(-\infty,k_{h_j})}(\varepsilon_{jt}^*)-u_j'\right]\prod_{j'\in M,j'\neq j}u_{j'}'\right\}\right]\\
&\quad -E\left[\left\{\sum_{i\in M}\left[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-u_i\right]\prod_{i'\in M,i'\neq i}u_{i'}\right\}\left[\prod_{j\in M}1_{(-\infty,k_{h_j})}(\varepsilon_{jt}^*)-\prod_{j\in M}u_j'\right]\right]\\
&\quad +E\left[\left\{\sum_{i\in M}\left[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-u_i\right]\prod_{i'\in M,i'\neq i}u_{i'}\right\}\right.\\
&\quad \left.\times\left\{\sum_{j\in M}\left[1_{(-\infty,k_{h_j})}(\varepsilon_{jt}^*)-u_j'\right]\prod_{j'\in M,j'\neq j}u_{j'}'\right\}\right].
\end{aligned}
$$

Regarding the first term,

$$
\begin{aligned}
&E\left\{\left[\prod_{i\in M}1_{(-\infty,k_{h_i}]}(\varepsilon_{it}^*)-\prod_{i\in M}u_i\right]\left[\prod_{j\in M}1_{(-\infty,k_{h_j}]}(\varepsilon_{jt}^*)-\prod_{j\in M}u_j'\right]\right\}\\
=\;&E\left\{\left[\prod_{i\in M}1_{(-\infty,k_{h_i}]}(\varepsilon_{it}^*)\right]\left[\prod_{j\in M}1_{(-\infty,k_{h_j}]}(\varepsilon_{jt}^*)\right]\right\}+\left(\prod_{i\in M}u_i\right)\left(\prod_{j\in M}u_j'\right)\\
&-E\left\{\left(\prod_{j\in M}v_j\right)\left[\prod_{i\in M}1_{(-\infty,k_{h_i}]}(\varepsilon_{it}^*)\right]\right\}-E\left\{\left(\prod_{i\in M}u_i\right)\left[\prod_{j\in M}1_{(-\infty,k_{h_j}]}(\varepsilon_{jt}^*)\right]\right\}\\
=\;&\prod_{i\in M}\min(u_i,u_i')+\prod_{i\in M}u_iu_i'-2\prod_{i\in M}u_iu_i'\\
=\;&\prod_{i\in M}\min(u_i,u_i')-\prod_{i\in M}u_iu_i',
\end{aligned}
$$

where the second equality follows from expanding the product, and the last one from the fact that

$$E[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)]=u_i \text{ for } i\in M,$$

and

$$E[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)1_{(-\infty,k_{h_i'})}(\varepsilon_{it}^*)]=\min(u_i,u_i') \text{ for } i\in M. \tag{A.1}$$

Similarly, the second term becomes

$$
\begin{aligned}
& -E\left\{\left[\prod_{i\in M} 1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-\prod_{i\in M} u_i\right]\left[\sum_{j\in M}\left[1_{(-\infty,k_{h_j})}(\varepsilon_{jt}^*)-u_j'\right]\prod_{j'\in M,j'\neq j} u_{j'}'\right]\right\} \\
= & -\sum_{i\in M} E\left[\left[\prod_{i\in M} 1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)\right]\left\{\left[1_{(-\infty,k_{h_i'})}(\varepsilon_{it}^*)-u_i'\right]\prod_{i'\in M,i'\neq i} u_{i'}'\right\}\right] \\
= & \ \mathrm{M}\prod_{i\in M} u_i u_i' - \sum_{i\in M}\min(u_i,u_i')\left(\prod_{i'\in M,i'\neq i} u_{i'}\right)\left(\prod_{i'\in M,i'\neq i} u_{i'}'\right),
\end{aligned}
$$

where the first equality follows from the fact that

$$
E[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)-u_i]=0 \quad \text{for} \quad i\in M, \tag{A.2}
$$

and the last one from (A.1). By symmetry, the third term is the same.

Finally,

$$
\begin{aligned}
& E\left[\left\{\sum_{i\in M}\left[1_{(-\infty,k_{h_i}]}(\varepsilon_{it}^*)-u_i\right]\prod_{i'\in M,i'\neq i} u_{i'}\right\}\left\{\sum_{j\in M}\left[1_{(-\infty,k_{h_i}]}(\varepsilon_{jt}^*)-u_j'\right]\prod_{j'\in M,j'\neq j} u_{j'}'\right\}\right] \\
= & \ E\left[\left\{\sum_{i\in M}\left[1_{(-\infty,k_{h_i}]}(\varepsilon_{it}^*)-u_i\right]\prod_{i'\in M,i'\neq i} u_{i'}\right\}\left[1_{(-\infty,k_{h_i'}]}(\varepsilon_{it}^*)-u_i'\right]\prod_{i'\in M,i'\neq i} u_{i'}'\right] \\
= & \sum_{i\in M}\min(u_i,u_i')\left(\prod_{i'\in M,i'\neq i} u_{i'}\right)\left(\prod_{i'\in M,i'\neq i} u_{i'}'\right)-\mathrm{M}\prod_{i\in M} u_i u_i',
\end{aligned}
$$

where we have used (A.2) in the first equality and (A.1) in the second one.

Collecting the four terms, we finally get the desired result. □

Proposition 3

It follows from Proposition 1 in Amengual, Fiorentini and Sentana (2022), and Lemmas 1 and 2 in Supplemental Appendix B. □

Proposition 4

Let $\mathbf{u}_M$ denote the vector containing all the u_i's such that $i\in M$. Using the independence copula as weighting function, so that $\varpi(\mathbf{u})=1\ \forall\mathbf{u}$, we have to compute

$$
\langle q_t,q_s\rangle=\int_{[0,1]^{\mathrm{M}}} q_t(\mathbf{u}_M)\,q_s(\mathbf{u}_M)\,d\mathbf{u}_M,
$$

with

$$q_t\left(\mathbf{u}_M\right) q_s\left(\mathbf{u}_M\right) = \prod_{i\in M} 1_{(0,u_i)}(\epsilon_{it}) 1_{(0,u_i)}(\epsilon_{is}) - \prod_{i\in M} u_i 1_{(0,u_i)}(\epsilon_{it}) - \prod_{i\in M} u_i 1_{(0,u_i)}(\epsilon_{is}) + \prod_{i\in M} u_i^2, \quad \text{(A.3)}$$

where we have used (19) evaluated at the observations t and s. Next, we need to compute the integrals for each of the four terms of the right-hand side of (A.3). Regarding the first term, under the independence null,

$$\begin{aligned}\int_{[0,1]^{\mathrm{M}}} \left[\prod_{i\in M} 1_{(0,u_i)}(\epsilon_{it}) 1_{(0,u_i)}(\epsilon_{is})\right] d\mathbf{u}_M &= \prod_{i\in M}\left[\int_0^1 1_{(0,u_i)}(\max\{\epsilon_{it},\epsilon_{is}\}) du_i\right] \\ &= \prod_{i\in M}\left[\int_{\max\{\epsilon_{it},\epsilon_{is}\}}^1 du_i\right] = \prod_{i\in M}[1-\max(\epsilon_{it},\epsilon_{is})] \equiv d_1(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_s).\end{aligned}$$

As for the second and third ones,

$$\begin{aligned}\int_{[0,1]^{\mathrm{M}}} \left[\prod_{i\in M} 1_{(0,u_i)}(\epsilon_{it}) u_i\right] d\mathbf{u}_M &= \prod_{i\in M}\left[\int_0^1 1_{(0,u_i)}(\epsilon_{it}) u_i du_i\right] \\ &= \prod_{i\in M}\left[\int_{u_{it}}^1 u_i du_i\right] = \left(\frac{1}{2}\right)^{\mathrm{M}} \prod_{i\in M}(1-\epsilon_{it}^2) \equiv d_2(\boldsymbol{\epsilon}_t),\end{aligned}$$

Next, integrating the fourth term,

$$\int_{[0,1]^{\mathrm{M}}} \left(\prod_{i=1}^N u_i^2\right) d\mathbf{u}_M = \prod_{i\in M}\left(\int_0^1 u_i^2 du_i\right) = \left(\frac{1}{3}\right)^m \equiv d_3.$$

Finally, collecting them in $d_{ts} = d_1(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_s) - d_2(\boldsymbol{\epsilon}_t) - d_2(\boldsymbol{\epsilon}_s) + d_3$ delivers the desired result. □

Proposition 5

It follows from Proposition 4 and Lemma 3 in Supplemental Appendix B. □

Table 1: Monte Carlo size of independence tests based on quantiles

	Discrete grid tests							Continuous grid tests		
	Asymptotic critical values			Resampling critical values				Resampling critical values		
	10%	5%	1%	10%	5%	1%		10%	5%	1%
Panel A: $N = 2$, $T = 250$										
$H = 2$	8.4	4.0	0.6	9.2	4.6	0.7	$\alpha = 10^{-5}$	9.6	4.4	0.8
$H = 3$	8.4	4.0	0.9	9.2	4.5	0.8	$\alpha = 10^{-6}$	9.4	4.5	0.8
$H = 4$	8.1	4.0	0.7	8.8	4.3	0.9	$\alpha = 10^{-7}$	9.4	4.5	0.9
$H = 5$	8.6	4.4	0.8	9.2	4.4	0.8	$\alpha = 10^{-8}$	9.3	4.5	0.9
Panel B: $N = 3$, $T = 250$										
$H = 2$	8.0	3.7	0.7	8.9	4.1	0.9	$\alpha = 10^{-5}$	8.2	3.8	0.6
$H = 3$	8.2	3.8	0.8	9.2	4.3	0.9	$\alpha = 10^{-6}$	8.3	3.8	0.7
$H = 4$	8.3	3.8	0.7	9.3	4.4	0.8	$\alpha = 10^{-7}$	8.0	4.1	0.7
$H = 5$	8.5	4.1	0.9	9.5	4.5	0.9	$\alpha = 10^{-8}$	8.0	4.1	0.7
Panel C: $N = 2$, $T = 1,000$										
$H = 2$	9.8	4.7	1.0	10.1	4.9	1.1	$\alpha = 10^{-5}$	10.7	5.4	1.2
$H = 3$	9.2	4.3	0.9	9.4	4.8	0.9	$\alpha = 10^{-6}$	10.7	5.5	1.3
$H = 4$	9.7	5.0	1.0	10.2	5.3	1.1	$\alpha = 10^{-7}$	10.6	5.5	1.2
$H = 5$	9.6	4.6	0.8	10.3	5.2	1.0	$\alpha = 10^{-8}$	10.6	5.4	1.2

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 10,000 replications. Details on the data generating processes: DGP 0: ε^*_{1t} follows a Student t with 10 degrees of freedom (and kurtosis coefficient equal to 4), and ε^*_{2t} is generated as an asymmetric t with kurtosis and skewness coefficients equal to 4 and $-.5$, respectively; in addition, in the trivariate case ε^*_{3t} follows an asymmetric t with the same kurtosis but opposite skewness coefficient as ε^*_{2t}. See sections 3 and 4 for a detailed description of the discrete and continuous grid test statistics, respectively. The asymptotic distribution of the discrete grid test is chi-squared with H^N degrees of freedom. See section 5.1.3 for a description of the resampling procedures used for the discrete and continuous grid test statistics.

Table 2: Monte Carlo power of independence tests DGP 1: Scale mixture of two normals.

	$N=2$ $T=250$			$N=3$ $T=250$			$N=2$ $T=1,000$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete grid tests									
$H=2$	62.3	48.2	22.0	88.2	82.6	62.9	100.0	100.0	99.6
$H=3$	54.8	41.4	16.4	90.8	84.3	61.4	100.0	99.9	99.2
$H=4$	37.8	24.4	8.6	71.7	58.8	30.2	99.8	99.6	96.1
$H=5$	36.6	23.0	7.1	61.6	46.2	22.8	99.8	99.7	96.7
Panel B: Continuous grid tests									
$\alpha=10^{-5}$	48.2	35.0	12.4	69.3	54.5	24.2	99.5	98.5	91.1
$\alpha=10^{-6}$	46.5	33.0	11.7	68.5	53.9	23.3	98.6	96.7	83.5
$\alpha=10^{-7}$	45.9	32.3	11.5	67.6	53.6	23.3	97.6	95.3	79.9
$\alpha=10^{-8}$	45.7	32.2	11.4	66.9	51.9	21.6	97.4	94.8	78.4
Panel C: Integer moment tests									
Co-cov	8.2	3.7	0.8	9.6	5.0	1.3	11.3	6.0	1.2
Co-skew	9.4	4.4	0.6	9.6	4.7	1.0	11.6	6.0	1.5
Co-kurt	14.3	8.0	1.7	13.2	7.2	1.3	50.4	36.8	15.7
Joint	12.6	7.0	1.3	12.8	6.7	1.2	41.7	30.3	12.4
Panel D: Other tests									
MT	15.1	7.2	1.0	19.6	9.0	1.6	99.8	98.4	72.2
KS	18.9	9.6	2.0	11.3	5.7	1.0	82.0	69.9	36.1

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 2,500 replications. Details on the data generating process DGP 1: Standardised scale mixture of two zero mean normals –with scalar covariance matrix– in which the higher variance component has probability $\lambda = 0.2$ and the ratio of the variances is $\varkappa = 0.05$. For Panels A and B, see sections 3 and 4 for a detailed description of the discrete and continuous grid test statistics, respectively, and Amengual, Fiorentini and Sentana (2022) for a description of the integer moment tests in Panel C. In Panel D, MT and KS denote the Matteson and Tsay (2017) and Kolmogorov-Smirnov testing procedures, respectively. See section 5.1.3 for a description of the resampling procedures we used for discrete and continuous grid test statistics.

Table 3: Monte Carlo power of independence tests DGP 2: Finite normal mixture.

	$N=2$, $T=250$			$N=3$, $T=250$			$N=2$, $T=1,000$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete grid tests									
$H=2$	70.0	58.6	32.5	90.0	82.6	61.1	100.0	99.7	97.4
$H=3$	89.3	81.7	55.1	98.1	96.2	86.3	100.0	100.0	100.0
$H=4$	91.6	85.2	62.2	99.2	98.1	90.3	100.0	100.0	100.0
$H=5$	90.4	81.9	56.1	98.4	96.1	83.0	100.0	100.0	100.0
Panel B: Continuous grid tests									
$\alpha=10^{-5}$	95.7	91.6	72.1	99.6	99.0	93.1	100.0	100.0	100.0
$\alpha=10^{-6}$	95.0	90.1	69.8	99.6	98.8	92.7	100.0	100.0	100.0
$\alpha=10^{-7}$	94.7	90.0	69.8	99.6	98.7	92.6	100.0	100.0	100.0
$\alpha=10^{-8}$	94.7	90.1	69.7	99.3	98.6	91.9	100.0	100.0	100.0
Panel C: Integer moment tests									
Cov	28.7	19.8	8.4	33.2	23.1	9.0	33.9	26.0	14.4
Co-skew	32.7	21.3	7.7	38.8	25.8	9.4	84.4	76.0	53.4
Co-kurt	28.8	17.8	5.2	38.6	26.3	8.1	80.4	71.9	40.2
Joint	36.2	23.4	7.2	45.5	30.7	8.7	92.2	85.5	58.2
Panel D: Other tests									
MT	89.7	79.9	48.2	95.3	89.1	59.7	100.0	100.0	100.0
KS	76.2	65.5	40.4	28.6	16.8	4.5	99.9	99.9	98.5

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 2,500 replications. Details on the data generating process DGP 2: Multivariate discrete mixture of two normals with mixing probability $\lambda = 0.7$, relative-means difference $\boldsymbol{\delta}_3 = (0.5,\text{-}0.5,0)'$ and relative-covariance difference such that $\aleph_3$ is lower triangular with $vech(\aleph_3) = 0.2\boldsymbol{\ell}_6$ (see Appendix D in Amengual, Fiorentini and Sentana (2023) for details). For Panels A and B, see sections 3 and 4 for a detailed description of the discrete and continuous grid test statistics, respectively, and Amengual, Fiorentini and Sentana (2022) for a description of the integer moment tests in Panel C. In Panel D, MT and KS denote the Matteson and Tsay (2017) and Kolmogorov-Smirnov testing procedures, respectively. See section 5.1.3 for a description of the resampling procedures we used for discrete and continuous grid test statistics.

Table 4: Monte Carlo power of independence test: DGP 3: Asymmetric Student t.

	$N=2$ $T=250$			$N=3$ $T=250$			$N=2$ $T=1,000$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete grid tests									
$H=2$	29.5	19.0	6.2	24.4	15.2	2.0	87.1	78.4	53.4
$H=3$	34.9	23.0	7.0	19.6	11.2	2.8	95.1	90.2	72.6
$H=4$	33.5	22.3	7.6	20.4	10.8	2.8	97.2	93.4	78.6
$H=5$	33.5	21.1	6.3	16.8	9.6	1.2	96.6	93.6	78.4
Panel B: Continuous grid tests									
$\alpha=10^{-5}$	43.8	30.8	12.3	57.6	40.0	16.4	95.3	92.2	74.9
$\alpha=10^{-6}$	43.9	30.1	12.1	58.4	44.4	20.0	94.1	89.8	69.3
$\alpha=10^{-7}$	43.1	29.8	12.1	58.8	47.2	20.4	93.5	88.7	67.6
$\alpha=10^{-8}$	42.9	29.8	11.8	58.4	45.2	21.2	93.5	88.8	67.0
Panel C: Integer moment tests									
Cov	62.0	54.2	33.4	78.8	68.4	48.8	85.8	82.2	73.5
Co-skew	89.1	80.9	52.5	94.8	88.4	69.2	100.0	100.0	99.6
Co-kurt	60.4	48.8	23.0	73.2	66.4	40.8	95.7	91.1	70.4
Joint	84.2	67.7	26.0	94.0	78.4	43.2	100.0	100.0	96.3
Panel D: Other tests									
MT	70.4	58.3	29.6	76.0	63.2	32.4	100.0	100.0	99.7
KS	34.4	21.7	6.1	32.8	15.6	2.8	97.2	92.4	67.7

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 2,500 replications. Details on the data generating process DGP 3: Asymmetric Student t with skewness vector $\boldsymbol{\beta}=-10\boldsymbol{\ell}_N$ and degrees of freedom parameter $\nu=12$ (see Mencía and Sentana (2012) for details). For panels A and B, see sections 3 and 4 for a detailed description of the discrete and continuous grid test statistics, respectively, and Amengual, Fiorentini and Sentana (2022) for a description of the integer moment tests in Panel C. In Panel D, MT and KS denote the Matteson and Tsay (2017) and Kolmogorov-Smirnov testing procedures, respectively. See section 5.1.3 for a description of the sampling procedure we used for discrete and continuous grid test statistics.

Figure 1: Copula contours for spherically symmetric, Hermite polynomial expansions of bivariate normal

Figure 1a: $d_2 = 0$ and $d_3 = -0.35$

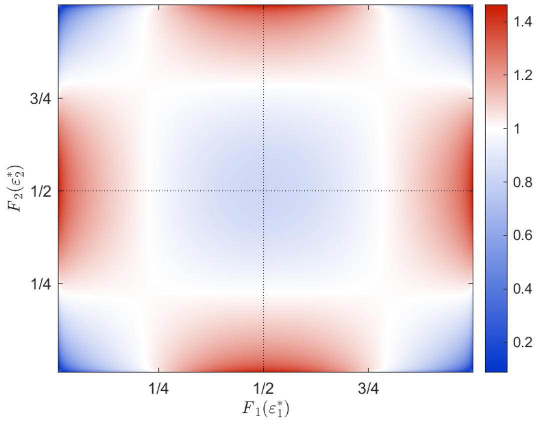
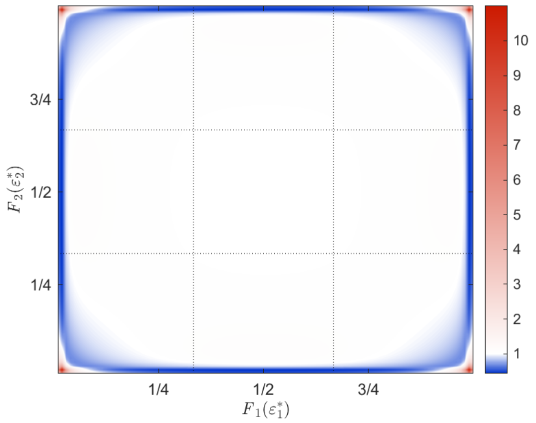


Figure 1b: $d_2 = 0.61$ and $d_3 = -0.39$



Notes: The copula density is given by

$$c(u_1, u_2; d_2, d_3) = \frac{f_2[F_1^{-1}(u_1; d_2, d_3), F_1^{-1}(u_2; d_2, d_3); d_2, d_3]}{f_1[F_1^{-1}(u_1; d_2, d_3); d_2, d_3] f_1[F_1^{-1}(u_2; d_2, d_3); d_2, d_3]},$$

where f_1 and f_2 denote the densities of spherically symmetric univariate and bivariate Hermite expansions of univariate and bivariate Gaussian distributions, respectively, which are obtained as Laguerre expansions of the corresponding generating χ_N^2 random variates ς, namely

$$h_N(\varsigma) = \frac{1}{2^{N/2}\Gamma(N/2)} \varsigma_t^{N/2-1} \exp\left(-\frac{1}{2}\varsigma\right) P_N(\varsigma), \text{ for } N = 1 \text{ and } N = 2,$$

and where $P_N(\varsigma) = \left[1 + d_2 p_{N/2-1,2}(\varsigma) + d_3 p_{N/2-1,3}(\varsigma)\right]$, with $p_{N/2-1,j}(.)$ denoting the generalized Laguerre polynomial of order j and parameter $N/2-1$ (see Amengual, Fiorentini and Sentana (2013) for the detailed expressions). In turn, $F_1^{-1}(u; d_2, d_3)$ denotes the corresponding inverse cdf of the univariate distribution.

Figure 2: Bivariate copula contours associated to the DGPs in section 5

Figure 2a: Independence (DGP 0)

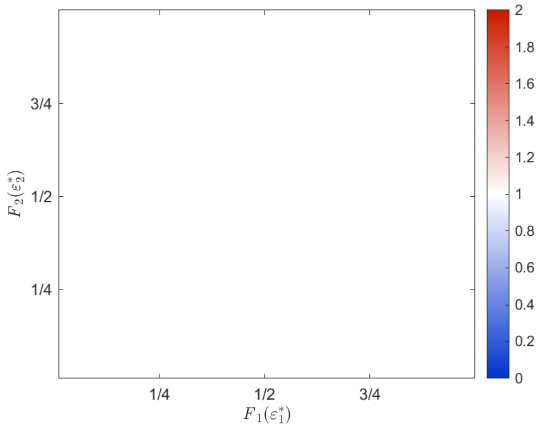


Figure 2b: Scale mixture of normals (DGP 1)

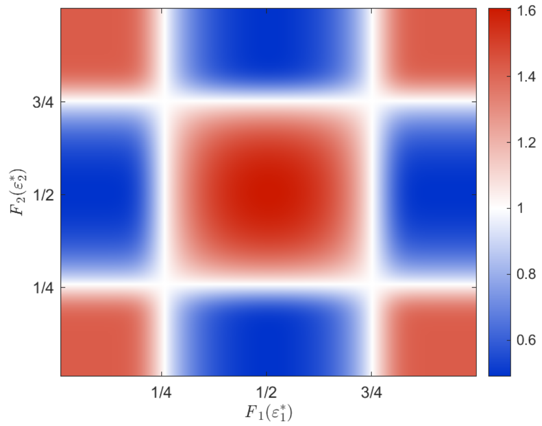


Figure 2c: Mixture of normals (DGP 2)

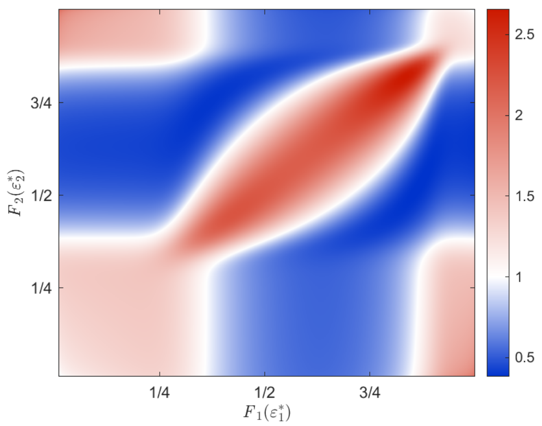
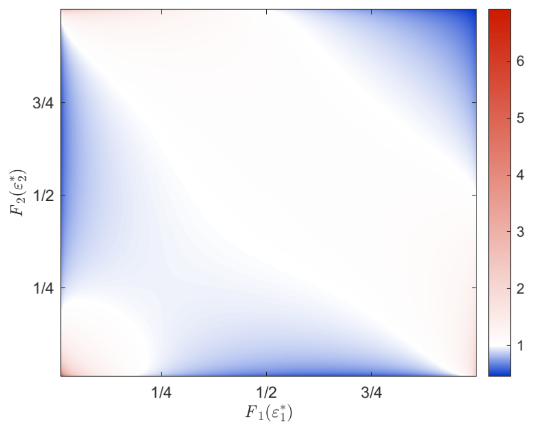


Figure 2d: Asymmetric t (DGP 3)



Notes: Details on the copula densities: DGP 0: ε_{1t}^* follows a Student t with 10 degrees of freedom (and kurtosis coefficient equal to 4), and ε_{2t}^* is generated as an asymmetric t with kurtosis and skewness coefficients equal to 4 and $-.5$, respectively; DGP 1: Standardised scale mixture of two zero mean normals –with scalar covariance matrix– in which the higher variance component has probability $\lambda = 0.2$ and the ratio of the variances is $\varkappa = 0.05$; DGP 2: Multivariate discrete mixture of two normals with mixing probability $\lambda = 0.7$, relative-means difference $\boldsymbol{\delta}_2 = (0.5,\text{-}0.5)'$ and relative-covariance difference such that $\aleph_2$ is lower triangular with $vech(\aleph_2) = 0.2\boldsymbol{\ell}_2$ (see Appendix D in Amengual, Fiorentini and Sentana (2023) for details); and DGP 3: Asymmetric Student t with skewness vector $\boldsymbol{\beta} = -10\boldsymbol{\ell}_2$ and degrees of freedom parameter $\nu = 12$ (see Mencía and Sentana (2012) for details).

Figure 3: Data and structural shocks

Figure 3a: Volatility index series (logs)

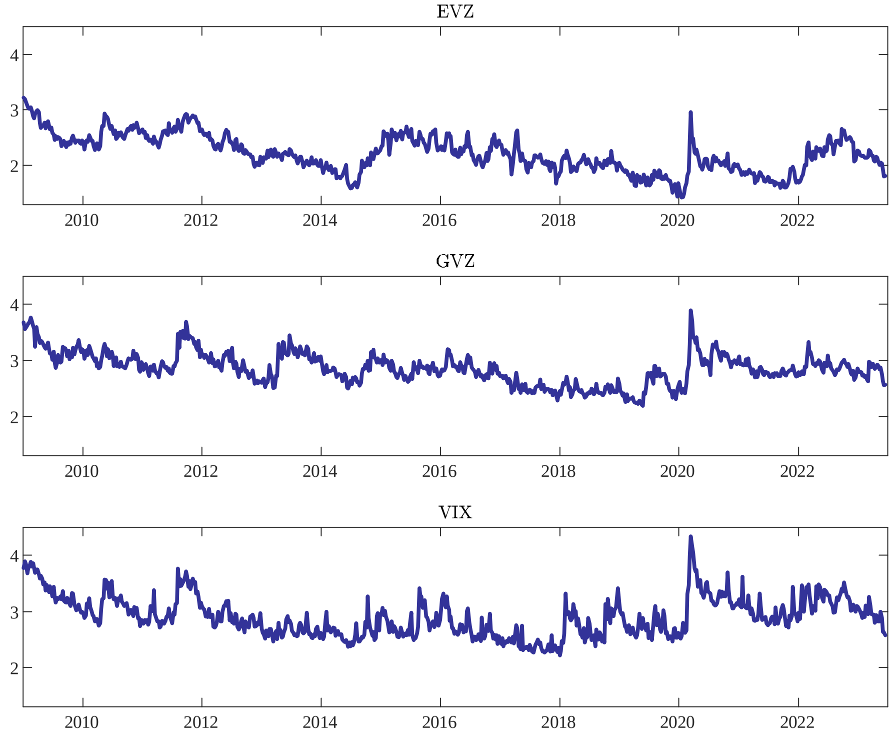
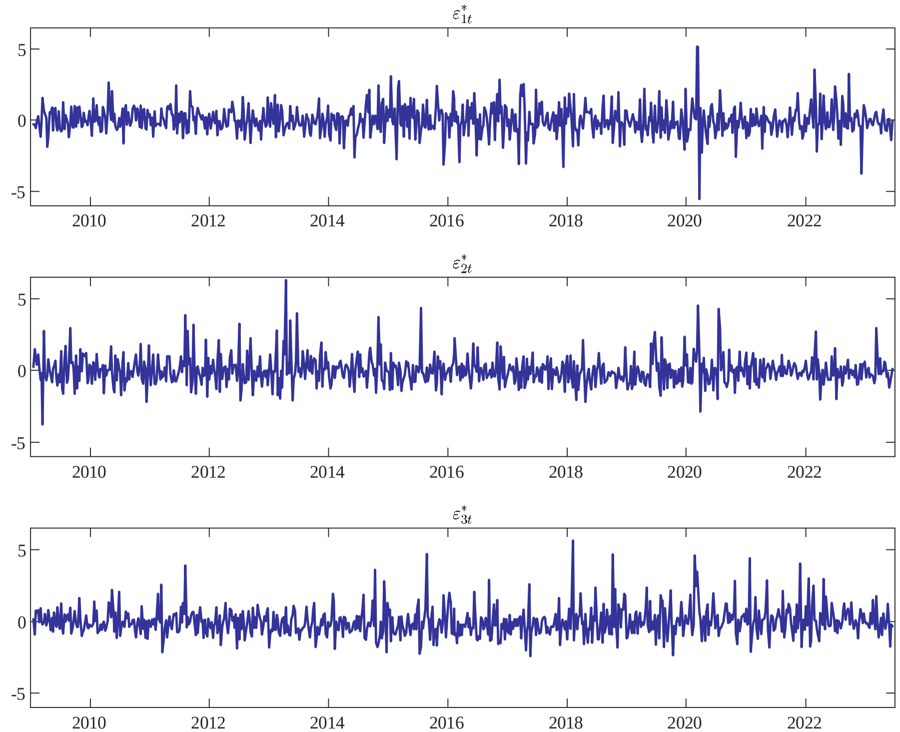
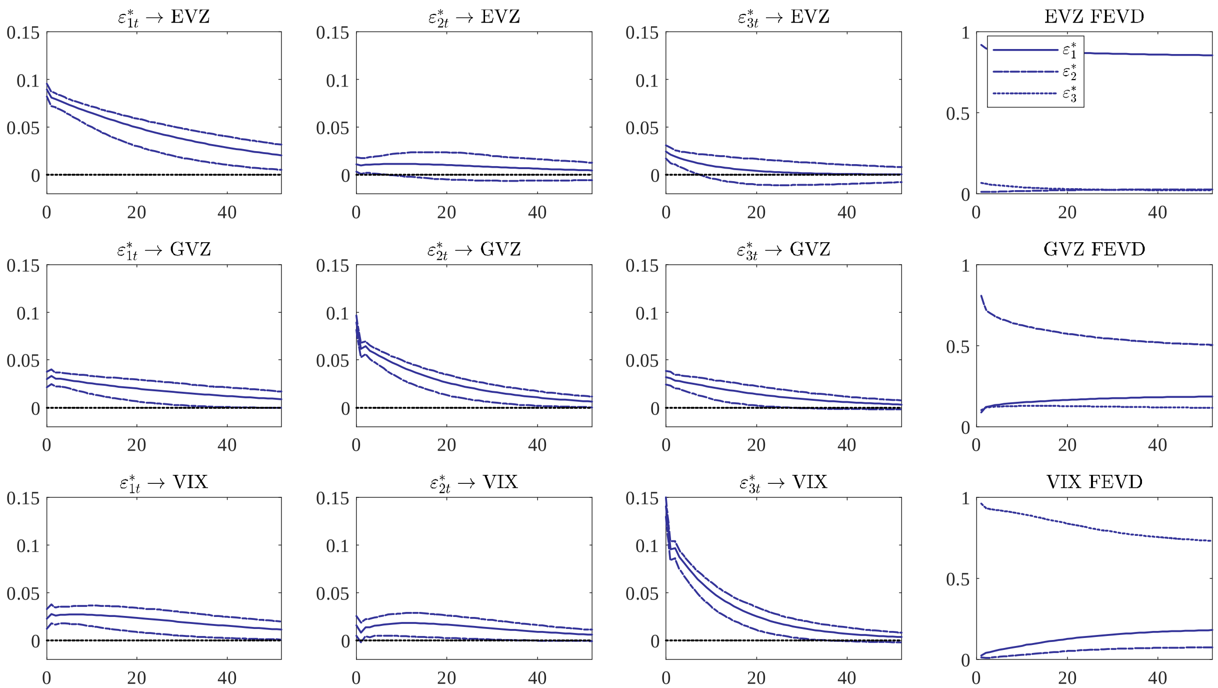


Figure 3b: Non-Gaussian structural shocks



Notes: Figure 3a: Three weekly series of market-based implied volatilities: the VIX index, the EVZ EuroCurrency ETF volatility index, and the GVZ Gold ETF volatility index, which represent three of the most actively traded asset classes: stocks, exchange rates and commodities. Sample spans 2009/01/21 to 2023/06/21 (753 Wednesday observations). Figure 3b: Standardised structural shocks based on the trivariate SVAR model $\mathbf{x}_t = \boldsymbol{\tau} + \mathbf{A}_1\mathbf{x}_{t-1} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \mathbf{C}\boldsymbol{\varepsilon}^*_t$ in which we selected $p = 2$ by looking at the Akaike information criterion and the likelihood ratio test for lack of residual serial correlation. We estimate $(\boldsymbol{\tau}, \mathbf{a}, \mathbf{c})$ jointly with the shape parameters by PML assuming that $\varepsilon^*_{it} \sim i.i.d.\ DMN(\delta_i, \kappa_i, \lambda_i)$.

Figure 4: Impulse response functions and FEVDs



Notes: Results based on a trivariate SVAR model $\mathbf{x}_t = \boldsymbol{\tau} + \mathbf{A}_1\mathbf{x}_{t-1} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \mathbf{C}\boldsymbol{\varepsilon}^*_t$, in which we selected $p = 2$ by looking at the Akaike information criterion and the likelihood ratio test for lack of residual serial correlation. We estimate $(\boldsymbol{\tau}, \mathbf{a}, \mathbf{c})$ jointly with the shape parameters by PML assuming that $\varepsilon^*_{it} \sim i.i.d.\ DMN(\delta_i, \kappa_i, \lambda_i)$.

Supplemental Appendices for

Specification tests for non-Gaussian structural vector autoregressions

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

Gabriele Fiorentini
Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy
<gabriele.fiorentini@unifi.it>

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

December 2022
Revised: March 2024

A The test in practice

We recommend following these steps for computing the discrete grid test statistics in a given sample:

1. Estimate the model by non-Gaussian PMLE assuming that the shocks follow independent univariate finite Gaussian mixtures, and compute the estimated structural residuals $\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}_T)$'s evaluated at the PMLEs $\hat{\boldsymbol{\theta}}_T$ using expression (4) for the unique ordering and signs of the matrix $\mathbf{C}$ obtained using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne et al. (2017). Importantly, the fact the structural shocks are only identified up to permutations is numerically irrelevant for the test statistic as long as one uses the same quantile grid for all of them since they only affect their labelling. Similarly, a change in the sign of one shock is also numerically irrelevant as long as one adjusts its quantiles accordingly. In fact, there is no need for such an adjustment if one uses equally spaced quantiles (say terciles, quartiles or quintiles) for all shocks.
2. For the q version of the test, partition the $[0, 1]$ interval with knots $(0, u_1, u_2, ..., u_H, 1)$, where $u_h = \frac{1}{2}(2h - 1) H^{-1}$ in the equally-spaced case, and obtain the corresponding marginal quantiles of each estimated shock $\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}_T)$, namely $[k_{i1}(u_1), \dots, k_{iH}(u_H)]$, $i = 1, \dots, N$ using Matlab's linear interpolation method. One could then replace the u's with the marginal empirical cdf of each shock computed at the estimated quantiles to take into account the linear interpolation method, but this would generate slightly different partitions of the unit interval for different shocks.

   For the p version of the tests, define H points, $k_1 < \cdots < k_h < \cdots < k_H$, together with $k_0 = -\infty$ and $k_{H+1} = \infty$, and estimate the marginal empirical cdf for each shock as $p_{ih} = T^{-1} \sum_{t=1}^{T} I[\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}_T) \le k_h]$. One could then replace k_h with its marginal empirical quantile at the estimated p_{ih} for each shock using Matlab's linear interpolation method, but again this would generate slightly different partitions of the real line for different shocks.
3. For the q version, estimate the joint cdf at the Cartesian product of the empirical quantiles as $q_{ij} = T^{-1} \sum_{t=1}^{T} I[\varepsilon_{1t}^*(\hat{\boldsymbol{\theta}}_T) \le k_{ih}(u_h); \dots, \varepsilon_{Mt}^*(\hat{\boldsymbol{\theta}}_T) \le k_{Mh'}(u_{h'})]$, while for the p version do the same but evaluate them at the N-ary Cartesian product of $(k_1, \dots, k_H)'$.
4. Compute the H^N influence functions underlying the test as the difference between the joint and the product of the marginal empirical cdfs.
5. Compute the $H^N \times H^N$ matrix whose elements are given by (13).
6. Estimate the asymptotic covariance matrix of the score and the expected Hessian of the pseudo log-likelihood function replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions that appear in Appendices C.3 and C.4, respectively, including (C33)-(C38) and (C39)-(C44), and use them in the

sandwich formula $\mathcal{A}^{-1}\mathcal{B}\mathcal{A}$, retaining the $N \times N$ blocks corresponding to the elements of $vec(\mathbf{C})$. The consistency of the estimators of $\mathcal{A}$ and $\mathcal{B}$ follows from Lemma 4.3 in Newey and McFadden (1994), while that of $\mathcal{A}^{-1}\mathcal{B}\mathcal{A}$ from their Theorem 4.1.

7. Estimate the $H^N \times N$ expected Jacobian matrix of the influence functions with respect to the elements of $vec(\mathbf{C})$ replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions in Lemma 1, using Silverman's (1986) robust rule-of-thumb bandwidth to obtain Gaussian kernel estimates of the true density of the shocks that appear in expression (B1). Despite involving the indicator function, the consistency of this procedure follows once again from Lemma 4.3 in Newey and McFadden (1994).
8. Estimate the $H^N \times N$ asymptotic covariance matrix between the influence functions and the scores with respect to the elements of $vec(\mathbf{C})$ replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions that appear in Lemma 2, including (B3)-(B9). As before, Lemma 4.3 in Newey and McFadden (1994) guarantees the consistency of the resulting estimators despite the indicator function appearing in the influence functions.
9. Combine these matrices to estimate $\mathcal{W}$ using (15), and replace this estimated matrix in (14) to obtain the discrete grid test statistic. Theorems 2.2 and 2.3 in Newey (1985) guarantee the consistent estimation of $\mathcal{W}$ and the asymptotic χ^2 distribution of (14), respectively.

Given that the continuous grid test can be regarded as a regularised version of the discrete grid test computed at the finest partition of the unit interval that remains meaningful when there are T observations, its computation shares several of the elements that we have just described. Specifically:

1. Estimate the model by non-Gaussian PMLE assuming that the shocks follow independent univariate finite Gaussian mixtures, and compute the estimated structural residuals $\varepsilon^*_{it}(\hat{\boldsymbol{\theta}}_T)$'s evaluated at the PMLEs $\hat{\boldsymbol{\theta}}_T$ using expression (4) for the unique ordering and signs of the matrix $\mathbf{C}$ obtained using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne et al. (2017). The fact the structural shocks are only identified up to permutations and sign changes is numerically irrelevant for the continuous test statistic as it effectively depends on the homogeneous, equally-spaced "discrete" grid $u_\tau = \frac{1}{2}(2\tau - 1)T^{-1}$, $\tau = 1, \ldots, T$.
2. Compute the empirical uniform ranks using expression (17) and use them to obtain the elements of the $T \times T$ matrix $\mathcal{D}$ in (25).
3. Estimate the $T \times T$ matrix $\mathcal{C}$ by replacing the integrals in (27) by sums over the empirical cdfs of the shocks. Specifically, if we denote by $\boldsymbol{\epsilon}^*_t(\hat{\boldsymbol{\theta}}) = [\epsilon^*_{it}(\hat{\boldsymbol{\theta}}), ..., \epsilon^*_{\mathrm{Mt}}(\hat{\boldsymbol{\theta}})]$ the vector containing the empirical ranks of the t^{th} observation of each of the estimated shocks that

appear in M, we can estimate the rank one matrix $\mathcal{C}$ as

$$\widehat{\mathcal{C}} = \boldsymbol{\ell}_T\boldsymbol{\ell}_T' \cdot \sum_{\tau^1=1}^{T} \cdots \sum_{\tau^{\mathrm{M}}=1}^{T} \mathbf{j}'[\epsilon^*_{\tau^1}(\hat{\boldsymbol{\theta}}), ..., \epsilon^*_{\tau^{\mathrm{M}}}(\hat{\boldsymbol{\theta}})]\mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})\mathcal{B}(\hat{\boldsymbol{\theta}})\mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{j}[\epsilon^*_{\tau^1}(\hat{\boldsymbol{\theta}}), ..., \epsilon^*_{\tau^{\mathrm{M}}}(\hat{\boldsymbol{\theta}})],$$

where $\boldsymbol{\ell}_T$ is a vector of T ones, while $\mathcal{A}(\hat{\boldsymbol{\theta}})$, $\mathcal{B}(\hat{\boldsymbol{\theta}})$ and $\mathbf{j}[\epsilon^*_{\tau^1}(\hat{\boldsymbol{\theta}}), ..., \epsilon^*_{\tau^{\mathrm{M}}}(\hat{\boldsymbol{\theta}})]$ are the consistent estimators that we mention in points 6 and 7 of the description of the discrete grid test, with the latter evaluated at $u_\tau = \frac{1}{2}(2\tau - 1)T^{-1-1}$, $\tau = 1, \ldots, T$. Given that sums over increasingly finer grids converge to the relevant integral, $\widehat{\mathcal{C}}$ will be consistent.

4. Finally, we consistently estimate $\mathcal{E}$ by adding up the consistent estimators of $\mathcal{C}$ and $\mathcal{D}$, which we then replace in expression (26) for a given choice of the regularization parameter α. Interestingly, the fact that $\widehat{\mathcal{C}}$ is proportional to $\boldsymbol{\ell}_T\boldsymbol{\ell}_T'$ implies that the expression (26) is numerically unaffected if we replace the two $\mathcal{E}s$ that appear at the extremes of this quadratic form with $\mathcal{D}s$.

B Lemmata

Lemma 1 *If model (2) satisfies Assumption 1, then the non-zero elements of the expected Jacobian matrix of the linearised $m_t(\mathbf{u})$ evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u_h^i in (10) are given by*

$$\mathrm{J}_{p_{\mathbf{h}}c_{ii'}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa(u_i)], \quad \text{for } \; i \neq i', \tag{B1}$$

where $\eta_{h_{i'}} = E_0[\varepsilon^*_{it}1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})]$ for $i \in M$.

Proof. From (12), we have that

$$\begin{aligned}
\frac{\partial m_t(\mathbf{u})}{\partial \theta} &= E\left[\frac{\partial}{\partial\theta}\left[\prod_{i\in M} 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it}) - \prod_{i\in M} u_i\right]\right. \\
&\quad \left. -\frac{\partial}{\partial\theta}\left\{\sum_{i\in M}\left[1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it}) - u_i\right]\prod_{i'\in M, i'\neq i} u_{i'}\right\}\right] \\
&= -\sum_{i\in M}\left[\prod_{i'\in M, i'\neq i} 1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\right]\left[1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it}) - u_i\right]\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial\varepsilon^*_{it}}\frac{\partial\varepsilon^*_{it}}{\partial\theta}.
\end{aligned}$$

Moreover, it is worth noticing that

$$\begin{aligned}
\frac{\partial\varepsilon^*_{it}(\boldsymbol{\theta})}{\partial\boldsymbol{\tau}'} &= -\mathbf{c}^{i\cdot}, \\
\frac{\partial\varepsilon^*_{it}(\boldsymbol{\theta})}{\partial\mathbf{a}_j'} &= -(\mathbf{y}_{t-j}' \otimes \mathbf{c}^{i\cdot}) \quad \text{for } j = 1, ..., p, \text{ and} \\
\frac{\partial\varepsilon^*_{it}(\boldsymbol{\theta})}{\partial\mathbf{c}'} &= -[\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}) \otimes \mathbf{c}^{i\cdot}].
\end{aligned} \tag{B2}$$

Therefore, under the independence null,

$$E\left[\frac{\partial m_t(\mathbf{u})}{\partial \theta_i}\right]=0$$

except for the off-diagonal elements of $\mathbf{C}$, namely,

$$\begin{aligned}
&E\left\{\sum_{i\in M}\left(\sum_{i'\in M,i'\neq i}1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\right)\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}\frac{\partial \varepsilon^*_{it}}{\partial \mathbf{c}'}\right\}\\
=&\ -E\left\{\sum_{i\in M}\left(\sum_{i'=1,i'\neq i}1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\right)\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}[\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\otimes \mathbf{c}^{i\cdot}]\right\}\\
=&\ -\sum_{i\in M}\sum_{i'\in M,i'\neq i}E\left(\prod_{i''\in M,i''\neq i'\neq i}1_{(-\infty,\varkappa(u_{i''}))}(\varepsilon^*_{i''t})\right)\\
&\times E[1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\varepsilon^*_{i't}]E\left[\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}\right](\mathbf{e}'_j\otimes \mathbf{c}^{i\cdot})\\
=&\ -\sum_{i\in M}\sum_{i'\in M,i'\neq i}\left(\prod_{i''\in M,i''\neq i'\neq i}u_{i''}\right)\eta_{u_{i'}}f[\varkappa(u_i)]
\end{aligned}$$

where the first equality uses (B2), the second one follows from the cross-sectional independence of the shocks, and the last one implicitly defines $\eta_{u_j}=E[\varepsilon^*_{jt}1_{(-\infty,\varkappa(u_j))}(\varepsilon^*_{jt})]$. □

Lemma 2 *If model (2) satisfies Assumption 1, then the non-zero elements of the covariance matrix between the linearised influence function* $m_t(\mathbf{u})$ *evaluated at* $\boldsymbol{\theta}_0$ *and the estimated values of* u^i_h *in (10) with the pseudo log-likelihood scores evaluated at the pseudo true values* $\boldsymbol{\phi}_\infty$ *is given by*

$$cov[m_t(\mathbf{u}),\mathbf{s}_{c_{ii'}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{v}_0]=E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty,\boldsymbol{v}_0)],$$

where

$$\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty,\boldsymbol{v}_0)=\begin{bmatrix}\mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{bmatrix}\begin{bmatrix}\mathbf{0}\\ \mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)\\ \mathbf{0}\end{bmatrix},$$

where $\mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ *is a* $N^2\times 1$ *vector whose entries* $s=N(i-1)+i'$ *for* $i,i'=1,...,N$ *are*

$$\mathrm{K}_{m,s}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)=-\sum_{i\in M}\sum_{i'\in M}\left(\prod_{i''\in M,i''\neq i,i''\neq i'}u_{i''}\right)\eta_{i'}E\left\{1_{(\varepsilon^*_{it}\leq\varkappa(u_i))}\cdot\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i}\bigg|\boldsymbol{\theta}_0,\boldsymbol{v}_0\right\},$$

for $i\neq i'$*, and zero otherwise.*

Proof. We start by computing the covariance of the influence functions underlying our test with the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$, namely

$$cov[m_t(\mathbf{u}),\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{v}_0]=\mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\phi}_\infty,\boldsymbol{v}_0)=E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty,\boldsymbol{v}_0)]$$

and

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = \mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0)],$$

where

$$\mathcal{K}_{\cdot t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathcal{K}_{\cdot lt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot st}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot rt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \end{bmatrix}.$$

Exploiting the cross-sectional independence of the shocks, we get for the mean parameters

$$\begin{aligned} \mathrm{K}_{p_k l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \le \varkappa(u_i))} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \qquad \text{(B3)} \\ \mathrm{K}_{p_\mathbf{k} l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial {\varepsilon_i}^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i' \in M, i' \neq i} u_{i'} \right) E\left\{ 1_{(\varepsilon_{it}^* \le \varkappa(u_i))} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \qquad \text{(B4)} \end{aligned}$$

and zero otherwise.

Similarly, $\mathcal{K}_{\cdot s}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ is a $N^2 \times 1$ vector whose entries are such that for i with $j_i > 0$,

$$\begin{aligned} \mathrm{K}_{p_k s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p(\boldsymbol{\varepsilon}_t^*; \mathbf{k}, \mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \ge k_{h_i})} \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \qquad \text{(B5)} \\ \mathrm{K}_{p_\mathbf{k} s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i' \in M, i' \neq i} \pi_{i'} \right) E\left\{ \varepsilon_{it}^* 1_{(\varepsilon_{it}^* \le k_{h_i})} \cdot \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \qquad \text{(B6)} \\ \mathrm{K}_{p_\mathbf{k} s_2}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{i't}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i'' \in M, i'' \neq i, i'' \neq i'} u_{i''} \right) \eta_{i'} E\left\{ 1_{(\varepsilon_{it}^* \le k_{h_i})} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \qquad \text{(B7)} \end{aligned}$$

and zero otherwise.

Finally, $\mathcal{K}_{kr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = \mathrm{K}'_{kr} vecd(\mathbf{I}_n)$, where K_{kr} another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$,

$$\begin{aligned} \mathrm{K}_{p_k r}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= E\left\{ 1_{(\varepsilon_{it}^* \le \varkappa(u_i))} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \qquad \text{(B8)} \end{aligned}$$

$$\mathrm{K}_{p_{\mathbf{k}}r}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) = cov\left\{ m_t(\mathbf{u}), \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\}$$

$$= \left(\prod_{i'\in M, i'\neq i} u_{i'}\right) E\left\{ 1_{(\varepsilon_{it}^*\leq k_{h_i})}\cdot \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\} \quad \text{(B9)}$$

and zero otherwise, again because of the cross-sectional independence of the shocks and the fact that $E[\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_\infty)/\partial \varepsilon_i^*|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = 0$.

Next, to obtain the covariance of the influence function evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u_h^i in (10) with the pseudo log-likelihood scores evaluated at the true values $\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0$, we can make use of (12) to write

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = cov\{p(\boldsymbol{\varepsilon}_t^*;\mathbf{k},\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\} \quad \text{(B10)}$$

$$-\sum_{i\in M}\left(\prod_{i'\in M, i'\neq i} u_{i'}\right) cov\left\{p_{k_i}(\varepsilon_{it}^*), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\}.$$

Then, substituting (B3) and (B4) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}$$

and

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\mathbf{a}_j t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}, \text{ for } j=1,...,p.$$

Similarly, substituting (B5) and (B6) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{c_{ii}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = 0, \text{ for } i=1,...,N;$$

and substituting (B8) and (B9) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\varrho}_i t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}, \text{ for } i=1,...,N.$$

Finally, substituting (B5) and (B7) into (B10), we get the result stated in the statement. □

Lemma 3 *If model (2) satisfies Assumption 1, then the adjustment of the covariance operator that accounts for the estimation of* $\boldsymbol{\theta}$ *is given by (27).*

Proof. From 1, the expected Jacobian with respect to $\boldsymbol{\theta}$ of the influence functions linearised with respect to the $\varkappa$'s can be written as

$$E\left[\frac{\partial n_t(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] = -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa(u_i)](\mathbf{e}_{i'}'\otimes \mathbf{c}^{i\cdot}),$$

where

$$n_t(\mathbf{u}_M) = \left[\prod_{i\in M} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - \prod_{i\in M} u_i\right] - \sum_{i\in M}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in M, i'\neq i} u_{i'}.$$

We are after

$$\int_{[0,1]^{\mathrm{M}}}\left\{n_t(\mathbf{u}_M)-E\left[\frac{\partial n_t(\mathbf{u}_M)}{\partial\boldsymbol{\theta}'}\right]\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right\}\left\{n_s(\mathbf{u}_M)-E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial\boldsymbol{\theta}'}\right]\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right\}d\mathbf{u}_M.$$

Let us consider each of the four terms separately. The first one, namely

$$\int_{[0,1]^{\mathrm{M}}} n_t(\mathbf{u}_M)n_s(\mathbf{u}_M)d\mathbf{u}_M,$$

is given in (13). Next, we have the cross-terms, which are of the form

$$-\int_{[0,1]^{\mathrm{M}}} E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial\boldsymbol{\theta}'}\right]\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)n_t(\mathbf{u}_M)d\mathbf{u}_M.$$

If we then use the fact that

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

we can see that

$$-\frac{1}{\sqrt{T}}\int_{[0,1]^{\mathrm{M}}} E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial\boldsymbol{\theta}'}\right]\left(\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\sum_{\tau=1}^{T}\mathbf{s}_{\boldsymbol{\theta}\tau}\right)n_t(\mathbf{u}_M)d\mathbf{u}_M=o_p(1)$$

because of the scaling factor $1/\sqrt{T}$ and the fact that the ε's entering into $\mathbf{s}_{\boldsymbol{\theta}\tau}(\boldsymbol{\phi})$ are asymptotically independent of the ones that appear in $n_t(\mathbf{u}_M)$ and $E\left[\partial n_s(\mathbf{u}_M)/\partial\boldsymbol{\theta}'\right]$. Therefore, the covariance of the linearised influence function with the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$ is asymptotically negligible.

Finally, regarding the last term, we obtain (27), as desired. □

C ML estimators with cross-sectionally independent shocks

In this appendix, we derive analytical expressions for the conditional variance of the score and the expected value of the Hessian of SVAR models with cross-sectionally independent non-Gaussian shocks when the distributions assumed for estimation purposes may well be misspecified, but all the parameters that characterise the conditional mean and covariance functions are consistently estimated, as in the case of finite normal mixtures. Fiorentini and Sentana (2023) consider the general case.

C.1 Log-likelihood, its score and Hessian

Given the linear mapping between structural shocks and reduced form innovations, the contribution to the conditional log-likelihood function from observation t will be given by

$$l_t(\mathbf{y}_t;\boldsymbol{\varphi})=-\ln|\mathbf{C}|+l[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]+\ldots+l[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N],\tag{C11}$$

where $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \mathbf{C}^{-1}(\mathbf{y}_t - \boldsymbol{\tau} - \mathbf{A}_1\mathbf{y}_{t-1} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p})$ and $l(\varepsilon_{it}^*; \boldsymbol{\varrho}_i) = \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_i)$ is the log of the univariate density function of ε_{it}^*, which we assume twice continuously differentiable with respect to both its arguments, although this is stronger than necessary, as the Laplace example illustrates.

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. Given that the mean vector and covariance matrix of (2) conditional on I_{t-1} are

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, \tag{C12a}$$

$$\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{C}\mathbf{C}', \tag{C12b}$$

respectively, we can use the expressions in Supplemental Appendix D.1 of Fiorentini and Sentana (2021) with $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}) = \mathbf{C}$ to show that

$$\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\frac{\partial vec'(\mathbf{C})}{\partial\boldsymbol{\theta}} vec(\mathbf{C}^{-1\prime}) = -\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} vec(\mathbf{C}^{-1\prime}) = -\mathbf{Z}_{st}'(\boldsymbol{\theta}) vec(\mathbf{I}_N) \tag{C13}$$

and

$$\begin{aligned} \frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\mathbf{C}^{-1}\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{C}^{-1}]\frac{\partial vec(\mathbf{C})}{\partial\boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}, \end{aligned} \tag{C14}$$

where

$$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix} \mathbf{C}^{-1\prime}, \tag{C15}$$

$$\mathbf{Z}_{st}(\boldsymbol{\theta}) = \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N^2\times N^2} \\ \vdots \\ \mathbf{0}_{N^2\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}), \tag{C16}$$

which confirms that the conditional mean and variance parameters are variation free. In addition,

$$\begin{aligned} \mathbf{s}_t(\boldsymbol{\phi}) &= \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_t(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\phi}), \end{aligned} \tag{C17}$$

where

$$\mathbf{e}_{lt}(\boldsymbol{\phi}) = -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} = -\begin{bmatrix} \partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]/\partial \varepsilon_1^* \\ \partial \ln f_2[\varepsilon_{2t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_2]/\partial \varepsilon_2^* \\ \vdots \\ \partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]/\partial \varepsilon_N^* \end{bmatrix}, \tag{C18}$$

$$\begin{aligned} \mathbf{e}_{st}(\boldsymbol{\phi}) &= -vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \\ &= -vec\left\{\begin{array}{ccc} 1 + \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & 1 + \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array}\right\} \end{aligned} \tag{C19}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi}) = \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}} = \left\{\begin{array}{c} \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array}\right\} = \begin{bmatrix} \mathbf{e}_{r_1t}(\boldsymbol{\phi}) \\ \mathbf{e}_{r_2t}(\boldsymbol{\phi}) \\ \vdots \\ \mathbf{e}_{r_Nt}(\boldsymbol{\phi}) \end{bmatrix} \tag{C20}$$

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

Let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial \mathbf{s}_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'$. Supplemental Appendix D.1 of Fiorentini and Sentana (2021) implies that

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} \\ &+ \left[\mathbf{e}_{lt}'(\boldsymbol{\phi}) \otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi}) \otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, \end{aligned} \tag{C21}$$

where $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ and $\mathbf{Z}_{st}(\boldsymbol{\theta})$ are given in (C15) and (C16), respectively. Therefore, we need to obtain $\partial vec(\mathbf{C}^{-1\prime})/\partial \boldsymbol{\theta}'$ and $\partial vec(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})/\partial \boldsymbol{\theta}'$.

Let us start with the former. Given that

$$dvec(\mathbf{C}^{-1\prime}) = -vec[\mathbf{C}^{-1\prime}d(\mathbf{C}')\mathbf{C}^{-1\prime}] = -(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})dvec(\mathbf{C}') = -(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN}dvec(\mathbf{C}),$$

where $\mathbf{K}_{NN}$ is the commutation matrix (see Magnus and Neudecker (2019)), we immediately get that

$$\frac{\partial vec(\mathbf{C}^{-1\prime})}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \mathbf{0}_{N^2\times(N+pN^2)} & -(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix},$$

so that

$$\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} = \left[\mathbf{I}_N \otimes \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix}\right]\frac{\partial vec(\mathbf{C}^{-1\prime})}{\partial \boldsymbol{\theta}'}$$

$$= \left[\mathbf{I}_N \otimes \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix} \right] \begin{bmatrix} \mathbf{0}_{N^2 \times (N+pN^2)} & (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix}.$$

Similarly, given that

$$vec(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}) = \{[(\mathbf{I}_N \otimes \mathbf{K}_{NN})(vec(\mathbf{I}_N) \otimes \mathbf{I}_N)] \otimes \mathbf{I}_N\} vec(\mathbf{C}^{-1\prime})$$

so that

$$\begin{aligned} vec(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}) &= ((\mathbf{I}_N \otimes \mathbf{K}_{NN})(vec(\mathbf{I}_N) \otimes \mathbf{I}_N) \otimes \mathbf{I}_N) dvec(\mathbf{C}^{-1\prime}) \\ &= -\{[(\mathbf{I}_N \otimes \mathbf{K}_{NN})(vec(\mathbf{I}_N) \otimes \mathbf{I}_N)] \otimes \mathbf{I}_N\}(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} dvec(\mathbf{C}), \end{aligned}$$

we will have that

$$\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} = \frac{\partial vec}{\partial \boldsymbol{\theta}'} \left[\begin{pmatrix} \mathbf{0}_{(N+pN^2)\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}) \right].$$

But

$$\begin{aligned} &\left[\mathbf{I}_{N^2} \otimes \begin{pmatrix} \mathbf{0}_{(N+pN^2)\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} \right] \frac{\partial vec(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})}{\partial \boldsymbol{\theta}'} \\ = \; & -\left[\mathbf{I}_{N^2} \otimes \begin{pmatrix} \mathbf{0}_{(N+pN^2)\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} \right] [\, \mathbf{0} \quad \{[(\mathbf{I}_N \otimes \mathbf{K}_{NN})(vec(\mathbf{I}_N) \otimes \mathbf{I}_N)] \otimes \mathbf{I}_N\}(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \,]. \end{aligned}$$

In addition,

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\} \quad \text{(C22)}$$

and

$$\begin{aligned} \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{ \mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= \left\{ [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right] \right\} \\ &\quad \times \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\}. \qquad \text{(C23)} \end{aligned}$$

The assumed independence across innovations implies that

$$\frac{\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{(\partial \varepsilon_1^*)^2} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{(\partial \varepsilon_N^*)^2} \end{bmatrix}, \qquad \text{(C24)}$$

which substantially simplifies the above expressions.

Moreover,

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'},$$

where

$$\begin{aligned}\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'} &= -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}, \\ \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}.\end{aligned}$$

with

$$\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^* \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^* \partial \boldsymbol{\varrho}_N'} \end{bmatrix} \tag{C25}$$

because of the cross-sectional independence assumption.

As for the shape parameters of the independent margins,

$$\mathbf{h}_{\boldsymbol{\varrho}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1 \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N \partial \boldsymbol{\varrho}_N'} \end{bmatrix}. \tag{C26}$$

Finally, regarding the Jacobian term $-\ln|\mathbf{C}|$, we have that differentiating (C13) once more yields

$$-\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} dvec(\mathbf{C}^{-1\prime}) = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} dvec(\mathbf{C}),$$

so

$$\frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} \begin{bmatrix} \mathbf{0}_{N^2 \times (N+pN^2)} & (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix}.$$

As usual, the pseudo true values of the parameters of a globally identified model, $\boldsymbol{\phi}_\infty$, are the unique values that maximise the expected value of the log-likelihood function over the admissible parameter space, which is a compact subset of $\mathbb{R}^{\dim(\boldsymbol{\phi})}$, where the expectation is taken with respect to the true distribution of the shocks. Under standard regularity conditions (see e.g., White (1982)), those pseudo true values will coincide with the values of the parameters that set to 0 the expected value of the pseudo-log likelihood score.

More formally, if we define $\boldsymbol{\upsilon}_0$ as the true values of the shape parameters, and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0)$, we would normally expect that

$$E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathbf{0}.$$

Let us now consider the alternative parametrisation $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$ studied in Fiorentini and Sentana (2021, 2023), so that the parameters of interest become $\boldsymbol{\tau}$, $\mathbf{a}_j = vec(\mathbf{A}_j)$ $(j = 1, \ldots, p)$, $\mathbf{j} = veco(\mathbf{J})$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$, where $veco(.)$ stacks by columns all the elements of the zero-diagonal matrix $\mathbf{J} - \mathbf{I}_N$ except those that appear in its diagonal, and $vecd(.)$ places the elements in the main diagonal of $\boldsymbol{\Psi}$ in a column vector (see Magnus and Sentana (2020) for some useful properties of these operators). Given that a pseudo log-likelihood function based on finite Gaussian mixtures for the shocks will lead to consistent estimators for all these parameters regardless of the true distribution, $\mathbf{e}_t(\boldsymbol{\phi}_\infty)$ will be serially independent and not just martingale difference sequences. Moreover, given that

$$\mathbf{Z}(\boldsymbol{\theta}) = E[\mathbf{Z}_t(\boldsymbol{\theta})|\boldsymbol{\varphi}_0] = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times q} \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{q\times N} & \mathbf{0}_{q\times N^2} & \mathbf{I}_q \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_d(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \quad \text{(C27)}$$

has full column rank,

$$E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0] = \mathbf{0} \quad \text{(C28)}$$

because

$$\mathbf{0} = E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = E\{E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0]|\boldsymbol{\varphi}_0\} = \mathbf{Z}(\boldsymbol{\theta})E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0] = \mathbf{Z}(\boldsymbol{\theta})E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0].$$

Furthermore, the diagonality of $\boldsymbol{\Psi}$ means that the pseudo-shocks $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)$ will also inherit the cross-sectional independence of the true shocks $\boldsymbol{\varepsilon}_t^*$. In addition, given that the estimators of $\boldsymbol{\theta}$ that we consider are consistent, we will have that under standard regularity conditions

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 0 \text{ and} \quad \text{(C29)}$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^{*2}(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^{*^2}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 1, \quad \text{(C30)}$$

where $\hat{\boldsymbol{\theta}}$ are the PMLEs of the conditional mean and variance parameters.

C.2 Asymptotic distribution

For simplicity, we assume henceforth that there are no unit roots in the autoregressive polynomial, so that the SVAR model (2) generates a covariance stationary process in which $rank(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p) = N$. If the autoregressive polynomial $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)$ had some unit roots, then $\mathbf{y}_t$ would be a (co-) integrated process, and the estimators of the conditional mean parameters would have non-standard asymptotic distributions, as some (linear combinations) 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)).

We also assume that the regularity conditions A1-A6 in White (1982) are satisfied, although like in his Theorems 3.1 and 3.2, we drop Assumption A3(b) when talking about the negative definiteness of the expected Hessian or the asymptotic normality of the PML estimators because they are both local rather than global results. These conditions are only slightly stronger than those in Crowder (1976), which guarantee that MLEs will be consistent and asymptotically normally distributed under correct specification. In particular, Crowder (1976) requires: (i) $\boldsymbol{\phi}_0$ is locally identified and belongs to the interior of the admissible parameter space, which is a compact subset of $\mathbb{R}^{\dim(\phi)}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\phi}_0$; (iii) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\boldsymbol{\phi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right]T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi}) \xrightarrow{p} \mathbf{I}_{p+q}$, where $E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$.

We can use the law of iterated expectations to compute

$$\mathcal{A}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = E[-\mathbf{h}_{\phi\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0] = E\left[\mathcal{A}_t(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0)\right]$$

and

$$V[\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathcal{B}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = E\left[\mathcal{B}_t(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi})\right].$$

In this context, the asymptotic distribution of the PMLEs of $\boldsymbol{\phi}$ under the regularity conditions A1-A6 in White (1982) will be given by

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

As we explained before, analogous expressions apply *mutatis mutandi* to a restricted PML estimator of $\boldsymbol{\theta}$ that fixes $\boldsymbol{\varrho}$ some a priori chosen value to $\bar{\boldsymbol{\varrho}}$. In that case, we would simply need to replace $\boldsymbol{\theta}_\infty$ by $\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\varrho}})$ and eliminate the rows and columns corresponding to the shape parameters $\boldsymbol{\varrho}$ from the $\mathcal{A}$ and $\mathcal{B}$ matrices.

If we write $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, then the chain rule for first derivatives implies that the gradient with respect to the parameters in $\mathbf{C}$ will be a linear combination of those corresponding to $\mathbf{j} = veco(\mathbf{J} - \mathbf{I}_N)$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$.

Therefore, we can invoke Proposition 3 in Fiorentini and Sentana (2023), which shows the consistency of the Gaussian mixture-based Pseudo MLEs of $\mathbf{j}$ and $\boldsymbol{\psi}$, to show that

$$E\left[\left.\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon^*_i}\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0$$

and

$$E\left[1 + \left.\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon^*_i}\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0 \qquad \text{(C31)}$$

for $i=1,...,N$. Moreover, the maintained assumption of cross-sectional independence of the shocks also implies that

$$E\left[\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \varepsilon_i^*}\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right]=0$$

As a consequence,

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0]=\mathbf{0} \quad \text{and} \quad E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0]=\mathbf{0}.$$

C.3 Variance of the score

If we maintain that $\boldsymbol{\theta}_\infty=\boldsymbol{\theta}_0$ because of the aforementioned consistency, and adapt Proposition D.2 in Fiorentini and Sentana (2023) to a PMLE context, we can show that

$$V[\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0]=\mathcal{B}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)=E\left[\mathcal{B}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)\right]$$

where

$$\mathcal{B}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)=\mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{O}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty), \tag{C32}$$

$$\mathbf{Z}_t(\boldsymbol{\theta})=\begin{pmatrix}\mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_s(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{pmatrix},$$

and

$$\mathcal{O}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)=\begin{bmatrix}\mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)\\ \mathcal{O}_{ls}'(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)\\ \mathcal{O}_{lr}'(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{sr}'(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)\end{bmatrix},$$

with

$$\begin{aligned}
\mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=V[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0],\\
\mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{st}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0],\\
\mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=V[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0],\\
\mathcal{O}_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{rt}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0],\\
\mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}_{rt}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0], \text{ and}\\
\mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)&=V[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0].
\end{aligned}$$

$\mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)$ will be a diagonal matrix of order N with typical element

$$\mathrm{o}_{ll}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=V\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\right|\boldsymbol{\upsilon}_0\right], \tag{C33}$$

$\mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)=\mathrm{O}_{ls}\mathbf{E}_N'$, where $\mathbf{E}_N'$ is the so-called diagonalization matrix and O_{ls} is a diagonal matrix of order N with typical element

$$\mathrm{o}_{ls}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=cov\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*},\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\varepsilon_{it}^*\right|\boldsymbol{\upsilon}_0\right], \tag{C34}$$

$\mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)$ is the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Upsilon}$

of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\mathbf{\Upsilon}_i(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \begin{bmatrix} \text{o}_{ll,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{o}_{ll,i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \text{o}_{ss}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) - 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \text{o}_{ll,i+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \text{o}_{ll,N} \end{bmatrix},$$

where $\text{o}_{ll,i} = \text{o}_{ll}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0)$ to shorten the expressions and

$$\text{o}_{ss}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = V\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\varepsilon_{it}^*\bigg| \boldsymbol{v}_0\right], \tag{C35}$$

$\mathcal{O}_{lr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)$ is an $N \times q$ block diagonal matrix with typical diagonal block of size $1 \times q_i$

$$\text{o}_{lr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -cov\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}, \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i}\bigg| \boldsymbol{v}_0\right], \tag{C36}$$

$\mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \mathbf{E}_N \text{O}_{sr}$, where O_{sr} another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$

$$\text{o}_{sr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -cov\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\varepsilon_{it}^*, \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i}\bigg| \boldsymbol{v}_0\right], \tag{C37}$$

and $\mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)$ is a $q \times q$ block diagonal matrix with typical block of size $q_i \times q_i$

$$\text{o}_{rr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = V\left[\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i}\bigg| \boldsymbol{v}_0\right]. \tag{C38}$$

C.4 Expected Hessian

We can also show that

$$E[-\mathbf{h}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{v}_0] = \mathcal{A}(\boldsymbol{\phi}_\infty, \boldsymbol{v}_0) = E\left[\mathcal{A}_t(\boldsymbol{\phi}_\infty, \boldsymbol{v}_0)\right]$$

where

$$\mathcal{A}_t(\boldsymbol{\phi}_\infty, \boldsymbol{v}_0) = \mathbf{Z}_t(\boldsymbol{\theta}_0)\mathcal{H}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)\mathbf{Z}_t'(\boldsymbol{\theta}_0),$$

$$\mathcal{H}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \begin{bmatrix} \mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{lr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \\ \mathcal{H}_{ls}'(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \\ \mathcal{H}_{lr}'(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{sr}'(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) & \mathcal{H}_{rr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) \end{bmatrix},$$

$$\begin{aligned} \mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) &= -E\left[\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\bigg| \boldsymbol{v}_0\right] \\ \mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) &= -E\left[\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}(\boldsymbol{\varepsilon}_t^{*\prime} \otimes \mathbf{I}_N)\bigg| \boldsymbol{v}_0\right] \\ \mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) &= -E\left[\left\{[\boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_N]\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^*}\right]\right\}[\boldsymbol{\varepsilon}_t^{\prime *} \otimes \mathbf{I}_N]\bigg| \boldsymbol{v}_0\right] \end{aligned}$$

$$\begin{aligned}\mathcal{H}_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0) &= E\left[\left.\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\right|\boldsymbol{v}_0\right]\\ \mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0) &= E\left[[\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N]\left.\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\right|\boldsymbol{v}_0\right]\end{aligned}$$

$\mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ will be a diagonal matrix of order N with typical element

$$\mathrm{H}_{ll}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = -E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{(\partial\varepsilon_i^*)^2}\right|\boldsymbol{v}_0\right], \tag{C39}$$

$\mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ $=\mathrm{H}_{ls}\mathbf{E}_N'$, H_{ls} is a diagonal matrix of order N with typical element

$$\mathrm{H}_{ls}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = -E\left[\frac{\partial^2 \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{(\partial\varepsilon_i^*)^2}\cdot\left.\varepsilon_{it}^*\right|\boldsymbol{v}_0\right], \tag{C40}$$

Given (C31),

$$-E\left[\left\{\left[\mathbf{I}_N\otimes\frac{\partial \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}[\boldsymbol{\varepsilon}_t'^*\otimes\mathbf{I}_N]\right|\left.\boldsymbol{v}_0\right] = \mathbf{K}_{NN},$$

so $\mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ will be the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Gamma}$ of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\boldsymbol{\Gamma}_i(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0) = \begin{bmatrix} \mathrm{H}_{ll,1} & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & \ddots & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \mathrm{H}_{ll,i-1} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \mathrm{H}_{ss}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \mathrm{H}_{ll,i+1} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{H}_{ll,N}\end{bmatrix},$$

where $\mathrm{H}_{ll,i}$ $=\mathrm{H}_{ll}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0)$ to shorten the expressions and

$$\mathrm{H}_{ss}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = -E\left\{\frac{\partial^2 \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_\infty)}{(\partial\varepsilon_i^2)}(\varepsilon_{it}^*)^2\right|\left.\boldsymbol{v}_0\right\}. \tag{C41}$$

$\mathcal{H}_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ is an $N\times q$ block diagonal matrix with typical diagonal block of size $1\times q_i$

$$\mathrm{H}_{lr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*;_{i\infty})}{\partial\varepsilon_i^*\partial\boldsymbol{\varrho}_i'}\right|\boldsymbol{v}_0\right], \tag{C42}$$

$\mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0) = \mathbf{E}_N\mathrm{H}_{sr}$, where H_{sr} another block diagonal matrix of order $N\times q$ with typical block of size $1\times q_i$

$$\mathrm{H}_{sr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = E\left[\frac{\partial^2 \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon_i^*\partial\boldsymbol{\varrho}_i'}\left.\varepsilon_i^*\right|\boldsymbol{v}_0\right], \tag{C43}$$

and $\mathcal{H}_{rr}(\boldsymbol{\varrho}_\infty,\boldsymbol{v}_0)$ is a $q\times q$ block diagonal matrix with typical block of size $q_i\times q_i$

$$\mathrm{H}_{rr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{v}_0) = -E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial\boldsymbol{\varrho}_i\partial\boldsymbol{\varrho}_i'}\right|\boldsymbol{v}_0\right]. \tag{C44}$$

Additional references

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society* B, 38, 45-53.

Fiorentini, G., and Sentana, E. (2021): "Specification tests for non-Gaussian maximum likelihood estimators", *Quantitative Economics* 12, 683-742.

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

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

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.

Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1–25.