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

Abstract

We propose specification tests for independent component analysis and structural vector autoregressions that assess the assumed cross-sectional independence of the non-Gaussian shocks. Our tests effectively compare their joint cumulative distribution with the product of their marginals at discrete or continuous grids of values for its arguments, the latter yielding a consistent test. We explicitly consider the sampling variability from using consistent estimators to compute the shocks. We study the finite sample size of our tests in several simulation exercises, with special attention to resampling procedures. We also show that they have non-negligible power against a variety of empirically plausible alternatives.

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

JEL: C32, C52

*We would like to thank participants at the XII Time Series Econometrics Workshop (Zaragoza, March 2022), the National University of Seoul, the NBER - NSF Time Series Conference (Boston, September 2022) and CEMFI's Fall Econometrics Conference (November 2022) for useful comments and suggestions. 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. Specifically, if in 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

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

then the matrix $\mathbf{C}$ of loadings of the observed variables on the latent ones can be identified (up to column permutations and sign changes) from an *i.i.d.* sample of observations on $\mathbf{y}$ provided the following assumption holds:[2]

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 ($\mathbf{I}$) and margins, but also the same non-linear dependence structure, rendering $\mathbf{C}$ underidentified.

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

[1] See Fiorentini and Sentana (2022a) for a selected list of recent Svar papers that exploit the non-Gaussian features of the structural shocks.

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

cross-moments, which should be equal to the product of the corresponding marginal moments under independence. 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.

The purpose of this paper is to provide alternative moment tests of independence which are 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 relate our discrete grid test to the classical Pearson's independence test statistic for categorical variables in contingency tables.

Importantly, though, 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 (2020) and the references therein), in this paper we consider the discrete mixtures of normals-based pseudo maximum likelihood estimators (PMLEs) in Fiorentini and Sentana (2022a) for three main reasons. First, they are consistent for the model parameters under standard regularity conditions provided that Assumption 1 holds regardless of the true marginal distributions of the shocks. Second, they seem to be rather efficient, the rationale being that finite normal mixtures can provide good approximations to many univariate distributions. And third, the influence functions on which they are based are the scores of the pseudo log-likelihood, which we can easily compute in closed-form. As 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 many empirical finance applications of SVARs, the number of observations is sufficiently large for asymptotic approximations to be reliable. In contrast, the limiting distributions of our tests may be a poor guide for the smaller samples typically used in macroeconomic applications. For that reason, we thoroughly study the finite sample size of our tests in several Monte Carlo exercises. We also discuss some resampling procedures that seem to improve their reliability. Finally, we show that our tests have non-negligible power against a variety of empirically plausible alternatives in which the cross-sectional independence of the shocks no longer holds.

The rest of the paper is organised as follows. Section 2 discusses the model and the estimation procedure. Then, we present our 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. Finally, we present our conclusions and suggestions for further research in section 6, and relegate proofs, auxiliary results and some technical material to the appendix.

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)) and the generalised error (or Gaussian) distribution, which includes normal, Laplace and uniform as special cases.

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 (2022a), where we paid particular attention to third and fourth cross-moments. Nevertheless, 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 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 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

Importantly, we focus on moment conditions for the latent shocks rather than the observed variables because Assumption 1 is written in terms of the ε^*'s instead of the y_i's. Let $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\mathbf{A}_1), \ldots, vec'(\mathbf{A}_p), vec'(\mathbf{C})]' = (\boldsymbol{\tau}', \mathbf{a}_1', \ldots, \mathbf{a}_p', \mathbf{c}') = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$ denote the structural parameters characterising the first two conditional moments of $\mathbf{y}_t$. If we knew the true values of $\boldsymbol{\theta}_0$, we could easily recover the true shocks from the observed variables using the expression

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

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 (2022a) 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.[3] Let $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1', \ldots, \boldsymbol{\varrho}_N')'$ denote those shape parameters, so that $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho}')'$. 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 $\boldsymbol{\phi}$.

[3] 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 (2022b).

3 Discrete grid tests

For pedagogical reasons, in sections 3.1 and 3.2 we first assume that $\boldsymbol{\theta}_0$ is known but later explain how to correct the covariance matrix of the relevant influence functions for the PML estimation of these parameters.

3.1 An event-based approach

Our first test is based on the joint probability of events defined before the sample is observed 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 $k_{H+1} = \infty$.[4] We then collect the indices of the shocks involved in the set $I = \{i, i', ..., i^{\mathrm{I}}\}$, where I denotes the cardinality of the set I, so that we can test for pairwise independence, joint independence of the entire vector of structural innovations, 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 $x \in A$. Finally, we denote by $\upsilon_{\mathbf{h}}^{\diamond}$ the difference between

$$\Pr\left(\bigcap_{i \in I} \{P_{h_i t}^{\diamond i} = 1\}\right),$$

which is the joint probability of the event defined by the vector $\mathbf{h}$ with typical element h_i, and the product of the marginal probabilities $\pi_{h_i}^{\diamond i} = \Pr(P_{h_i t}^{\diamond i} = 1)$, so that $\upsilon^{\diamond}(\mathbf{h}) = \mathbf{0}$ under independence. Using this notation, we could in principle test the null on the basis of the following influence function:

$$p_{\mathbf{h}}^{\diamond}(\boldsymbol{\varepsilon}_t^*) = \prod_{i \in I} P_{h_i t}^{\diamond i} - \prod_{i \in I} \pi_{h_i}^{\diamond i} - \upsilon^{\diamond}(\mathbf{h}). \tag{5}$$

However, (5) is not computable unless one knows the marginal probabilities, as 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

$$E[p_{h_i}^{\diamond}(\varepsilon_{it}^*)] = 0, ..., E[p_{h_{i^{\mathrm{I}}}}^{\diamond}(\varepsilon_{i^{\mathrm{I}}t}^*)] = 0,$$

where

$$p_h^{\diamond}(\varepsilon_{it}^*) = P_{ht}^{\diamond i} - \pi_h^{\diamond i}, \text{ for } i \in I,\ h = 1, ..., H, \tag{6}$$

[4] 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.

which results in the analogue estimator $\hat{\pi}_h^{\diamond i} = \frac{1}{T}\sum_{t=1}^{T} P_{ht}^{\diamond i}$, a fact that we need to take into account in computing the asymptotic covariance matrix of the feasible version of (5) that adequately reflects the sampling uncertainty in $\hat{\pi}_h^{\diamond i}$ for all intervals and shocks.

If we then consider an N-dimensional contingency table whose cells are the Cartesian product of the different marginal partitions, we will end up with a GMM version of Pearson's joint (or multi-way) independence test, which is in fact numerically identical to Pearson's original test statistic (see Sentana (2022)).[5]

3.1.1 A re-interpretation in terms of cdfs

Consider now replacing 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 $h = 1, ..., H+1$. Two things are immediately obvious regardless of the independence between the shocks. First,

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

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

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

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

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

because

$$\Pr\left[\bigcap_{i\in I}\{k_{j_i-1} \leq \varepsilon_{it}^* \leq k_{j_i}\}\right] = \prod_{i\in I}\Pr(k_{j_i-1} \leq \varepsilon_{it}^* \leq k_{j_i}) \quad \forall i \in I \text{ and } \forall k_{j_i}, j \in H.$$

Thus, the usual Pearson test for independence can be easily re-written in our context as a moment test of independence of the cdf at a finite grid of points because the influence functions of the latter are a simple full-rank linear transformation of the former with known coefficients. This re-interpretation will allow us to extend our tests to a continuous grid in section 4.

[5] 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 I shocks, but the choice of excluded category for each shock is arbitrary.

For practical purposes, let us define

$$P_{kt}^{i}=1_{(-\infty,k)}(\varepsilon_{it}^{*}),$$

and

$$p_{k}(\varepsilon_{it}^{*})=P_{kt}^{i}-u_{k}^{i}, \tag{7}$$

where $u_k^i = E(P_{kt}^i) = F_i(k)$, as the new dummy variables and marginal influence functions, respectively, which trivially give rise to the analogue estimator

$$\hat{u}_{k}^{i}=\frac{1}{T}\sum_{t=1}^{T}P_{kt}^{i}. \tag{8}$$

Let us also define the joint influence function

$$p_{\mathbf{k}}(\boldsymbol{\varepsilon}_{t}^{*})=\prod_{i\in I}P_{k_{i}t}^{i}-\prod_{i\in I}u_{k_{i}}^{i}-\upsilon(\mathbf{k}), \tag{9}$$

where $\mathbf{k}=(k_i,k_{i'},\ldots,k_{i^{\text{I}}})'$, which is such that $\upsilon(\mathbf{k})=0$ under the independence null.

Importantly, the fact that the estimating moment conditions (7) exactly identify the relevant u_h^i's implies that there is no efficiency loss in sequentially estimating the $\upsilon(\mathbf{k})$'s from (9) by replacing the marginal cdfs by their sample counterparts relative to estimating them jointly from (7) and (9), which in turn implies that the non-centrality parameters of corresponding moment tests that impose $\upsilon(\mathbf{k})=0$ will coincide.

The following proposition is crucial to compute the relevant test statistics:

Proposition 1 *If the shocks defined by I are stochastically independent, then the asymptotic covariance of the influence functions $p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*)$ and $p_{\mathbf{k}'}(\boldsymbol{\varepsilon}_t^*)$ evaluated at the estimated values of u_k^i and $v_{k'}^i$ in (8), will be given by*

$$\prod_{i\in I}\min(u_{k_i}^{i},v_{k_i'}^{i})+(\text{I}-1)\prod_{i\in I}u_{k_i}^{i}v_{k_i'}^{i}-\sum_{i\in I}\min(u_{k_i}^{i},v_{k_i'}^{i})\left(\prod_{i'\in I,i'\neq i}u_{k_{i'}}^{i'}\right)\left(\prod_{i'\in I,i'\neq i}v_{k_{i'}'}^{i'}\right), \tag{10}$$

where $u_{k_i}^i = F_i(k_{h_i})$ and $v_{k_i'}^i = F_i(k_{h_i'})$.

As we show in the proof of the proposition, expression (10) coincides with the covariance between versions of the influence functions $p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*)$ and $p_{\mathbf{k}'}(\boldsymbol{\varepsilon}_t^*)$ linearised with respect to u_k^i and $v_{k'}^i$, respectively. This linearised versions are particularly useful when $\boldsymbol{\theta}_0$ is unknown and has to be estimated, a topic to which we turn next.

3.1.2 Adjustments for the estimation of $\boldsymbol{\theta}$

Let $\mathbf{m}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ denote a vector of influence functions that depend on $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, a parameter-dependent transformation of the data given by (4). Similarly, let $\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi})$ 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, and $\boldsymbol{\phi}_\infty = (\boldsymbol{\theta}'_0, \boldsymbol{\varrho}'_\infty)'$ the pseudo-true values of the estimated parameters. We can then use the theory of moment tests (see Newey (1985) and Tauchen (1985)) to derive the asymptotic covariance matrix of the (scaled) sample averages of $\mathbf{m}[\varepsilon^*_t(\hat{\boldsymbol{\theta}})]$ that appears in Lemma 1 in Appendix B by combining the joint asymptotic covariance matrix of the (scaled) sample averages of $\mathbf{m}[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)]$ and $\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)$ with the limiting expected value of $\partial\mathbf{m}[\varepsilon^*_t(\boldsymbol{\theta}_0)]/\partial\boldsymbol{\theta}'$.

Let us now apply this general result to $p_{\mathbf{k}}[\varepsilon^*_t(\hat{\boldsymbol{\theta}})]$ and $p_{\mathbf{k}'}[\varepsilon^*_t(\hat{\boldsymbol{\theta}})]$. Specifically, if we combine the expressions in Proposition 1, Lemma 2 and Lemma 3 with the expressions for the covariance matrix of the pseudo log-likelihood score and the expected value of its Hessian in appendix C, we can easily obtain the adjusted covariance matrix of these influence functions evaluated at the estimated values of u^i_k and $v^i_{k'}$ in (8). Thus, our proposed test will differ from Pearson's independence test in that it takes into account not only the estimation of the marginal cdfs, as in Proposition 1, but also because those probabilities will be computed on the basis of estimated ε^*_t's that replace $\boldsymbol{\theta}_0$ with its PMLE $\hat{\boldsymbol{\theta}}$.

Unfortunately, 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^{I} degrees of freedom. Intuitively, a too fine 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, an issue which we will revisit in section 4. 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.[6]

In turn, the choice of the k's will also crucially affect power even though it does not affect the (first-order) asymptotic distribution of the test under the null. Therefore, it would be useful to adapt the grid to the marginal distribution of the shocks. For that reason, in the next section we suggest a simple way to choose the partition which achieves precisely that goal.

3.2 A copula-based approach

Suppose that, 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, we begin by collecting the relevant indices in $I = \{i, i', ...i^{\text{I}}\}$ as in

[6] In principle, we could further deviate from Pearson's test by not necessarily including all 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.

section 3.1, and select H probabilities $0 < u_1 < \cdots < u_H < 1$.[7] Let $\varkappa_i(u_h)$ be the u_h-quantile of ε^*_{it} for $h = 1, ..., H$ after suitably defining $u_0 = 0$ and $u_{H+1} = 1$, with $\varkappa_i(0) = -\infty$ and $\varkappa_i(1) = \infty$. Next, we define the dummy variables

$$Q_{iu_h}(\varepsilon^*_{it}) = 1_{(-\infty, \varkappa_i(u_h))}(\varepsilon^*_{it})$$

for each $i \in I$. In this notation, a straightforward independence test for the shocks could be alternatively computed on the basis of the influence functions

$$q_{\mathbf{u}}(\boldsymbol{\varepsilon}^*_t) = \prod_{i \in I} Q_{iu_h}(\varepsilon^*_{it}) - \prod_{i \in I} u_{h_i} - \upsilon(\mathbf{u}) \tag{11}$$

because $\upsilon_{\mathbf{u}} = 0$ under the null, where $\mathbf{u} = (u_i, u_{i'}, ..., u_{i^{\mathrm{I}}})$. 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.

However, (11) is not computable unless one knows the marginal quantiles. Therefore, in practice the quantiles $\varkappa_i(u_{h_i})$ for the chosen probabilities u_h for each $i \in I$ will usually be estimated in turn from the exactly identified moment conditions

$$E[q_{h_i}(\varepsilon^*_{it})] = 0, ..., E[q_{h_{i^{\mathrm{I}}}}(\varepsilon^*_{i^{\mathrm{I}}t})] = 0,$$

where

$$q_{u_{h_i}}(\varepsilon^*_{it}) = Q_{iu_{h_i}}(\varepsilon^*_{it}) - u_{h_i} \text{ for } i \in I, \text{ and } h = 1, \ldots, H \tag{12}$$

which yields the sample marginal quantiles of the shocks involved as the natural analogue estimators.

Once again, the fact that the estimating moment conditions (12) exactly identify the relevant quantiles implies that there is no efficiency loss in sequentially estimating the $\upsilon(\mathbf{u})$'s from (11) by replacing the marginal quantiles by their sample counterparts relative to estimating them jointly from (11) and (12), which in turn implies that the non-centrality parameters of the corresponding moment tests that impose $\upsilon(\mathbf{u}) = 0$ will also coincide.

An obvious question at this stage is whether practitioners should rely on the event-based approach in section 3.1 or the copula-flavoured test in this one. 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,

[7] For notational simplicity, we again maintain the assumption that the number of intervals and their limits are common across shocks even though it would be straightforward to relax it.

it turns out that

Proposition 2 *If the shocks defined by I are stochastically independent, then the asymptotic covariance of the influence functions $q_{\mathbf{u}}(\boldsymbol{\varepsilon}_t^*)$ and $q_{\mathbf{v}}(\boldsymbol{\varepsilon}_t^*)$ evaluated at the estimated values of $\varkappa_i(u_{h_i})$ and $\varkappa_i(v_{h'_i})$ in (12) will be given by (10).*

In particular, this means that if we chose the limits of the intervals of the test in section 3.1 so that they exactly matched the theoretical quantiles of the test in the 3.2, or in simpler terms, so that the two population partitions were identical, the moment tests for independence based on (9) and (12) would be asymptotically equivalent because the asymptotic covariance matrices that correct for the estimation of the marginal probabilities or the marginal quantiles would also be identical.

As we show in the proof of Proposition 2, expression (10) also coincides with the covariance between versions of the influence functions $q_{\mathbf{u}}(\boldsymbol{\varepsilon}_t^*)$ and $q_{\mathbf{v}}(\boldsymbol{\varepsilon}_t^*)$ linearised with respect to $\varkappa_i(u_{h_i})$ and $\varkappa_i(v_{h'_i})$ when these are in turn linearised with respect to u_h^i and $v_{h'}^i$. Once more, these linearised versions are particularly useful when $\boldsymbol{\theta}_0$ is unknown and has to be estimated, a topic to which we turn next.

3.2.1 Adjusting for the estimation of $\boldsymbol{\theta}$

Although in this case we know the marginal probabilities by construction, our proposed test will once more differ from Fisher (1922) exact independence test because it takes into account not only the estimation of the quantiles of the marginal densities, but also that those quantiles will be computed on the basis of estimated $\boldsymbol{\varepsilon}_t^*$'s that replace $\boldsymbol{\theta}_0$ with its PMLE $\hat{\boldsymbol{\theta}}$. To do so, we can again apply Lemma 1 in Appendix B to $q_{\mathbf{u}}[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}})]$ and $q_{\mathbf{v}}[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}})]$. Specifically, if we combine the expressions in Proposition 2, Lemma 4 and Lemma 5 with the expressions for the covariance matrix of the pseudo log-likelihood score and the expected value of its Hessian in appendix C, we can obtain the adjusted covariance matrix of these influence functions evaluated at the estimated values of $\varkappa_i(u_{h_i})$ and $\varkappa_i(v_{h'_i})$ in (12).

Importantly, it turns out that the sampling variability in estimating the mean parameters or the diagonal elements of the matrix $\mathbf{C}$, which characterise the scale of the different shocks, is totally irrelevant. Intuitively, the reason is that a contingency table based on quantiles is numerically invariant to affine linear transformations of each shock because the new quantiles are the same affine transformation of the original ones. Therefore, the only parameters whose sampling variability matter are the off-diagonal elements of $\mathbf{C}$.

4 A continuous grid

Unfortunately, the tests discussed in sections 3.1 and 3.2 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 if we only considered 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. For that reason, 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 also considered Kolmogorov-Smirnov-type tests based on the maximum absolute discrepancy.[8] However, those tests rely on specific functionals of the difference, while the discrete grid tests 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 (1961) test would do, but more importantly, the covariance between those influence functions for different values of the arguments.

4.1 Moment tests with a continuum of moments

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 $\mathbb{R}^{\text{I}}$ 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.

[8] 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.

Specifically, we can regard (9) under $H_0 : \upsilon(\mathbf{k}) = 0$ as the sample version of

$$p(\mathbf{k}) = F_{\mathbf{h}}(\mathbf{k}) - \prod_{i \in I} F_i(k_i),$$

which should be identically 0 for all $\mathbf{k}$ if and only if the underlying random variables are independent for any random vector with continuous joint cdf. In practice, the marginal cdfs $F_i(x_i)$ will typically be unknown, but we can similarly estimate them by regarding the expected value of (7) as yet another continuum of exactly identified moment conditions, which effectively lead to the empirical cdf of the i^{th} shock.

Importantly, a straightforward extension of the arguments in section 3.2 implies that the continuum of moments test that looks at (9) 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. In effect, we can do so by transforming ε^*_{it} into its empirical uniform rank

$$\epsilon^*_{it} = \frac{1}{T} \sum_{s=1}^{T} 1_{(-\infty, \varepsilon^*_{it})}(\varepsilon^*_{is}),$$

so that $1_{(-\infty, \varkappa(u_i))}(\varepsilon^*_{it})$ becomes $1_{(0,u_i)}(\epsilon_{it})$. Thus, we can define

$$q_{it}(u_i) = 1_{(0,u_i)}(\epsilon^*_{it}) - u_i, \tag{13}$$

$$q_t(\mathbf{u}) = \prod_{i \in I} 1_{(0,u)}(\epsilon^*_{it}) - \prod_{i \in I} u_i, \tag{14}$$

and

$$\bar{q}_T(\mathbf{u}) = \frac{1}{T} \sum_{t=1}^{T} q_t(\mathbf{u}). \tag{15}$$

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

$$\langle f, g \rangle = \int_{[0,1]^{\mathrm{I}}} f(\mathbf{u})\, g(\mathbf{u}) \varpi(\mathbf{u})\, d\mathbf{u}.$$

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, as T goes to infinity

$$\sqrt{T} \bar{q}_T(\mathbf{u}) \Rightarrow \mathcal{N}(0, K)$$

in $L^2(\varpi)$, where $\mathcal{N}(0,K)$ denotes 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

$$(Kf)(\mathbf{u}) = \int_{[0,1]^{\mathrm{I}}} k(\mathbf{u},\mathbf{v})f(\mathbf{v})\varpi(\mathbf{v})\,d\mathbf{v} \tag{16}$$

whose kernel $k(\mathbf{u},\mathbf{v}) = E[p_t(\mathbf{u})p_t(\mathbf{v})]$ is given by (10).

As we mentioned before, we are interested in applying an overidentifying restrictions 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, 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 $j,k = 1,2,\ldots$.[9] 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

$$\left\langle \sqrt{T}\bar{q}_T, K^{-1}\bar{q}_T\right\rangle = \sum_j\sum_k \frac{1}{\lambda_{kj}}\left|\left\langle \sqrt{T}\bar{q}_T, \mu_{jk}\right\rangle\right|^2. \tag{17}$$

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.

[9] It is not worth extending their results for $\mathrm{I} > 2$ because they apply to observed variables rather than estimated shocks.

Thus, the regularised version of the J-type test will be

$$\left\|(K^{\alpha})^{-1/2}\sqrt{T}\bar{q}_T\right\|^2 = \sum_j \sum_k \frac{\lambda_{jk}}{\lambda_{jk}^2+\alpha}\left|\left\langle \sqrt{T}\bar{q}_T, \mu_{jk}\right\rangle\right|^2. \tag{18}$$

Comparing the expressions (17) and (18), 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 (18), which uses as eigenvalues and eigenfunctions those of K, in terms of certain matrices and vectors (see Carrasco et al (2007) for analogous expressions for K under time series dependence). Specifically, we use the following computationally convenient expression for (18):

$$\mathbf{W}'\{\alpha\mathbf{I}_T + [(\mathbf{I}_T - \ell_T\ell_T'/T)\mathbf{D}^2(\mathbf{I}_T - \ell_T\ell_T'/T)]^2\}^{-1}\mathbf{W} \tag{19}$$

where $\mathbf{W}$ is a $T\times 1$ vector whose t^{th} element is $w_t = \int q_t(\mathbf{u})\,\bar{q}_T(\mathbf{u})\,\varpi(\mathbf{u})\,d\mathbf{u}$, $\mathbf{D}$ is a $T\times T$ matrix whose $(t,s)^{th}$ element is $d_{ts} = \langle q_t, q_s\rangle/T$, and ℓ_T is a $T\times 1$ vector of ones. In practice, only $\mathbf{D}$ is needed in order to compute the test statistic since (19) is equivalent to

$$\ell_T'\mathbf{D}(\mathbf{I}_T - \ell_T\ell_T'/T)\{\alpha\mathbf{I}_T + [(\mathbf{I}_T - \ell_T\ell_T'/T)\mathbf{D}^2(\mathbf{I}_T - \ell_T\ell_T'/T)]^2\}^{-1}(\mathbf{I}_T - \ell_T\ell_T'/T)\mathbf{D}\ell_T.$$

The following proposition provides analytical expressions for the elements of the matrix $\mathbf{D}$:

Proposition 3 *If the* I *shocks in* I *are stochastically independent, then*

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

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

4.2 Adjusting for the estimation of $\boldsymbol{\theta}$

Although in the previous section we have already considered the effects of estimating the marginal cdfs of the shocks on the covariance operator, in practice we must take again take into account the sampling variability in estimating $\boldsymbol{\theta}$ by PML. 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. Otherwise, all the expressions that we have derived in sections 3.1.2 and 3.2.1 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 (14) 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 section 3.2.1, though, 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 3 with Lemma 6 in Appendix B. To use this result in practice, though, we need to replace the integrals in (B20) by sums over the empirical cdfs of the shocks. For example, if we denote by $\boldsymbol{\epsilon}_t^*(\hat{\boldsymbol{\theta}}) = [\epsilon_{it}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\mathrm{I}t}^*(\hat{\boldsymbol{\theta}})]$ the vector containing the empirical ranks of the t^{th} observation of each of the estimated shocks that appear in I, we can estimate the scalar $\mathcal{C}$ that appears in Lemma 6 as

$$\widehat{\mathcal{C}} = \sum_{\tau^i=1}^{T} \cdots \sum_{\tau^{\mathrm{I}}=1}^{T} E\left\{\frac{\partial n_t[\epsilon_{\tau^i}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\tau^{\mathrm{I}}}^*(\hat{\boldsymbol{\theta}})]}{\partial \boldsymbol{\theta}'}\right\} \mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})\mathcal{B}(\hat{\boldsymbol{\theta}})\mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})E\left\{\frac{\partial n_t[\epsilon_{\tau^i}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\tau^{\mathrm{I}}}^*(\hat{\boldsymbol{\theta}})]}{\partial \boldsymbol{\theta}}\right\},$$

where $n_t(\cdot)$ denotes the linearised version of (14) that accounts for estimation of the quantiles through (15), whose explicit expression is given by (A6) in the proof of Lemma 6.

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 the power of our proposed tests to that of the integer moment based tests in Amengual, Fiorentini and Sentana (2022a).

5.1 Design and computational details

To keep CPU time within bounds, we focus on bivariate and trivariate static models, as the sampling variability of estimating the VAR coefficients is irrelevant for the copula-based tests that we propose. Specifically, we generate samples of size T from the following static processes

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 1/2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \end{pmatrix} \tag{20}$$

and

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \\ \varepsilon_{3t}^* \end{pmatrix}. \tag{21}$$

Our PML estimation procedure, though, assumes that the matrix of the impact multipliers is fully unconstrained and does not exploit the restriction that the loading matrix of the shocks is upper triangular. Importantly, given that we can easily prove that the estimated shocks are numerically invariant to affine transformations of the y's, and that the same is true of the different test statistics, our results below do not depend on our choice of $\boldsymbol{\tau}$ or $\mathbf{C}$.

We consider both $T = 250$, which is realistic in most macro applications with monthly or quarterly data, and $T = 1,000$, which is representative of financial applications with daily data. 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, 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

$$\varepsilon_{it}^* = \begin{cases} N[\mu_1^*(\boldsymbol{\varrho}_i), \sigma_1^{*2}(\boldsymbol{\varrho}_i)] \text{ with probability } \lambda_i \\ N[\mu_2^*(\boldsymbol{\varrho}_i), \sigma_2^{*2}(\boldsymbol{\varrho}_i)] \text{ with probability } 1-\lambda_i \end{cases} \quad (22)$$

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

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

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.[10]

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

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

where $\phi(\varepsilon; \mu, \sigma^2)$ denotes the pdf of a Gaussian random variable with mean μ and variance σ^2 evaluated at ε. We maximise the log-likelihood with respect to the N elements of $\boldsymbol{\tau}$, 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.[11]

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.[12] To exploit this property, we start the iterations by obtaining consistent initial

[10] We can trivially extend this procedure to three or more components if we replace the normal random variable in the first branch of (22) by a k-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\varrho})$ and $\sigma_1^{*2}(\boldsymbol{\varrho})$, respectively, so that the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

[11] 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 $\overline{\lambda} = 1 - 2/T$.

[12] 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

estimators of $\boldsymbol{\tau}$ and $\mathbf{C}$, $\overline{\boldsymbol{\tau}}_{FICA}$ and $\overline{\mathbf{C}}_{FICA}$ say, using the *FastICA* algorithm of Gävert, Hurri, Särelä, and Hyvärinen.[13] In addition, 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 $\overline{\boldsymbol{\varepsilon}}^*_{t,FICA} = \overline{\mathbf{C}}^{-1}_{FICA}(\mathbf{y}_t - \overline{\boldsymbol{\tau}}_{FICA})$.[14]

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, Meitz and Saikkonen (2017). In addition, we impose that $diag(\mathbf{C})$ is positive by simply changing the sign of all the elements of the relevant columns. Naturally, we apply the necessary changes to the shape parameters estimates, and in particular to the sign of δ_i.

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), while in the trivariate case, ε^*_{3t} follows an asymmetric t with the same kurtosis but opposite skewness coefficient as ε^*_{2t}.

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 $\lambda = 0.2$ 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 (2022b) for details).

control the slope of the functions, which we set to 1.

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

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

DGP 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).

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

The theoretical results in Beran (1988) imply that if the usual Gaussian asymptotic approximation provides a reliable guide to the finite sample distribution of the sample version of the moments being tested, critical values obtained by resampling should not only be valid, but also their errors should be of a lower order of magnitude under additional regularity conditions that guarantee the validity of a higher-order Edgeworth expansion. For that reason, we explicitly analyse the performance of applying resampling methods to our proposed tests.

Specifically, we follow Matteson and Tsay (2017) and Davis and Ng (2022) in reshuffling the estimated standardised residuals as follows. For each Monte Carlo sample, we generate another N_{boot} 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

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

where $\tilde{\varepsilon}_{is}^* = \hat{\varepsilon}_{iR_{is}}^*$ and $\hat{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T) = \hat{\mathbf{C}}_T^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\tau}}_T)$ are the estimated residuals in said Monte Carlo sample.[15]

5.1.4 Simulation results

To gauge the finite sample size and power of our proposed independence tests, we generate $5,000$ samples for the designs under the null and $1,000$ for those under the alternative. For each sample, we also compute $N_{boot} = 99$ random permutation samples, as explained in the previous subsection.

In Table 1 we report the results on the finite sample size of the independence tests proposed in sections 3.2.1 and 4.2 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 all cases, although the continuous grid test with $N = 3$ and $T = 250$

[15] 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.

is moderately undersized. Interestingly, the size of discrete Q 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 Q test is not very sensitive to the choice of the regularization 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 cases $N = 2, T = 250$, $N = 3, T = 250$, and $N = 2, T = 1,000$, respectively. For comparison, we have also included the power of the integer moment tests based on the influence functions (3) in Amengual, Fiorentini and Sentana (2022a). Once again, we use the resampling procedures described in the previous subsection to effectively size-adjust the critical values of all tests.

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. The discrete grid test is better than the continuous one when $H \leq 3$, but it becomes worse for larger values of H. 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 $T = 1,000$.

When the true distribution is a mixture of two multivariate Gaussian components (DGP 2), the power of the continuous Q 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 $T = 1,000$.

Under DGP 1 (asymmetric Student t), the integer moment test is the most powerful, with most of its power coming from the co-skewness component. This is perhaps not surprising given that the integer moments that this test uses coincide with the ones underlying the LM tests for copulas in Amengual and Sentana (2020). Nevertheless, the continuous Q test performs reasonably well and it is better than the discrete grid version. When $T = 1,000$, both quantile-based 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 in most cases, a fact that is most evident for the continuous Q test, which on average, is the best of the three reflecting its consistency property.

6 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, 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 (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 (2022a), 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. 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.

An obvious extension of our work would be the calculation of our proposed tests in some of the increasing number of empirical applications that rely on the cross-sectional independence of the shocks. Before doing so, though, it is important to remember that most of those applications rely on two-step estimators for the parameters of the static ICA model (1) or the dynamic SVAR (2) which differ from the discrete mixture of normals-based PMLEs we have considered in this paper. Although the specifications tests that we have proposed could also be applied to shocks computed on the basis of those alternative estimators, the asymptotic covariance matrices that take into account their sampling variability will differ from the ones we have derived.

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 cross-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-Smirnoff functionals that he used.

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 $\mathbf{C}$ 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 (2022) constitutes an interesting topic for further research.

Another important question is what would happen to our proposed tests in the other extreme case in which the true joint distribution of the shocks is Gaussian. 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 because 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 interesting 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. (2022a): "Moment tests for independent components", *SERIEs* 13, 429-474.

Amengual, D., Fiorentini, F. and Sentana, E. (2022b): "PML vs minimum χ^2: the comeback", CEMFI Working Paper 2210.

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.

Beran, R. (1988): "Prepivoting test statistics: a bootstrap view of asymptotic refinements", *Journal of the American Statistical Association* 83, 687–697.

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 Statististics* 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.

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.

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

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

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

Davis, R. and Ng, S. (2022): "Time series estimation of the dynamic effects of disaster shocks", mimeo, Columbia University.

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. (2022a): "Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions", forthcoming in the *Journal of Econometrics.*

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

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.

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.

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in*

Statistics and Econometrics, 3rd edition, Wiley.

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

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.

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

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

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 Letters* 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.

Appendices

A Proofs of Propositions

A.1 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

$$E_0[\delta_{\{\varepsilon_{it}^*-k\}}] = f_i(k). \tag{A1}$$

We will also exploit the fact that, analogously,

$$E_0[\varepsilon_{it}^*\delta_{\{\varepsilon_{it}^*-k\}}] = kf_i(k). \tag{A2}$$

Proposition 1

To simplify the notation, let $u_i = u_{h_i}^i$ and $v_i = u_{h_i'}^i$. Linearising the influence function (8) yields

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

Regarding the influence function (9), we have that

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

Then, the linearised influence function that takes into account the estimation of u_i, for $i \in I$, becomes

$$m_t(\mathbf{u}) = \left[\prod_{i\in I} 1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*) - \prod_{i\in I} u_i\right] - \sum_{i\in I}\left[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in I, i'\neq i} u_{i'}. \tag{A3}$$

Next, we have to compute

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

Regarding the first term,

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

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

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

and

$$E[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)1_{(-\infty,k_{h_i'})}(\varepsilon_{it}^*)]=\min(u_i,v_i) \text{ for } i\in I. \tag{A4}$$

Similarly, the second term becomes

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

where the first equality follows from the fact that

$$
E[1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*) - v_i] = 0 \quad \text{for} \quad i \in I, \tag{A5}
$$

and the last one from (A4). By symmetry, the third term is equal.

Finally,

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

where we have used (A5) in the first equality and (A4) in the second one.

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

Proposition 2

As in Proposition 1, linearising the influence function for $\hat{\varkappa}_i(u_i)$ yields

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

Regarding the influence function (11), we have that

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

Then, the linearised influence function that takes into account the estimation of $\varkappa_i$ becomes

$$n_t(\mathbf{u}) = \left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - \prod_{i\in I} u_i\right] - \sum_{i\in I}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - u_i\right] \prod_{i'\in I, i'\neq i} u_{i'}. \tag{A6}$$

Next, we have to compute

$$\begin{aligned}
E[n_t(\mathbf{u})n_t(\mathbf{v})] &= E\left\{\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - \prod_{i\in I} u_i\right]\left[\prod_{j\in I} 1_{(-\infty,\varkappa_j(v_j))}(\varepsilon^*_{jt}) - \prod_{j\in I} v_j\right]\right\} \\
&\quad -E\left[\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - \prod_{i\in I} u_i\right]\left\{\sum_{j\in I}\left[1_{(-\infty,\varkappa_j(v_j))}(\varepsilon^*_{jt}) - v_j\right]\prod_{j'\in I, j'\neq j} v_{j'}\right\}\right] \\
&\quad -E\left[\left\{\sum_{i\in I}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - u_i\right]\prod_{i'\in I, i'\neq i} u_{i'}\right\}\left[\prod_{j\in I} 1_{(-\infty,\varkappa_j(v_j))}(\varepsilon^*_{jt}) - \prod_{j\in I} v_j\right]\right] \\
&\quad +E\left[\left\{\sum_{i\in I}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - u_i\right]\prod_{i'\in I, i'\neq i} u_{i'}\right\}\right. \\
&\quad \times \left.\left\{\sum_{j\in I}\left[1_{(-\infty,\varkappa_j(v_j))}(\varepsilon^*_{jt}) - v_j\right]\prod_{j'\in I, j'\neq j} v_{j'}\right\}\right]
\end{aligned}$$

Regarding the first term,

$$\begin{aligned}
& E\left\{\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i)]}(\varepsilon^*_{it}) - \prod_{i\in I} u_i\right]\left[\prod_{j\in I} 1_{(-\infty,\varkappa_j(v_j)]}(\varepsilon^*_{jt}) - \prod_{j\in I} v_j\right]\right\} \\
=\ & E\left\{\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i)]}(\varepsilon^*_{it})\right]\left[\prod_{j\in I} 1_{(-\infty,\varkappa_j(v_j)]}(\varepsilon^*_{jt})\right]\right\} + \left(\prod_{i\in I} u_i\right)\left(\prod_{j\in I} v_j\right) \\
& -E\left\{\left(\prod_{j\in I} v_j\right)\left[\prod_{i\in I} 1_{(-\infty,\varkappa(u_i)]}(\varepsilon^*_{it})\right]\right\} - E\left\{\left(\prod_{i\in I} u_i\right)\left[\prod_{j\in I} 1_{(-\infty,\varkappa(v_j)]}(\varepsilon^*_{jt})\right]\right\} \\
=\ & \prod_{i\in I}\min(u_i, v_i) + \prod_{i\in I} u_i v_i - 2\prod_{i\in I} u_i v_i \\
=\ & \prod_{i\in I}\min(u_i, v_i) - \prod_{i\in I} u_i v_i
\end{aligned}$$

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

$$E[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it})] = u_i \text{ for } i \in I,$$

and

$$E[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it})1_{(-\infty,\varkappa_j(v_j))}(\varepsilon^*_{jt})] = \min(u_i, v_j) \text{ for } i, j \in I.$$

Similarly, the second term becomes

$$
\begin{aligned}
&-E\left\{\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - \prod_{i\in I} u_i\right]\left[\sum_{j\in I}\left[1_{(-\infty,\varkappa_j(v_j))}(\varepsilon_{jt}^*) - v_j\right]\prod_{j'\in I, j'\neq j} v_{j'}\right]\right\} \\
= \; & -\sum_{i\in I} E\left[\left[\prod_{i\in I} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*)\right]\left\{\left[1_{(-\infty,\varkappa_i(v_i))}(\varepsilon_{it}^*) - v_i\right]\prod_{i'\in I, i'\neq i} v_{i'}\right\}\right] \\
= \; & \mathrm{I}\prod_{i\in I} u_i v_i - \sum_{i\in I}\min(u_i, v_i)\left(\prod_{i'\in I, i'\neq i} u_{i'}\right)\left(\prod_{i'\in I, i'\neq i} v_{i'}\right)
\end{aligned}
$$

where the first equality follows from the fact that

$$
E[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - u_i] = 0 \;\; \text{for } i \in I,
$$

and the last one from (A4). By symmetry, the third term is equal.

Finally,

$$
\begin{aligned}
&E\left[\left\{\sum_{i\in I}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in I, i'\neq i} u_{i'}\right\}\left\{\sum_{j\in I}\left[1_{(-\infty,\varkappa_j(v_j))}(\varepsilon_{jt}^*) - v_j\right]\prod_{j'\in I, j'\neq j} v_{j'}\right\}\right] \\
= \; & E\left[\left\{\sum_{i\in I}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in I, i'\neq i} u_{i'}\right\}\left[1_{(-\infty,\varkappa_i(v_i))}(\varepsilon_{it}^*) - v_i\right]\prod_{i'\in I, i'\neq i} v_{i'}\right] \\
= \; & \sum_{i\in I}\min(u_i, v_i)\left(\prod_{i'\in I, i'\neq i} u_{i'}\right)\left(\prod_{i'\in I, i'\neq i} v_{i'}\right) - \mathrm{I}\prod_{i\in I} u_i v_i,
\end{aligned}
$$

where we have used (A5) in the first equality and (A4) in the second one.

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

Proposition 3

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

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

with

$$
q_t(\mathbf{u}_I)\, q_s(\mathbf{u}_I) = \prod_{i\in I} 1_{(0,u_i)}(\epsilon_{it}) 1_{(0,u_i)}(\epsilon_{is}) - \prod_{i\in I} u_i 1_{(0,u_i)}(\epsilon_{it}) - \prod_{i\in I} u_i 1_{(0,u_i)}(\epsilon_{is}) + \prod_{i\in I} u_i^2, \quad \text{(A7)}
$$

where we have used (14) evaluated at the observations t and s. Next, we have to compute the integrals for each of the four terms of the right-hand side of (A7). Regarding the first term, under the independence null,

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

As for the second and third ones,

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

Next, integrating the fourth term,

$$\begin{aligned}\int_{[0,1]^{\mathrm{I}}}\left(\prod_{i=1}^N u_i^2\right)d\mathbf{u}_I &= \prod_{i\in I}\left(\int_0^1 u_i^2 du_i\right)\\ &= \left(\frac{1}{3}\right)^i\\ &\equiv d_3.\end{aligned}$$

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. □

B Lemmata

Lemma 1 *Let* $\mathbf{m}[\varepsilon_t^*(\boldsymbol{\theta})]$ *denote a* $K\times 1$ *vector containing a collection of influence functions and* $\hat{\boldsymbol{\theta}}$ *a consistent estimator of* $\boldsymbol{\theta}$*. Under standard regularity conditions*

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

where

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

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

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

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

with

$$\mathcal{A}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = -E[\partial \mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)/\partial \boldsymbol{\phi}')|\boldsymbol{\varphi}_0] \tag{B8}$$

and

$$\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = V[\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]. \tag{B9}$$

Proof. It follows from applying the same steps as in Proposition 1 in Amengual, Fiorentini and Sentana (2022a). □

Lemma 2 *Suppose that model (2) satisfies Assumption 1. Then, the non-zero elements of the expected Jacobian matrix of the linearised $p_{\mathbf{k}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u_h^i in (8) are be given by*

$$\mathrm{J}_{p_{\mathbf{h}} c_{ii'}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -\sum_{i\in I}\sum_{i'\in I, i'\neq i}\left(\prod_{i''\in I, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[k_{h_i}], \quad \textit{for} \quad i \neq i',$$

where $\eta_{h_{i'}} = E_0[\varepsilon_{it}^* 1_{(-\infty,k_{h_i})}(\varepsilon_{it}^*)]$ for $i \in I$.

Proof. From (A3), we have that

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

Moreover, notice that

$$\frac{\partial \varepsilon_{it}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}'} = -\mathbf{c}^{i\cdot},$$

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

and

$$\frac{\partial \varepsilon_{it}^*(\boldsymbol{\theta})}{\partial \mathbf{c}'} = -[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{c}^{i\cdot}]. \tag{B10}$$

Hence, under the independence null,

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

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

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

where the first equality uses (B10), the second one follows from 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 3 *Suppose that model (2) satisfies Assumption 1. Then, the non-zero elements of the covariance matrix between the linearised influence function $p_{\mathbf{k}}[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})]$ evaluated at at $\boldsymbol{\theta}_0$ and the estimated values of u^i_h in (8) and the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$ is given by*

$$cov\{p_{\mathbf{k}}[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)],\mathbf{s}_{c_{ii'}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\}=E[\mathcal{K}_{p_{\mathbf{h}}t}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)],$$

where

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

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

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

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

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

$$cov\{p_k[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)],\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\}=\mathcal{K}_{p_k}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)=E[\mathcal{K}_{p_kt}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)],$$

and

$$cov\{p_{\mathbf{k}}[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)],\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\}=\mathcal{K}_{p_{\mathbf{k}}}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)=E[\mathcal{K}_{p_{\mathbf{k}}t}(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)],$$

where

$$\mathcal{K}_{\cdot t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \left[\begin{array}{ccc} \mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right] \left[\begin{array}{c} \mathcal{K}_{\cdot lt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot st}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot rt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \end{array}\right],$$

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

$$\begin{aligned} \mathrm{K}_{p_k l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p_k(\boldsymbol{\varepsilon}_t^*), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \le k_{h_i})} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \end{aligned} \quad \text{(B11)}$$

and

$$\begin{aligned} \mathrm{K}_{p_{\mathbf{k}} l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i{}^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i' \in I, i' \neq i} u_{i'} \right) E\left\{ 1_{(\varepsilon_{it}^* \le k_{h_i})} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \end{aligned} \quad \text{(B12)}$$

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 $j_i > 0$,

$$\begin{aligned} \mathrm{K}_{p_k s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p_k(\boldsymbol{\varepsilon}_t^*), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \ge k_{h_i})} \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \end{aligned} \quad \text{(B13)}$$

$$\begin{aligned} \mathrm{K}_{p_{\mathbf{k}} s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i' \in I, i' \neq i} \pi_{i'} \right) E\left\{ \varepsilon_{it}^* 1_{(\varepsilon_{it}^* \le k_{h_i})} \cdot \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \end{aligned} \quad \text{(B14)}$$

$$\begin{aligned} \mathrm{K}_{p_{\mathbf{k}} s_2}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{i't}^* \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= -\left(\prod_{i'' \in I, i'' \neq i, i'' \neq i'} u_{i''} \right) \eta_{i'} E\left\{ 1_{(\varepsilon_{it}^* \le k_{h_i})} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\}, \end{aligned} \quad \text{(B15)}$$

and zero otherwise.

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

$$\begin{aligned} \mathrm{K}_{p_k r}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= cov\left\{ p_k(\boldsymbol{\varepsilon}_t^*), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \\ &= E\left\{ 1_{(\varepsilon_{it}^* \le k_{h_i})} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \bigg\} \end{aligned} \quad \text{(B16)}$$

$$\mathrm{K}_{p_{\mathbf{k}}r}(\boldsymbol{\varrho}_{\infty},\boldsymbol{\upsilon}_0) = cov\left\{p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*), \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\} = \left(\prod_{i'\in I, i'\neq i} u_{i'}\right) E\left\{1_{(\varepsilon_{it}^*\leq k_{h_i})}\cdot \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\} \tag{B17}$$

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 (8) and the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0$, we can make use of (A3) to write

$$cov\{m_t^u,\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\} = cov\{p_{\mathbf{k}}(\boldsymbol{\varepsilon}_t^*),\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\} - \sum_{i\in I}\left(\prod_{i'\in I,i'\neq i} u_{i'}\right) cov\left\{p_{k_i}(\varepsilon_{it}^*),\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\}. \tag{B18}$$

Then, substituting (B11) and (B12) into (B18), we get

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

and

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

Similarly, substituting (B13) and (B14) into (B18), we get

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

and substituting (B16) and (B17) into (B18), we get

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

Finally, substituting (B13) and (B15) into (B18), we get result stated in the proposition. □

Lemma 4 *Suppose that model (2) satisfies Assumption 1. Then, the non-zero elements of the expected Jacobian matrix of the linearised influence function $q_{\mathbf{u}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ evaluated at $\boldsymbol{\theta}_0$ and the estimated values of $\varkappa_i(u_i)$ evaluated are given by*

$$\mathrm{J}_{q_{\mathbf{u}}c_{ii'}}(\boldsymbol{\varrho}_{\infty},\boldsymbol{\varphi}_0) = -\sum_{i\in I}\sum_{i'\in I,i'\neq i}\left(\prod_{i''\in I,i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa_i(u_i)], \text{ for } i\neq i', \tag{B19}$$

where $\xi_{u_{i'}} = E_0[\varepsilon_{i't}^* 1_{(-\infty,\varkappa_{i'}(u_{i'})]}(\varepsilon_{i't}^*)]$.

Proof. The proof is analogous to the one of Lemma 2 by virtue of the symmetry between (A3) and (A6). □

Lemma 5 *Suppose that model (2) satisfies Assumption 1. Then, the non-zero elements of the covariance matrix between the linearised influence function* $q_{\mathbf{u}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$ *evaluated at* $\boldsymbol{\theta}_0$ *and the estimated values of* $\varkappa_i(u_h^i)$ *and the pseudo log-likelihood scores evaluated at the pseudo true values* $\boldsymbol{\phi}_\infty$ *is given by*

$$cov\{q_{\mathbf{u}}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)], \mathbf{s}_{c_{ii'}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\} = E[\mathcal{K}_{q_{\mathbf{u}}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0)],$$

where

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

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

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

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

Proof. The proof is analogous to the one of Lemma 3 by virtue of the symmetry between (A3) and (A6). □

Lemma 6 *Suppose that model (2) satisfies Assumption 1. Then, the adjusted covariance operator that accounts for estimation of* $\boldsymbol{\theta}$ *is given by*

$$\langle q_t, q_s \rangle / T + \mathcal{C},$$

where

$$\mathcal{C} = \int_{[0,1]^{\mathrm{I}}} E\left[\frac{\partial n_t(\mathbf{u}_I)}{\partial \boldsymbol{\theta}'}\right] T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' E\left[\frac{\partial n_s(\mathbf{u}_I)}{\partial \boldsymbol{\theta}'}\right] d\mathbf{u}_I. \tag{B20}$$

Proof. From 4, the Jacobian of the linearised (with respect to $\varkappa$'s) influence function with respect to $\boldsymbol{\theta}$ can be written as

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

We are after

$$\int_{[0,1]^i}\left\{n_t(\mathbf{u}_I) - E\left[\frac{\partial n_t(\mathbf{u}_I)}{\partial \theta}\right]\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\}\left\{n_s(\mathbf{u}_I) - E\left[\frac{\partial n_s(\mathbf{u}_I)}{\partial \theta}\right]\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\} d\mathbf{u}_I.$$

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

$$\int_{[0,1]^i} n_t(\mathbf{u}_I) n_s(\mathbf{u}_I) d\mathbf{u}_I,$$

is given in Proposition 2. Next, we have the cross-terms, which are of the form

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

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

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

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 in $n_t(\mathbf{u}_I)$ and $E\left[\partial n_s(\mathbf{u}_I)/\partial \boldsymbol{\theta}'\right]$. Hence, 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 (B20), 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 (2022a) 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

$$l_t(\mathbf{y}_t; \boldsymbol{\varphi}) = -\ln|\mathbf{C}| + l[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + l[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N], \tag{C21}$$

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

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, \tag{C22a}$$

$$\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{C}\mathbf{C}', \tag{C22b}$$

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

$$\frac{\partial d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial vec'(\mathbf{C})}{\partial \boldsymbol{\theta}} vec(\mathbf{C}^{-1\prime}) = -\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} vec(\mathbf{C}^{-1\prime}) = -\mathbf{Z}'_{st}(\boldsymbol{\theta}) vec(\mathbf{I}_N) \tag{C23}$$

and

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= -\mathbf{C}^{-1}\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{C}^{-1}]\frac{\partial vec(\mathbf{C})}{\partial \boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\}, \end{aligned} \tag{C24}$$

where

$$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\mu}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix} \mathbf{C}^{-1\prime}, \tag{C25}$$

$$\mathbf{Z}_{st}(\boldsymbol{\theta}) = \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{N^2 \times N^2} \\ \vdots \\ \mathbf{0}_{N^2 \times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}), \tag{C26}$$

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

$$\begin{aligned} \mathbf{s}_t(\boldsymbol{\phi}) &= \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_t(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\phi}), \end{aligned} \tag{C27}$$

where

$$\mathbf{e}_{lt}(\boldsymbol{\phi}) = -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} = -\begin{bmatrix} \partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]/\partial \varepsilon_1^* \\ \partial \ln f_2[\varepsilon_{2t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_2]/\partial \varepsilon_2^* \\ \vdots \\ \partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]/\partial \varepsilon_N^* \end{bmatrix}, \tag{C28}$$

$$\mathbf{e}_{st}(\boldsymbol{\phi}) = -vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}$$
$$= -vec\left\{\begin{array}{ccc} 1+\frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & 1+\frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array}\right\} \quad \text{(C29)}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi}) = \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}} = \left\{\begin{array}{c} \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array}\right\} = \left[\begin{array}{c} \mathbf{e}_{r_1t}(\boldsymbol{\phi}) \\ \mathbf{e}_{r_2t}(\boldsymbol{\phi}) \\ \vdots \\ \mathbf{e}_{r_Nt}(\boldsymbol{\phi}) \end{array}\right] \quad \text{(C30)}$$

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 (2021b) implies that

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'}$$
$$+\left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, \quad \text{(C31)}$$

where $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ and $\mathbf{Z}_{st}(\boldsymbol{\theta})$ are given in (C25) and (C26), 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

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

so that

$$\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} = \left[\mathbf{I}_N \otimes \left(\begin{array}{c} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes \mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{array}\right)\right]\frac{\partial vec(\mathbf{C}^{-1\prime})}{\partial \boldsymbol{\theta}'}$$
$$= \left[\mathbf{I}_N \otimes \left(\begin{array}{c} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes \mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{array}\right)\right]\left[\begin{array}{cc} \mathbf{0}_{N^2\times(N+pN^2)} & (\mathbf{C}^{-1}\otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{array}\right].$$

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

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

In addition,

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \quad \text{(C32)}$$

and

$$\begin{aligned} \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{ \mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= \left\{ [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right] \right\} \\ &\quad \times \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. \end{aligned} \quad \text{(C33)}$$

The assumed independence across innovations implies that

$$\frac{\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{(\partial \varepsilon_1^*)^2} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{(\partial \varepsilon_N^*)^2} \end{bmatrix}, \quad \text{(C34)}$$

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

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'},$$

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

with

$$\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^* \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^* \partial \boldsymbol{\varrho}_N'} \end{bmatrix} \tag{C35}$$

because of the cross-sectional independence assumption.

As for the shape parameters of the independent margins,

$$\mathbf{h}_{\boldsymbol{\varrho}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1 \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N \partial \boldsymbol{\varrho}_N'} \end{bmatrix}. \tag{C36}$$

Finally, regarding the Jacobian term $-\ln|\mathbf{C}|$, we have that differentiating (C23) 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

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

Given that the parameters $\boldsymbol{\tau}$, $\mathbf{a}_j = vec(\mathbf{A}_j)$ $(j = 1, \ldots, p)$, $\mathbf{j} = veco(\mathbf{J})$ and$\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$ are consistently estimated regardless of the true distribution, $\mathbf{e}_t(\boldsymbol{\phi}_\infty)$ will be serially independent and not just martingale difference sequences. Moreover, given that

$$\mathbf{Z}(\boldsymbol{\theta}) = E[\mathbf{Z}_t(\boldsymbol{\theta})|\boldsymbol{\varphi}_0] = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times q} \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{q\times N} & \mathbf{0}_{q\times N^2} & \mathbf{I}_q \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_d(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \quad \text{(C37)}$$

has full column rank,

$$E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0] = \mathbf{0} \quad \text{(C38)}$$

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

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 0 \text{ and} \quad \text{(C39)}$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^{*2}(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^{*2}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 1, \quad \text{(C40)}$$

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. 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 $\mathbb{R}^{\dim(\boldsymbol{\phi})}$; (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

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

and

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

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 $\mathbf{j} = vec(\mathbf{J})$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$.

Therefore, we can invoke Proposition 3 in Fiorentini and Sentana (2022a), 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

$$E\left[1 + \left.\frac{\partial \ln f[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon^*_i}\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0 \qquad \text{(C41)}$$

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

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

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 (2022a) 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

$$\mathcal{B}_t(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{O}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty),$$
$$\mathbf{Z}_t(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_s(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},$$

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

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

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

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

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

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

$\mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ 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:

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

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

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

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

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

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

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

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

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

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

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

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

with

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

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

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

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

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

Given (C41),

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

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

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

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

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

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

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

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

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

Table 1: Monte Carlo size of independence Q tests

	Discrete Q tests							Continuous Q tests		
	Asymptotic critical values			Bootstrap critical values				Bootstrap critical values		
	10%	5%	1%	10%	5%	1%		10%	5%	1%
Panel A: $N=2$, $T=250$										
$H=2$	8.3	3.8	0.6	9.2	4.5	0.8	$\alpha=1e^{-05}$	9.3	4.7	1.1
$H=3$	8.1	4.0	0.7	9.0	4.9	1.1	$\alpha=1e^{-06}$	9.6	4.6	1.0
$H=4$	8.1	4.1	0.9	9.3	4.8	1.0	$\alpha=1e^{-07}$	9.4	4.5	1.0
$H=5$	8.4	4.4	0.9	9.9	4.8	0.7	$\alpha=1e^{-08}$	9.3	4.6	1.0
Panel B: $N=3$, $T=250$										
$H=2$	8.0	3.9	0.6	8.7	4.2	1.0	$\alpha=1e^{-05}$	8.4	3.9	0.7
$H=3$	8.4	3.9	0.6	9.5	5.0	0.8	$\alpha=1e^{-06}$	8.6	4.0	0.7
$H=4$	8.6	4.1	0.8	9.2	4.8	0.9	$\alpha=1e^{-07}$	8.5	3.9	0.7
$H=5$	8.0	3.7	0.8	9.1	4.3	0.8	$\alpha=1e^{-08}$	8.6	4.1	0.7
Panel C: $N=2$, $T=1,000$										
$H=2$	9.9	4.2	0.7	10.1	4.8	0.8	$\alpha=1e^{-05}$	10.5	5.3	1.2
$H=3$	10.2	5.1	0.9	10.6	5.7	1.1	$\alpha=1e^{-06}$	10.4	5.6	1.3
$H=4$	9.8	4.6	0.9	10.7	5.2	1.1	$\alpha=1e^{-07}$	10.7	5.8	1.3
$H=5$	9.7	4.9	0.9	10.3	5.6	1.0	$\alpha=1e^{-08}$	10.7	5.7	1.3

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 5,000 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 section 3.2 and 4 for a detailed description of the Discrete and Continuous Q test statistics, respectively. The asymptotic distribution of the Discrete Q test statistic is chi-squared with H^N degrees of freedom. We describe the sampling procedure we use to implement the bootstrap for both, Discrete and Continuous Q test statistics, in section 5.1.3.

Table 2: Monte Carlo power of bivariate independence moment tests: Sample size $T = 250$.

	DGP 1 Scale mixture of two normals			DGP 2 Finite normal mixture			DGP 3 Asymmetric Student t		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete Q tests									
$H = 2$	69.0	56.8	27.0	71.1	57.1	30.0	31.5	21.3	5.5
$H = 3$	54.8	41.7	17.9	90.3	83.6	57.5	36.7	24.9	8.5
$H = 4$	39.2	27.2	7.6	92.4	87.1	64.5	37.1	25.7	8.8
$H = 5$	38.7	26.2	8.4	91.0	82.9	57.7	34.7	23.6	7.3
Panel B: Continuous Q tests									
$\alpha = 1e^{-05}$	55.4	39.9	16.2	96.9	93.0	75.9	46.7	33.4	13.6
$\alpha = 1e^{-06}$	51.9	38.9	14.9	96.7	91.8	74.1	46.3	33.5	13.4
$\alpha = 1e^{-07}$	51.4	37.5	14.6	96.0	91.5	73.5	46.0	33.1	12.7
$\alpha = 1e^{-08}$	51.1	37.5	14.6	95.8	91.8	73.9	45.9	32.9	12.5
Panel C: Integer moment tests									
Co-cov	8.6	4.2	0.8	27.8	20.7	8.1	63.1	56.0	35.6
Co-skew	8.9	4.5	0.7	32.9	22.6	6.8	90.6	83.0	54.0
Co-kurt	14.3	7.0	1.9	29.9	19.0	5.7	61.4	47.3	23.0
Joint	11.6	6.3	1.6	37.6	24.3	7.0	84.8	69.8	25.9

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 1,000 replications. Details on the data generating processes: 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 $vech(\aleph_2) = 0.2\ell_2$ (see Appendix D in Amengual, Fiorentini and Sentana (2022b) 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). See section 3.2 and 4 for a detailed description of the Discrete and Continuous Q test statistics, respectively, and Amengual, Fiorentini and Sentana (2022a) for a description of the Integer moment tests. See section 5.1.3 for a description of the sampling procedure we use to implement the bootstrap for both, Discrete and Continuous Q test statistics.

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

	DGP 1 Scale mixture of two normals			DGP 2 Finite normal mixture			DGP 3 Asymmetric Student t		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete Q tests									
$H = 2$	95.7	92.3	75.7	92.1	86.3	65.0	20.2	11.8	3.3
$H = 3$	95.3	91.8	70.5	99.1	98.1	90.7	23.3	13.2	3.1
$H = 4$	79.9	67.2	37.6	99.8	99.0	94.4	17.6	9.9	1.8
$H = 5$	71.3	57.4	29.1	99.3	98.9	87.9	16.1	7.8	1.7
Panel B: Continuous Q tests									
$\alpha = 1e^{-05}$	79.0	66.7	32.9	100.0	99.5	95.1	55.4	42.1	16.1
$\alpha = 1e^{-06}$	79.2	66.4	32.0	99.9	99.6	94.6	58.2	43.9	18.7
$\alpha = 1e^{-07}$	78.1	64.7	31.5	99.9	99.6	94.2	60.1	46.3	20.6
$\alpha = 1e^{-08}$	76.9	62.5	29.2	99.8	99.5	94.2	59.5	45.6	20.8
Panel C: Integer moment tests									
Cov	10.1	4.0	0.8	34.7	25.1	11.6	73.8	65.7	43.3
Co-skew	11.8	6.0	0.7	41.5	28.2	11.4	96.0	92.3	65.6
Co-kurt	14.8	8.2	1.6	39.7	25.6	8.3	74.1	61.4	33.2
Joint	13.6	8.4	1.6	48.2	33.3	9.7	89.8	74.9	31.3

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 1,000 replications. Details on the data generating processes: 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 $\boldsymbol{\delta}_3 = (0.5,\text{-}0.5, 0)'$ and relative-covariance difference such that $\aleph_3$ is lower triangular with $vech(\aleph_3) = 0.2\ell_6$ (see Appendix D in Amengual, Fiorentini and Sentana (2022b) for details); and DGP 3: Asymmetric Student t with skewness vector $\boldsymbol{\beta} = -10\boldsymbol{\ell}_3$ and degrees of freedom parameter $\nu = 12$ (see Mencía and Sentana (2012) for details). See section 3.2 and 4 for a detailed description of the Discrete and Continuous Q test statistics, respectively, and Amengual, Fiorentini and Sentana (2022a) for a description of the Integer moment tests. See section 5.1.3 for a description of the sampling procedure we use to implement the bootstrap for both, Discrete and Continuous Q test statistics.

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

	DGP 1 Scale mixture of two normals			DGP 2 Finite normal mixture			DGP 3 Asymmetric Student t		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Discrete Q tests									
$H = 2$	100.0	100.0	99.7	100.0	100.0	98.5	88.3	80.7	55.4
$H = 3$	100.0	99.8	99.2	100.0	100.0	100.0	95.7	91.7	75.2
$H = 4$	99.9	99.6	96.0	100.0	100.0	100.0	96.9	94.1	80.7
$H = 5$	100.0	99.7	96.5	100.0	100.0	100.0	97.6	95.3	80.9
Panel B: Continuous Q tests									
$\alpha = 1e^{-05}$	99.3	98.1	91.8	100.0	100.0	100.0	97.2	93.1	77.4
$\alpha = 1e^{-06}$	97.9	96.4	85.2	100.0	100.0	100.0	95.4	90.5	72.2
$\alpha = 1e^{-07}$	97.3	94.6	80.8	100.0	100.0	100.0	94.3	89.1	70.7
$\alpha = 1e^{-08}$	97.2	94.0	79.3	100.0	100.0	100.0	93.8	88.6	70.1
Panel C: Integer moment tests									
Cov	11.5	5.8	1.3	32.4	25.5	12.1	86.1	83.3	75.9
Co-skew	11.1	6.3	1.2	87.0	79.7	52.9	100.0	100.0	99.6
Co-kurt	49.1	38.2	14.8	80.1	72.0	41.9	93.8	90.7	70.4
Joint	41.5	28.7	10.7	92.8	85.7	57.3	100.0	100.0	94.9

Notes: Monte Carlo empirical rejection rates of SVAR specification tests based on 1,000 replications. Details on the data generating processes: 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 $vech(\aleph_2) = 0.2\ell_2$ (see Appendix D in Amengual, Fiorentini and Sentana (2022b) 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). See section 3.2 and 4 for a detailed description of the Discrete and Continuous Q test statistics, respectively, and Amengual, Fiorentini and Sentana (2022a) for a description of the Integer moment tests. See section 5.1.3 for a description of the sampling procedure we use to implement the bootstrap for both, Discrete and Continuous Q test statistics.

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

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

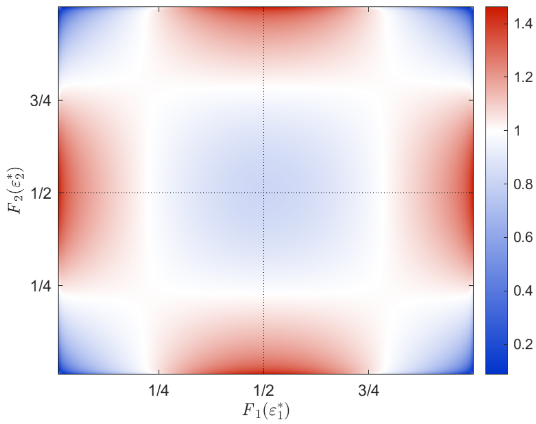
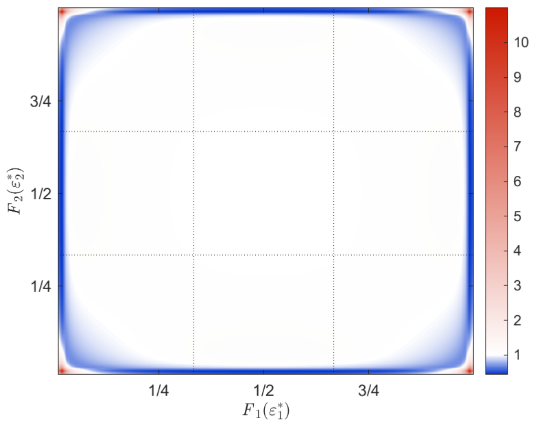


Figure 1b: $d_2 = 0.61$ and $d_3 = -0.39$



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)

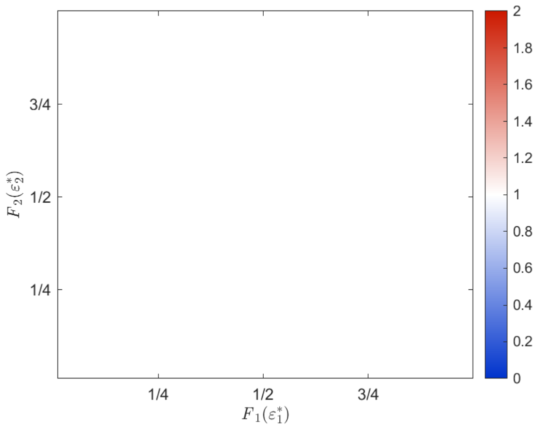


Figure 2b: Scale mixture of normals (DGP 1)

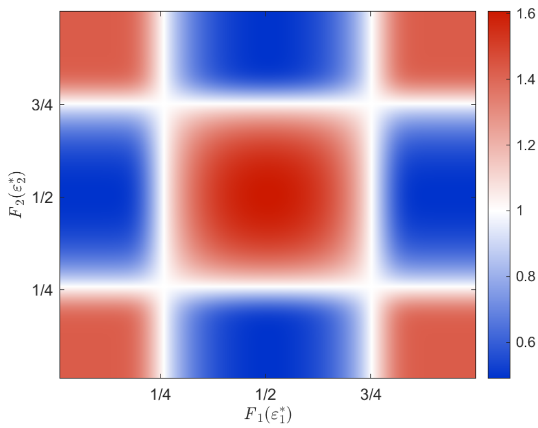


Figure 2c: Mixture of normals (DGP 2)

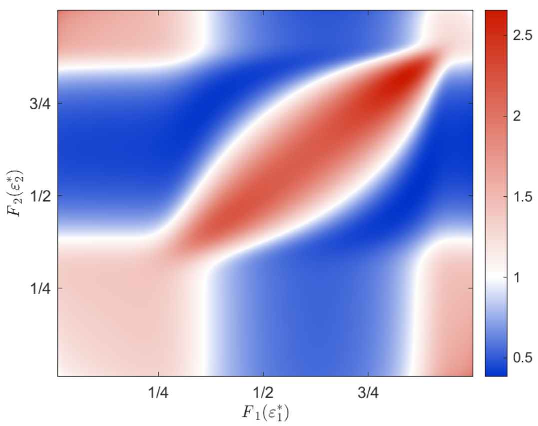
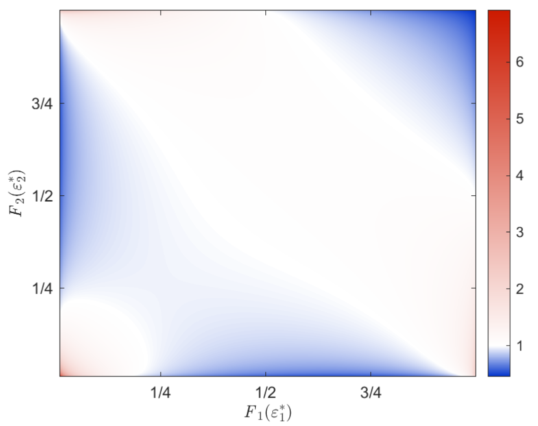


Figure 2d: Asymmetric t (DGP 3)



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 $vech(\aleph_2) = 0.2\ell_2$ (see Appendix D in Amengual, Fiorentini and Sentana (2022b) 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).