cemfi

working paper
2023

Discrete Mixtures of Normals Pseudo Maximum Likelihood Estimators of Structural Vector Autoregressions

Gabriele Fiorentini
Enrique Sentana

October 2020

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2023
October 2020

Discrete Mixtures of Normals Pseudo Maximum Likelihood Estimators of Structural Vector Autoregressions

Abstract

Likelihood inference in structural vector autoregressions with independent non-Gaussian shocks leads to parametric identification and efficient estimation at the risk of inconsistencies under distributional misspecification. We prove that autoregressive coefficients and (scaled) impact multipliers remain consistent, but the drifts and standard deviations of the shocks are generally inconsistent. Nevertheless, we show consistency when the non-Gaussian log-likelihood is a discrete scale mixture of normals in the symmetric case, or an unrestricted finite mixture more generally. Our simulation exercises compare the efficiency of these estimators to other consistent proposals. Finally, our empirical application looks at dynamic linkages between three popular volatility indices.

JEL Codes: C32, C46, C51, C58.

Keywords: Consistency, finite normal mixtures, pseudo maximum likelihood estimators, structural models, volatility indices.

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

Enrique Sentana
CEMFI
sentana

Acknowledgement

The second author acknowledges financial support from the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689.

1 Introduction

Statistical identification of the parameters of a structural vector autoregression (SVAR) through independent non-Gaussian shocks is becoming increasingly popular after Lanne, Meitz and Saikkonen (2017) and Gouriéroux, Monfort and Renne (2017).[1] A selected list of recent papers that exploit the non-Gaussian features of the structural shocks includes Lanne and Lütkepohl (2010), Hyvärinen et al (2013), Moneta et al (2013), Capasso and Moneta (2016), Herwartz and Plödt (2016), Guay and Normandin (2018), Herwartz (2018), Bernoth and Herwartz (2019), Coad and Grassano (2019), Herwartz (2019), Lanne and Luoto (2019), Puonti (2019), Tank, Fox and Shojaie (2019), Bekaert, Engstrom and Ermolov (2019, 2020), Gouriéroux, Monfort and Renne (2020) and Maxand (2020).

Maximum likelihood estimation and inference in SVAR models with independent non-Gaussian shocks is relatively simple to implement, and leads to efficient estimators of all the structural parameters as long as the assumed univariate distributions are correctly specified. Unfortunately, while Gaussian pseudo maximum likelihood estimators (PMLE) remain consistent for the conditional mean and variance reduced form parameters under relatively weak conditions when the true shocks are not Gaussian, the same is not true for many other distributions (see e.g. Newey and Steigerwald (1997)). Nevertheless, this does not mean that all the parameters are inconsistently estimated. In this respect, an important contribution of our paper is to prove that the autoregressive matrices of the VAR and the (scaled) matrix of impact multipliers, which jointly determine the temporal pattern of the Impulse Response Functions (IRFs), continue to be consistently estimated under distributional misspecification. In contrast, we show that in general the standard deviation of the structural shocks will be inconsistently estimated in those circumstances, which distorts the scale of the IRFs and the entire forecast error variance decompositions (FEVDs). Further, we prove that while the drifts of the VAR will also be consistently estimated when both the assumed and true distributions of the shocks are symmetric, they will be inconsistently estimated otherwise, thereby leading to biased forecasts.

In principle, semiparametric (SP) estimators seem to provide a very attractive solution in this context because under appropriate regularity conditions they would be not only consistent but also attain full efficiency for the subset of the parameters that continue to be consistently estimated under distributional misspecification. Unfortunately, SP estimators are usually computed using one BHHH iteration of the efficient score evaluated at a consistent estimator. But for SVARs the usual initial estimator, namely Gaussian PMLE, can only identify the elements

[1] The vast signal processing literature on Independent Component Analysis popularised by Comon (1994) exploits the same identification scheme.

of the impact multiplier matrix up to an orthogonal rotation of order N, so it is of no use.

In Fiorentini and Sentana (2019) (FS), we studied in detail the statistical properties of consistent estimators which replace the parameters that are inconsistently estimated by a misspecified non-Gaussian log-likelihood with the first and second sample moments of residuals readily generated by most software packages. In this respect, another important contribution of the present paper is to show that if the non-Gaussian log-likelihood is based on a discrete scale mixture of normals in the spherically symmetric case, or an unrestricted finite Gaussian mixture more generally, there is no need to replace any of the initial estimators because all the parameters are consistently estimated to begin with. Intuitively, the reason is that the discrete normal mixture-based maximum likelihood estimators of the unconditional mean vector and covariance matrix of an observed series coincide with the first and second sample moments.[2] Similarly, the discrete gamma mixture-based maximum likelihood estimators of the unconditional mean also coincides with the sample mean in the spherically symmetric case. In both cases, though, the shape parameters of the mixture, including the mixing proportions, must be estimated simultaneously with the mean and variance parameters.

Still, the fact that log-likelihoods based on discrete normal mixtures lead to consistent estimators for SVAR models with independent non-Gaussian shocks does not imply that these estimators are more efficient than the two-step FS estimators with an alternative parametric distribution, such as the popular Student t or the Laplace. We study this important issue by means of Monte Carlo simulations. Nevertheless, the fact that under certain conditions discrete mixture of normals with multiple components can provide good approximations to many other distributions (see Hamdan (2006) for scale mixtures of normals and Nguyen et al (2019) for general ones) suggests that the flexible parametric procedure we consider has the potential to achieve the semiparametric efficiency bound. We also compare our estimators to the two-step procedure in Gouriéroux, Monfort and Renne (2017), which estimates all the reduced form parameters by Gaussian PML, and the orthogonal rotation matrix mapping structural shocks and reduced form innovations by non-Gaussian PML.

We would like to emphasise that our results are valid not only for SVARs with cross-sectional independent structural shocks, but also for many dynamic conditionally heteroskedastic multivariate regression models routinely used in empirical finance and other fields, including ARCH-M models and multivariate regressions. In this respect, they provide an alternative justification for the model-specific consistency results in Lee and Lee (2009) and Ha and Lee (2011).

Finally, we apply our proposed estimators to the empirical analysis of the dynamic linkages

[2] This result was first noted by Behboodian (1970) for univariate mixtures but largely ignored in the subsequent literature (but see Supplemental Appendix E.7 in Fiorentini and Sentana (2020a)).

between three popular market-based volatility indices representative of some of the most actively traded asset classes: stocks, exchange rates and commodities. The empirical analysis of such linkages has become a very active area of research (see e.g. Diebold and Yilmaz (2014) and Barigozzi and Brownlees (2019)). Specifically, we analyse the omnipresent VIX, which captures the one-month ahead volatility of the S&P500 stock market index; the EVZ, which computes the 30-day volatility of the $US/Euro exchange rate from options on the CurrencyShares Euro Trust (Ticker - FXE); and the GVZ, which measures the market's expectation of 30-day volatility of gold prices by applying the VIX methodology to options on SPDR Gold Shares (Ticker - GLD) index futures.

The rest of the paper is organised as follows. In section 2 we introduce finite mixtures of normals and present two numerical results that their MLEs satisfy. Then in section 3 we discuss multivariate dynamic regression models with time-varying variances and covariances, and exploit those two numerical results to prove the consistency of the pseudo MLEs based on finite Gaussian mixtures. Section 4 analyses SVAR models with cross-sectionally independent shocks and characterises the parameters that remain consistently estimated under distributional misspecification. Next, in section 5 we present an extensive Monte Carlo exercise that combines several simulation and estimation densities, while in section 6 we carry out our empirical application to the aforementioned volatility indices. This is followed by our concluding remarks. Proofs and auxiliary results are gathered in appendices.

2 Discrete mixture of normals

2.1 General mixtures and their ML estimators

Let $\mathbf{s} = (s_1, \ldots, s_k, \ldots, s_K)$ denote a categorical random variable of dimension K, which is nothing other than a collection of K mutually exclusive Bernoulli random variables with $P(s_k = 1) = \lambda_k$ such that $\sum_{k=1}^K \lambda_k = 1$. If $\mathbf{z}|\mathbf{s}$ is $N(\mathbf{0}, \mathbf{I}_N)$, then

$$\mathbf{x} = \textstyle\sum_{k=1}^K s_k(\boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{1/2}\mathbf{z}), \tag{1}$$

is a K-component mixture of normals, whose first two unconditional moments are

$$\boldsymbol{\pi} = E(\mathbf{x}) = \textstyle\sum_{k=1}^K \lambda_k \boldsymbol{\mu}_k = E[E(\mathbf{x}|\mathbf{s})], \ and \tag{2}$$

$$\boldsymbol{\Omega} = V(\mathbf{x}) = \textstyle\sum_{k=1}^K \lambda_k[(\boldsymbol{\mu}_k\boldsymbol{\mu}_k') + \boldsymbol{\Sigma}_k] - (\sum_{k=1}^K \lambda_k\boldsymbol{\mu}_k)(\sum_{k=1}^K \lambda_k\boldsymbol{\mu}_k') = E[V(\mathbf{x}|\mathbf{s})] + V[E(\mathbf{x}|\mathbf{s})]. \tag{3}$$

The model parameters are $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k, \ldots, \lambda_K)$, subject to the unit simplex restrictions $\lambda_k \geq 0 \ \forall k$ and $\sum_{k=1}^K \lambda_k = 1$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_k', \ldots, \boldsymbol{\mu}_K')'$ and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1', \ldots, \boldsymbol{\sigma}_k', \ldots, \boldsymbol{\sigma}_K')'$, where $\boldsymbol{\sigma}_k = vech(\boldsymbol{\Sigma}_k)$. The representation in (1) is very general, and may give rise to substantially

deviations from multivariate normality through higher order moments.[3] In particular, it nests random vectors consisting of N independent univariate mixtures with K_i components each, in which case $K = \prod_{i=1}^{N} K_i$. Such mixtures play an important role in our analysis of SVARs with cross-sectional independent structural shocks in section 4.

If we observe a random sample of size T on $\mathbf{x}$, ML estimation of the model parameters by numerical methods is conceptually straightforward. Nevertheless, the log-likelihood function of a finite normal mixture has a pole for each observation. Specifically, it will go to infinity if we set $\hat{\boldsymbol{\mu}}_1 = \mathbf{x}_t$ and let $|\hat{\boldsymbol{\Sigma}}_1|$ go to 0. As a result, the ML estimator must be defined as the consistent root of the first order conditions (see Kiefer (1978)). In practice, one may deal with this issue by starting the numerical algorithm from many different values. In addition, there is a trivial identification issue that arises by exchanging the labels of the components, but this is also easy to fix. Boldea and Magnus (2009) provide analytical expressions for the score and Hessian matrix, and compare several numerical algorithms and asymptotic covariance matrix estimators.

However, it is usually convenient to start the recursions from sensibly chosen values. In this respect, the EM algorithm discussed by Dempster, Laird and Rubin (1977) allows us to obtain initial values as close to the MLEs as desired. In the unrestricted case, the recursions are as follows:

$$\hat{\boldsymbol{\mu}}_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}} \frac{1}{T} \sum_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)}) \mathbf{x}_t, \tag{4a}$$

$$\hat{\boldsymbol{\Sigma}}_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}} \frac{1}{T} \sum_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)}) \mathbf{x}_t \mathbf{x}_t' - \hat{\boldsymbol{\mu}}_k^{(n} \hat{\boldsymbol{\mu}}_k^{(n\prime}, \tag{4b}$$

$$\hat{\lambda}_k^{(n} = \frac{1}{T} \sum\nolimits_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)}) \tag{4c}$$

where

$$w_k(\mathbf{x}_t; \boldsymbol{\varphi}) = P(s_{kt} = 1 | \mathbf{x}_t) = \frac{\lambda_k |\boldsymbol{\Sigma}_k|^{-N/2} \phi_N[\boldsymbol{\Sigma}_k^{-1/2} (\mathbf{x}_t - \boldsymbol{\mu}_k)]}{\sum_{j=1}^{K} \lambda_j |\boldsymbol{\Sigma}_j|^{-N/2} \phi_N[\boldsymbol{\Sigma}_j^{-1/2} (\mathbf{x}_t - \boldsymbol{\mu}_j)]} \tag{5}$$

is the posterior probability that observation t comes from the k^{th} component, and $\phi_N(.)$ the spherical normal density of dimension N.[4]

The following proposition, which generalises the univariate result in Behboodian (1970), will prove fundamental for our consistency results:

[3] By the law of iterated expectations, third- and fourth-order raw moments, defined as $E[vec(\mathbf{x}\mathbf{x}')\mathbf{x}']$ and $E[vec(\mathbf{x}\mathbf{x}')vec(\mathbf{x}\mathbf{x}')']$ respectively, can be readily obtained as convex combinations of the third- and fourth-order raw moments of the K underlying Gaussian components. Subtracting the corresponding moments for a $N(\boldsymbol{\pi}, \boldsymbol{\Omega})$ random vectors yields the third- and fourth-order cumulants.

[4] These recursions had been proposed by several authors without appealing to the EM principle. For example, Hassenblad (1966) shows that they coincide with the steepest descent recursions, which confirms that they always lead to improvements in the log-likelihood function (see also Wolfe (1970) and Peters and Walker (1978)).

Proposition 1 *The (pseudo) maximum likelihood estimators of the unconditional mean vector (2) and covariance matrix (3) of the discrete unrestricted mixture of K multivariate normals in (1) are given by the sample mean vector and covariance matrix (with denominator T) of $\mathbf{x}_t$, respectively.*

As a result, if we reparametrise the model as $\mathbf{x}_t = \boldsymbol{\pi} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*$, where $\boldsymbol{\varepsilon}_t^*$ is a standardised discrete mixture of normals, then we can maximise the log-likelihood function with respect to $\boldsymbol{\lambda}$ and the free elements of this distribution keeping $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\Omega}}$ fixed at their Gaussian pseudo ML values.[5] Appendix C.2 first explains how to parametrise the distribution of $\boldsymbol{\varepsilon}_t^*$ so as to ensure that $E(\boldsymbol{\varepsilon}_t^*) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*) = \mathbf{I}_N$ when $K = 2$ as a function of N mean difference parameters $\boldsymbol{\delta}$, $N(N+1)/2$ relative variance parameters $\mathbf{K}$ and a single probability parameter λ, and then generalises this procedure for any K.

Given that Proposition 1 is a numerical result that holds for any sample size T and does not depend in any way on the true distribution of the data, the discrete mixture of normals Pseudo ML estimators of $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$ will continue to be consistent for $E(\mathbf{x})$ and $V(\mathbf{x})$ under distributional misspecification.

2.2 Scale mixtures and their ML estimators

Given that they are rather popular in empirical research, for completeness we also analyse scale mixtures of normals, which as we will see below, inherit the consistency properties of general mixtures under distributional misspecifications that preserve ellipticity.

The random vector $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\sqrt{\varsigma}\mathbf{u}$, where $\mathbf{u}$ is uniform on the unit sphere surface in $\mathbb{R}^N$, is distributed as a discrete scale mixture of normals (DSMN) if

$$\varsigma = \textstyle\sum_{k=1}^{K} s_k \kappa_k^{1/2} \varsigma_i^o, \tag{6}$$

where $\varsigma^o|\mathbf{s}$ is $\boldsymbol{\chi}_N^2$. This is a special case of (1) in which $\boldsymbol{\mu}_k = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_k = \kappa_k\boldsymbol{\Sigma}$ $\forall k$. Therefore, its unconditional mean is $\boldsymbol{\mu}$ while its unconditional variance will be

$$\begin{aligned} \boldsymbol{\Omega} &= V(\mathbf{x}) = \varpi\boldsymbol{\Sigma} = E[V(\mathbf{x}|\mathbf{s})], \\ \varpi &= E(\varsigma/N) = \textstyle\sum_{k=1}^{K} \lambda_k\kappa_k. \end{aligned} \tag{7}$$

As a result, we can easily standardise $\mathbf{x}$ by assuming that $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}_N$ and defining the relative variance parameters

$$\kappa_k^* = \kappa_k/\varpi, \ k = 1, \ldots, K.$$

DSMNs with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_N$ are a particular case of spherically symmetric random

[5] Interestingly, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only $\mathbf{x}_t$ but also $\mathbf{s}_t$.

vectors. Therefore, all their odd central moments will be 0, while their fourth-order moments, which exceed those of the multivariate normal, depend on a single parameter known as the multivariate coefficient of excess kurtosis, which is given by $E(\varsigma^2)/N(N+1)-1$. DSMNs approach the multivariate normal when $\kappa_k^* \to 1$ for all k, or when any $\lambda_k \to 1$. Near the limit, though, the distributions can be radically different. For instance, given that we can choose $\kappa_2/\kappa_1 \in (0,1]$ when $K=2$ without loss of generality, when $\lambda \to 0^+$ there are very few observations with very large variance ("outliers case"), while when $\lambda \to 1^-$ the opposite happens, very few observations with very small variance ("inliers case") (see Amengual and Sentana (2011) for further details).

It is also possible to apply the EM algorithm to DSMNs but the recursions are different. Specifically, they become:

$$\hat{\boldsymbol{\mu}}^{(n} = \frac{\sum_{t=1}^T w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})(\kappa_k^{(n})^{-1}\mathbf{x}_t}{\sum_{t=1}^T \sum_{j=1}^K w_j(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})(\kappa_j^{(n})^{-1}}, \tag{8a}$$

$$\hat{\boldsymbol{\Sigma}}^{(n} = \frac{\sum_{t=1}^T w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})(\kappa_k^{(n})^{-1}(\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n})(\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n})'}{\sum_{t=1}^T \sum_{j=1}^K w_j(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})(\kappa_j^{(n})^{-1}}, \tag{8b}$$

$$\kappa_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}}\frac{1}{TN}\sum_{t=1}^T w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})(\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n})'(\hat{\boldsymbol{\Sigma}}^{(n})^{-1}(\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n}) \tag{8c}$$

with $\hat{\lambda}_k^{(n}$ and $w_k(\mathbf{x}_t;\boldsymbol{\varphi})$ still given by (4c) and (5), respectively.[6]

But if we keep $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ fixed, then the recursions for the $\lambda's$ and $\kappa's$ simplify considerably. To understand why, it is convenient to work with the log-likelihood function of ς, which is a discrete mixture of K gamma random variables with common shape parameter $N/2$ and scale parameters $2\kappa_k$, so that their means are $N\kappa_k$.

Let

$$h_\varsigma(\varsigma_t;\boldsymbol{\eta}) = \frac{\varsigma_t^{N/2-1}}{2^{N/2}\Gamma(N/2)}\sum_{k=1}^K \lambda_k \kappa_k^{-N/2}\exp(-.5\kappa_k^{-1}\varsigma)$$

denote the marginal density of ς, where $\boldsymbol{\eta}$ contains the free elements of $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$. In this context, the EM recursions are given by

$$\kappa_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}}\frac{1}{NT}\sum_{t=1}^T w_k(\boldsymbol{\varsigma}_t;\boldsymbol{\eta}^{(n-1)})\boldsymbol{\varsigma}_t, \tag{9a}$$

$$\hat{\lambda}_k^{(n} = \frac{1}{T}\sum_{t=1}^T w_k(\boldsymbol{\varsigma}_t;\boldsymbol{\eta}^{(n-1)}) \tag{9b}$$

[6] Some overall scale normalisation is obviously required. For example, we could fix one κ_1 to 1, work with the relative variance parameters κ_k^* subject to the restriction $\sum_{k=1}^K \lambda_k\kappa_j^* = 1$ or fix $|\boldsymbol{\Omega}| = 1$, as explained in appendix B of Fiorentini and Sentana (2019). In the first case, the recursions (8a)-(8c) continue to be valid after excluding the relevant element. Given the invariance properties of ML estimators, we recommend the first normalisation, which can be changed after convergence has been achieved.

where

$$w_k(\varsigma_t;\boldsymbol{\varphi}) = P(s_{kt}=1|\varsigma_t) = \frac{\lambda_k \kappa_k^{-N/2}\exp(-.5\kappa_k^{-1}\varsigma)}{\sum_{j=1}^{K}\lambda_j\kappa_j^{-N/2}\exp(-.5\kappa_j^{-1}\varsigma)} \tag{10}$$

is the posterior probability that observation t comes from the k^{th} component. Not surprisingly, (9a) and (10) coincide with (8c) and (5), respectively, when

$$\varsigma_t(\hat{\boldsymbol{\mu}}^{(n},\hat{\boldsymbol{\sigma}}^{(n}) = (\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n})'(\hat{\boldsymbol{\Sigma}}^{(n})^{-1}(\mathbf{x}_t-\hat{\boldsymbol{\mu}}^{(n}).$$

The following proposition, which is the counterpart to Proposition 1, will also prove fundamental for our consistency results in the spherically symmetric case:

Proposition 2 *The (pseudo) maximum likelihood estimators of the unconditional mean (7) of the discrete unrestricted mixture of K gammas with common shape parameter $N/2$ and scale parameters $2\kappa_k$ in (6) is given by the sample mean of ς_t.*

Given that Proposition 2 is a numerical result that holds for any sample size T and does not depend in any way on the true distribution of the data,[7] the discrete scale mixture of normals Pseudo ML estimator of ϖ will continue to be consistent for $E(\varsigma/N)$ under distributional misspecification for any spherically symmetric distribution.

3 Multivariate dynamic regression models with time-varying variances and covariances

3.1 Model specification

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N observed variables, $\mathbf{y}_t$, is typically assumed to be generated as:

$$\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*,$$

where $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(I_{t-1};\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(I_{t-1};\boldsymbol{\theta})$, $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are $N\times 1$ and $N(N+1)/2\times 1$ vector functions describing the conditional mean vector and covariance matrix known up to the $p\times 1$ vector of parameters $\boldsymbol{\theta}$, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and possibly some contemporaneous conditioning variables, and $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. To focus on the effect of distributional misspecification, we maintain the assumption that the conditional mean and variance are correctly specified, in the sense that there is a true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, such that $E(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and $V(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$. We also maintain the high level regularity conditions in Bollerslev and Wooldridge (1992) because we want to leave unspecified

[7] Once again, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only ς_t but also $\mathbf{s}_t$.

the conditional mean vector and covariance matrix in order to achieve full generality. Primitive conditions for specific multivariate models can be found for example in Ling and McAleer (2003).

To complete the model, a researcher needs to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. For the sake of generality, we initially consider a situation in which she makes the assumption that, conditional on I_{t-1}, the distribution of $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed with mean vector equal to 0 and covariance matrix equal to the identity. Nevertheless, we will obtain stronger results below by assuming that either the components of $\boldsymbol{\varepsilon}_t^*$ are cross-sectionally independent, or this vector follows some particular member of the spherical family with a well defined density, or $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\eta} \sim i.i.d.\ s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ for short, where $\boldsymbol{\eta}$ denotes q additional shape parameters which effectively characterise the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. As is well known, spherical symmetry reduces to ordinary symmetry in the univariate case ($N = 1$).

In the first case, we follow Fiorentini and Sentana (2019) in assuming that it is possible to rewrite the model in this form:

Reparametrisation 1 *A homeomorphic transformation* $\mathbf{r}_g(.) = [\mathbf{r}'_{gc}(.),\mathbf{r}'_{gim}(.),\mathbf{r}'_{gic}(.)]'$ *of the mean-variance parameters* $\boldsymbol{\theta}$ *into an alternative set* $\boldsymbol{\phi} = (\boldsymbol{\phi}'_c,\boldsymbol{\phi}'_{im},\boldsymbol{\phi}'_{ic},)'$, *where* $\boldsymbol{\phi}_{im}$ *is* $N \times 1$, $\boldsymbol{\phi}_{ic} = vech(\boldsymbol{\Phi}_{ic})$, $\boldsymbol{\Phi}_{ic}$ *is an unrestricted positive definite symmetric matrix of order* N *and* $\mathbf{r}_g(\boldsymbol{\theta})$ *is twice continuously differentiable in a neighbourhood of* $\boldsymbol{\theta}_0$ *with* $\text{rank}\left[\partial\mathbf{r}'_g\left(\boldsymbol{\theta}_0\right)/\partial\boldsymbol{\theta}\right] = p$, *such that*

$$\left.\begin{array}{r}\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\phi}_c) + \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\phi}_c)\boldsymbol{\phi}_{im} \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\phi}_c)\boldsymbol{\Phi}_{ic}\boldsymbol{\Sigma}_t^{\diamond 1/2\prime}(\boldsymbol{\phi}_c)\end{array}\right\} \quad \forall t. \tag{11}$$

This parametrisations simply requires the pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\phi}_c) = \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\phi}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\phi}_c)] \tag{12}$$

to be *i.i.d.* with mean vector $\boldsymbol{\phi}_{im}$ and covariance matrix $\boldsymbol{\Phi}_{ic}$.

In the spherically case, in contrast, we are able to consider the existence of a less restricted reparametrisation.

Reparametrisation 2 *A homeomorphic transformation* $\mathbf{r}_s(.) = [\mathbf{r}'_{sc}(.),r'_{si}(.)]'$ *of the mean-variance parameters* $\boldsymbol{\theta}$ *into an alternative set* $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_c,\vartheta'_i)'$, *where* ϑ_i *is a positive scalar, and* $\mathbf{r}_s(\boldsymbol{\theta})$ *is twice continuously differentiable with* $\text{rank}\left[\partial\mathbf{r}'_s\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}\right] = p$ *in a neighbourhood of* $\boldsymbol{\theta}_0$, *such that*

$$\left.\begin{array}{r}\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_i\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c)\end{array}\right\} \quad \forall t. \tag{13}$$

Expression (13) simply requires that one can construct pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\circ}(\boldsymbol{\vartheta}_c) = \boldsymbol{\Sigma}_t^{\circ -1/2}(\boldsymbol{\vartheta}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\circ}(\boldsymbol{\vartheta}_c)] \tag{14}$$

which are *i.i.d.* $s(\mathbf{0},\vartheta_i\mathbf{I}_N,\boldsymbol{\eta})$, where ϑ_i is a global scale parameter, a condition satisfied by most static and dynamic models.

3.2 Consistency of discrete mixtures of normals ML estimators

Proposition 1 in Fiorentini and Sentana (2019) states that if (13) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\upsilon}_0$, is *i.i.d.* $s(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\upsilon}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution

assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the joint ML estimator of $\boldsymbol{\varphi} = (\boldsymbol{\vartheta}_c', \vartheta_i, \boldsymbol{\eta})'$, $\boldsymbol{\varphi}_\infty$, is such that $\boldsymbol{\vartheta}_{c\infty}$ is equal to the true value $\boldsymbol{\vartheta}_{c0}$. In this context, in Fiorentini and Sentana (2007) we proposed to estimate ϑ_i by $\vartheta_{iT}(\hat{\boldsymbol{\vartheta}}_{cT})$, where

$$\vartheta_{iT}(\boldsymbol{\vartheta}_c) = \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\varsigma_t^\circ(\boldsymbol{\vartheta}_c). \tag{15}$$

The rationale for this estimator comes from the fact that under normality the score for ϑ_i simplifies to:

$$\mathbf{s}_{\vartheta_i t}(\boldsymbol{\vartheta}, \mathbf{0}) = \frac{1}{2\vartheta_i}\left[\varsigma_t(\boldsymbol{\vartheta}) - N\right], \tag{16}$$

whose expected value when evaluated at $\boldsymbol{\vartheta}_0$ is 0 because the expected value of $\varsigma_t^\circ(\boldsymbol{\vartheta}_{c0}) = \boldsymbol{\varepsilon}_t^{\circ\prime}(\boldsymbol{\vartheta}_c)\boldsymbol{\varepsilon}_t^\circ(\boldsymbol{\vartheta}_c)$ in (14) is precisely $N\vartheta_{i0}$.

However, it turns out that Proposition 2 above implies that (15) numerically coincides the MLE of ϑ_i when the assumed spherical distribution is a discrete scale mixture of normals, so it is irrelevant whether we replace it or not. As a result, the ML estimators based on a discrete scale mixture of normals are consistent for all the parameters when the true distribution is spherical.

In turn, Proposition 2 in Fiorentini and Sentana (2019) states that if (11) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\upsilon}_0$ is *i.i.d.* $(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\upsilon}$ includes $\boldsymbol{\phi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the joint ML estimator of $\boldsymbol{\varphi} = (\boldsymbol{\phi}_c', \boldsymbol{\phi}_i', \boldsymbol{\varrho})'$, $\boldsymbol{\varphi}_\infty$, is such that $\boldsymbol{\phi}_{c\infty}$ is equal to the true value $\boldsymbol{\phi}_{c0}$. In this context, in Fiorentini and Sentana (2007) we proposed to estimate $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ as $\boldsymbol{\phi}_{imT}(\hat{\boldsymbol{\phi}}_{cT})$ and $\boldsymbol{\phi}_{icT}(\hat{\boldsymbol{\phi}}_{cT})$, respectively, where

$$\boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\phi}_c), \tag{17}$$

$$\boldsymbol{\phi}_{icT}(\boldsymbol{\phi}_c) = vech\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\phi}_c) - \boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c)\right]\left[\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\phi}_c) - \boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c)\right]'\right\}. \tag{18}$$

Once again, the rationale for these estimators comes from the fact that under normality the scores for $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ simplify to:

$$\begin{aligned}\mathbf{s}_{\psi_{im}t}(\boldsymbol{\phi}, \mathbf{0}) &= \frac{1}{2}\boldsymbol{\Phi}_{ic}^{-1/2\prime}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}),\\ \mathbf{s}_{\psi_{ic}t}(\boldsymbol{\phi}, \mathbf{0}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Phi}_{ic}^{-1/2\prime}\otimes\boldsymbol{\Phi}_{ic}^{-1/2\prime})vec\left\{\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\phi}) - \mathbf{I}_N\right\},\end{aligned}$$

where $\mathbf{D}_N$ is the duplication matrix (see Magnus and Neudecker (2019)), whose expected values at $\boldsymbol{\phi}_0$ are 0 because the expected value of

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}_{c0}, \boldsymbol{\phi}_i) = \boldsymbol{\Phi}_{ic}^{-1/2}(\boldsymbol{\phi}_{im0} - \boldsymbol{\phi}_{im}) + \boldsymbol{\Phi}_{ic}^{-1/2}\boldsymbol{\Phi}_{ic0}^{1/2}\boldsymbol{\varepsilon}_t^*$$

is 0 and the expected value of $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}_{c0}, \boldsymbol{\phi}_i)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\phi}_{c0}, \boldsymbol{\phi}_i)$ is $\mathbf{I}_N$ when $\boldsymbol{\phi}_i = \boldsymbol{\phi}_{i0}$.

However, it turns out that Proposition 1 above implies that (17) and (18) numerically coincide with the MLEs of $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ when the assumed distribution is an unrestricted discrete mixture of normals, so once again, it is irrelevant whether we replace them or not. As a result, the ML estimators based on an unrestricted discrete mixture of normals are consistent for all the parameters regardless of the true distribution

4 Application to structural vector autoregressions

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

$$\mathbf{y}_t = \boldsymbol{\tau} + \textstyle\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \mathbf{C}\boldsymbol{\varepsilon}_t^*, \quad \boldsymbol{\varepsilon}_t^* | I_{t-1} \sim i.i.d.\ (\mathbf{0}, \mathbf{I}_N), \tag{19}$$

where $\mathbf{C}$ is a matrix of impact multipliers and $\boldsymbol{\varepsilon}_t^*$ are "structural" shocks. In what follows, we will often reparametrise $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix whose elements contain the scale of the structural shocks, while the columns of $\mathbf{J}$, whose diagonal elements are normalised to 1, measure the relative impact effects of each of the structural shocks on all the remaining variables, so that the parameters of interest become $\mathbf{j} = veco(\mathbf{J} - \mathbf{I}_N)$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$.[8] Similarly, the drift $\boldsymbol{\tau}$ is often written as $(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)\boldsymbol{\mu}$ under the assumption of covariance stationarity, where $\boldsymbol{\mu}$ is the unconditional mean of the observed process.

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 $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}' = \mathbf{J}\boldsymbol{\Psi}^2\mathbf{J}'$. 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 so-called LQ matrix decomposition[9] to relate the matrix $\mathbf{C}$ to the Cholesky decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_L\boldsymbol{\Sigma}_L'$ as $\mathbf{C} = \boldsymbol{\Sigma}_L\mathbf{Q}$, 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 $|\mathbf{Q}| = 1$.[10,11] 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.[12]

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

[9] The LQ decomposition is intimately related to the QR decomposition. Specifically, $\mathbf{Q}'\boldsymbol{\Sigma}_L'$ provides the QR decomposition of the matrix $\mathbf{C}'$,which is uniquely defined if we restrict the diagonal elements of $\boldsymbol{\Sigma}_L$ to be positive (see e.g. Golub and van Loan (1993) for further details).

[10] See section 9 of Magnus, Pijls and Sentana (2020) for a detailed discussion of three ways of explicitly parametrising a rotation (or special orthogonal) matrix: (i) as the product of Givens matrices that depend on $N(N-1)/2$ Tait-Bryan angles, one for each of the strict upper diagonal elements; (ii) by using the so-called Cayley transform of a skew-symmetric matrix; and (c) by exponentiating a skew-symmetric matrix.

[11] If $|\mathbf{Q}| = -1$ instead, we can change the sign of the i^{th} structural shock and its impact multipliers in the i^{th} column of the matrix $\mathbf{C}$ without loss of generality as long as we also modify the shape parameters of the distribution of ε_{it}^* to alter the sign of all its non-zero odd moments.

[12] The identifying assumption of Proposition 1 in Lanne and Lütkepohl (2010) explicitly rules out scale mixtures

Nevertheless, Lanne, Meitz and Saikkonen (2017) show that statistical identification of both the structural shocks and $\mathbf{C}$ (up to column permutations and sign changes) is possible assuming (i) cross-sectional independence of the N shocks and (ii) a non-Gaussian distribution for at least $N-1$ of them. In what follows, we assume that the N structural shocks are cross-sectionally independent. In addition to finite normal mixtures, a particularly important example is $\varepsilon_{it}^*|I_{t-1} \sim i.i.d.\ t(0,1,\nu_i)$. Univariate t distributions are very popular in finance as a way of capturing fat tails while nesting the traditional Gaussian assumption, and their popularity is also on the rise in macroeconomics, as illustrated by Brunnermeier et al (2019). Other possibilities are Generalised Error distributions (GED), which nest both the normal and the Laplace (or double exponential).

Let $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\mathbf{A}_1), \ldots, vec'(\mathbf{A}_p), vec'(\mathbf{C})]' = (\boldsymbol{\tau}', \mathbf{a}_1', \ldots, \mathbf{a}_p', \mathbf{c}') = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$ denote the structural parameters characterising the first two conditional moments of $\mathbf{y}_t$. In addition, let $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_N)'$ denote the shape parameters, so that $\boldsymbol{\varphi} = (\boldsymbol{\theta}', \boldsymbol{\varrho}')'$.

Given the linear mapping between structural and reduced form shocks, the contribution to the conditional log-likelihood function from observation t ($t = 1, \ldots, T$) will be given by

$$l_t(\mathbf{y}_t; \boldsymbol{\varphi}) = -\ln|\mathbf{C}| + l[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i] + \ldots + l[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N], \tag{20}$$

where $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \mathbf{C}^{-1}(\mathbf{y}_t - \boldsymbol{\tau} - \mathbf{A}_1\mathbf{y}_{t-1} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p})$. We consider two ML estimators: a restricted one which fixes $\boldsymbol{\varrho}$ to its supposedly true value, and an unrestricted one, which simultaneously estimates all the elements of $\boldsymbol{\varphi}$ (see Fiorentini and Sentana (2020b) for further details).

It turns out that both the slope parameters $\mathbf{a}$ and the (scaled) impact multiplier coefficients $\mathbf{j}$ will continue to be consistently estimated by distributional misspecified ML estimators. More formally,

Proposition 3 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (19) is the product of N univariate densities but they are different from the ones assumed for estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (19) remain consistent for $\mathbf{a}$ and $\mathbf{j}$.*

Intuitively, the pseudo-standardised residuals $\mathbf{J}_0^{-1}(\mathbf{y}_t - \mathbf{A}_{10}\mathbf{y}_{t-1} - \ldots - \mathbf{A}_{p0}\mathbf{y}_{t-p})$ remain time series and cross-sectionally *i.i.d.* with means $\mathbf{J}_0^{-1}\boldsymbol{\tau}_0$ and covariance matrix $\boldsymbol{\Psi}_0^2$ under distributional misspecification, so in effect, the pseudo true values of $\boldsymbol{\tau}$ mop up the biases in the means of those residuals while the pseudo true values of $\boldsymbol{\psi}$ do the same for their standard deviations. This intuition also justifies that the consistent estimators in Fiorentini and Sentana (2019), which replace the non-Gaussian pseudo ML estimators of τ_i and ψ_i by the sample mean and variance of the i^{th} pseudo-standardised residual, will work in this context too.

of normals.

Proposition 3 also illustrates the practical consequences of distributional misspecification. Given that the IRFs of the structural VAR model in (19) under stationarity will be given by $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)^{-1}\mathbf{J}\boldsymbol{\Psi}$, where L is the usual lag operator, their temporal pattern will be consistent estimated. In contrast, the estimated scale of the IRFs, and the FEVDs will generally be inconsistent.

As we mentioned in the introduction, we can strengthen the consistency results in Proposition 3 by assuming that both the true univariate distributions of the structural shocks and the ones assumed for estimation purposes are symmetric, even though they do not necessarily coincide:

Proposition 4 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (19) is the product of N univariate symmetric densities but they are different from the symmetric ones assumed for estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (19) remain consistent for $\mathbf{a}$, $\mathbf{j}$ and $\boldsymbol{\tau}$.*

It is illustrative to compare Propositions 3 and 4 to the consistency results in Gouriéroux, Monfort and Renne (2017), who work with the alternative reparametrisation $\mathbf{C} = \boldsymbol{\Sigma}_L \mathbf{Q}(\boldsymbol{\omega})$. They show that regardless of the specific distributions assumed for estimation purposes, a non-Gaussian PMLE can usually consistently estimate the $N(N-1)/2$ underlying free elements of $\mathbf{Q}(\boldsymbol{\omega})$ when the true value of $\boldsymbol{\Sigma}_L$ is either known, or replaced by a consistent estimator such as the Gaussian PMLE. In contrast, we show that a non-Gaussian PMLE can consistently estimate the $N(N-1)$ elements of $\mathbf{J}$. In addition, we saw in the previous section that the non-Gaussian PMLE of the diagonal elements of $\boldsymbol{\Psi}$ will be consistently estimated if we assume for estimation purposes that the structural shocks follow univariate mixtures of normals. More generally, the FS estimators, which are effectively Gaussian PMLEs based on pseudo-standardised residuals, will provide consistent estimators of the N elements of $\boldsymbol{\psi}$.

5 Monte Carlo evidence

In this section, we assess the small sample behavior of the different estimators discussed in the previous section by means of an extensive Monte Carlo simulation exercise in which we generate samples from the following three-variate SVAR(1) process

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0.2 & 1 & 0 \\ 0.2 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \\ \varepsilon_{3t}^* \end{bmatrix}.$$

The main aim of the partial interchangeability of this design is to save space in presenting the simulation results by pooling several groups of parameters. Nevertheless, the estimators that we consider are fully unrestricted and do not exploit any of the restrictions resulting from the

fact that the true unconditional means are zero or the true loading matrix of the shocks has a triangular structure.

In accordance with the assumptions in section 4, the error terms $\boldsymbol{\varepsilon}_t^*$ are stochastically independent from each other with zero mean and unit variance. We simulate four different distributions, two of which are symmetric: (i) three heterogeneous univariate Student t distributions; (ii) three Laplace distributions; (iii) three heterogenous discrete location scale mixtures of two normals (DLSMN); and (iv) three heterogenous asymmetric Student ts (see Mencía and Sentana (2009) for details).

For each simulation design, we generate 5,000 samples of length $T = 2,000$ and estimate the model parameters with ten different estimators. In particular, we use: (1) the Student t MLE and (2) the corresponding consistent FS correction, (3) the MLE based on unrestricted two-component Gaussian mixtures (DLSMN), which are consistent regardless of the true distribution, as we have previously shown, (4) the MLE which assumes that the shocks are symmetric scale mixture of two normals (DSMN) and (5) the corresponding consistent FS correction, (6) the Laplace-based MLE and (7) its FS consistent correction. We also compute three versions of the two-step consistent procedure in Gouriéroux, Monfort and Renne (2017) (GMR). As we mentioned at the end of the previous section, in their first step, they estimate the $N + pN^2 + N(N+1)/2$ reduced form parameters $[\boldsymbol{\tau}', vec'(\mathbf{A}_1), \ldots, vec'(\mathbf{A}_p), vech'(\boldsymbol{\Sigma}_L)]'$ by Gaussian PML. Then, they compute the standardised reduced form residuals $\tilde{\mathbf{u}}_t^* = \tilde{\boldsymbol{\Sigma}}_L^{-1}(\mathbf{y}_t - \tilde{\boldsymbol{\tau}} - \sum_{j=1}^p \tilde{\mathbf{A}}_j \mathbf{y}_{t-j})$, on the basis of which they estimate by non-Gaussian PML the $N(N-1)/2$ free elements $\boldsymbol{\omega}$ of the orthogonal rotation matrix $\mathbf{Q}$, which maps structural shocks and reduced form innovations as $\mathbf{u}_t^* = \mathbf{Q}(\boldsymbol{\omega})\boldsymbol{\varepsilon}_t^*$. To level the playing field, in this second step we consider estimators based on the Student t, the DLSMN and the Laplace likelihoods. These three estimators, though, share the first step, so they only differ in the estimated values of $\mathbf{C}$ that they produce. As for the FS corrections, we use a Gaussian PMLE for the N parameters in $\boldsymbol{\psi}$ and the N parameters in $\boldsymbol{\tau}$, except when we use unrestricted finite mixtures of normals to compute our joint non-Gaussian ML estimators, in which case we estimate all the parameters in one go. In all cases, we 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 Table 1 we report the Monte Carlo mean absolute bias for several groups of parameters: the drifts $\boldsymbol{\tau}$, the diagonal elements of the autoregressive matrix $\{\mathbf{A}\}_{ii}$, the off diagonal elements $\{\mathbf{A}\}_{ij,i\neq j}$, the diagonal elements of the impact multiplier matrix $\{\mathbf{C}\}_{ii}$, and its lower and upper diagonal elements $\{\mathbf{C}\}_{ij,i>j}$ and $\{\mathbf{C}\}_{ij,i<j}$, respectively. Finally, we also report the biases of the

lower and upper diagonal elements of $\mathbf{J} = \mathbf{C}\boldsymbol{\Psi}^{-1}$, for which non-Gaussian PMLEs should be consistent according to Propositions 3 and 4.

When the structural shocks follow independent Student t distributions with 6, 12 and 8 degrees of freedom, all estimators are consistent except the Laplace-based MLE of $\mathbf{C}$. As expected, the Student t MLE and the corresponding FS correction dominate the others, but the Mixture-based MLE and the Student t based GMR perform rather well. In turn, when the errors follow independent Laplace distributions, the results are analogous, in that this time the bias appears in the Student t MLE of $\mathbf{C}$ with all other estimators showing extremely low finite sample bias. The third panel of Table 1 displays the results for the simulation with DLSMN shocks. In this case, the biases of Student t MLEs of $\mathbf{C}$ and $\boldsymbol{\tau}$ are large while the biases of the Laplace MLEs are more apparent in the drift estimators. Not surprisingly, the DLSMN MLE is the best but the two consistent FS corrections of the Student and Laplace MLEs are also very good and compare favourably even to the DLSMN version of GMR. Finally, when the error terms follow asymmetric Student t distributions, all estimators are based on misspecified likelihoods. Nevertheless, the last panel of Table 1 indicates that many of them perform rather well in terms of finite sample biases, with the DLSMN MLE being probably the best one.

Next, we evaluate the finite sample relative efficiency of the different consistent estimators using the Monte Carlo root mean squared errors (RMSE) in Table 2 for the same groups of parameters. For the Student t DGP, the Student t MLE is obviously the best but its FS correction also performs very well, and the same is true of the estimators that rely on a finite normal mixture. As for the Student-based GMR estimators, they are clearly inefficient for $\boldsymbol{\tau}$ and $\mathbf{a}$ but fully efficient for the elements of $\mathbf{C}$ because the information matrix is block diagonal between conditional mean and variance parameters (see Proposition B1 in Fiorentini and Sentana (2020b)). In contrast, the estimators that rely on a Laplace likelihood are the worst. Somewhat surprisingly, the Laplace MLEs is more precise for $\mathbf{J}$ than the Laplace GMR estimator even though the asymptotic covariance matrix of this estimator is likely to be block diagonal between the conditional mean and variance parameters in view of the symmetry of the true distribution.

As expected, the second panel of Table 2 confirms that the relative performance of the Student t and Laplace estimator is by and large the mirror image of the first panel. The main difference is that the Laplace-based GMR estimators are noticeably less efficient for $\mathbf{C}$ than the Laplace-based MLEs.

Once more, the MLE based on the correct distribution is the best performer when we simulate DLSMN shocks, but the GMR-DLSMN estimator of the diagonal elements of $\mathbf{C}$ is also very precise. In contrast, this estimator is again suboptimal when we look at the elements of

the autoregressive matrix $\mathbf{A}$ since it relies on the first-step Gaussian PMLE, which is clearly dominated by both the Student and Laplace PMLEs.

Finally, we can see in the last panel of Table 2 that the DLSMN MLE is the best performer in terms of precision when the true shocks follow asymmetric Student ts even though all estimators are based on misspecified likelihood functions. Among the remaining consistent non-Gaussian PMLE estimators, the FS correction to the Laplace MLEs shows more finite sample variability than the others, with the GMR-Laplace being even worse, especially for $\mathbf{J}$ and the off-diagonal elements of $\mathbf{C}$.

In summary, our Monte Carlo exercises confirm the efficiency under distributional misspecification of our proposed ML estimators based on unrestricted discrete mixtures of normals relative to other consistent proposals.

6 Empirical application to volatility indices

We consider three daily series of market-based implied volatilities as measured by the VIX index, the EVZ EuroCurrency ETF volatility index and the GVZ Gold ETF volatility index. The series are compiled by the Chicago Board of Options Exchange (CBOE) and can be freely downloaded from the St. Louis FRED site. They represent three of the most actively traded asset classes, namely stocks, exchange rates and commodities, and since their inception have become incredibly popular among academics, financial market practitioners and commentators. Our sample spans from June 2nd 2008 to September 24th 2020 for a total of 3,101 observations.

Let $\mathbf{x}_t = (x_{VIX,t}, x_{EVZ,t}, x_{GVZ,t})'$ denote the log-transformation of these volatility indexes, which we depict in Figure 1. A preliminary univariate data analysis confirms their high persistence, with a first-order autocorrelation above 0.98 and a slow rate of decay for higher orders. This is hardly surprising, as it is well known that the temporal pattern of volatility indices at the daily frequency shows mean reversion over the long run but persistent deviations from the mean during extended periods. This is confirmed by the fact that ARMA(2,1) models, which correspond to the exact discretisation of the stationary central tendency process in continuous time considered by Mencía and Sentana (2013), provide a good representation for the three series.

Given that our interest is to study the dynamic linkages between these volatility indices, we estimate the following three-variate SVAR(5) model

$$\mathbf{x}_t = \boldsymbol{\tau} + \mathbf{A}_1\mathbf{x}_{t-1} + \ldots + \mathbf{A}_5\mathbf{x}_{t-5} + \mathbf{C}\boldsymbol{\varepsilon}_t^*,$$

where we have selected the lag order by looking at the Akaike information criterion and the

likelihood ratio test for the null hypothesis of lack of residual serial correlation.

We estimate the structural parameters using three of the consistent estimators in the previous section. The first estimator assumes that $\varepsilon_{it}^{*} \sim i.i.d.\ t(0,1,\nu_i)$, where v_i denotes the Student t degrees of freedom parameter, but then we apply the FS correction, which is consistent even if the true shock distributions are asymmetric. In turn, the second estimator assumes that $\varepsilon_{it}^{*} \sim i.i.d.$ $DLSMN(\delta_i, \kappa_i, \lambda_i)$ and estimates all the parameters jointly. Finally, the third estimator employs the GMR two-step strategy with the same unrestricted finite mixture of normals assumption in the second step.

As for initial values, we use standard Gaussian $PMLE$, which is equivalent to running OLS regression for each of the three variables and computing the covariance matrix of the estimated residuals. Thus, we obtain

$$\hat{\boldsymbol{\mu}}_{tFS} = \begin{bmatrix} 2.895 \\ 2.254 \\ 2.880 \end{bmatrix}; \qquad \hat{\boldsymbol{\mu}}_{DLSMN} = \begin{bmatrix} 2.893 \\ 2.253 \\ 2.877 \end{bmatrix}; \qquad \hat{\boldsymbol{\mu}}_{GMR} = \begin{bmatrix} 2.902 \\ 2.265 \\ 2.886 \end{bmatrix}.$$

where $\boldsymbol{\mu} = \boldsymbol{\tau}(\mathbf{I} - \mathbf{A}_1 - \ldots - \mathbf{A}_5)^{-1}$ are the unconditional means. Notice that by construction $\widehat{\boldsymbol{\mu}}_{GMR}$ is numerically the same as the corresponding OLS estimator. As expected from the results in Table 1, the three estimators provide very similar point estimates.

As for the structural impact multipliers matrix, we find that

$$\begin{aligned} \hat{\mathbf{C}}_{tFS} &= \begin{bmatrix} 0.0766 & 0.0074 & 0.0007 \\ 0.0123 & 0.0497 & 0.0033 \\ 0.0210 & 0.0118 & 0.0502 \end{bmatrix}; \\ \hat{\mathbf{C}}_{DLSMN} &= \begin{bmatrix} 0.0769 & 0.0052 & 0.0016 \\ 0.0135 & 0.0493 & 0.0033 \\ 0.0206 & 0.0111 & 0.0505 \end{bmatrix}; \\ \hat{\mathbf{C}}_{GMR} &= \begin{bmatrix} 0.0766 & 0.0064 & 0.0025 \\ 0.0133 & 0.0493 & 0.0034 \\ 0.0207 & 0.0122 & 0.0506 \end{bmatrix}, \end{aligned}$$

which are also rather similar, confirming once again the findings of the Monte Carlo simulation exercise in the previous section.

The estimated structural shocks are shown in Figure 2. Reassuringly, they appear to be serially $i.i.d.$ but highly non-normal. To help with the interpretation of the structural shocks, it is convenient to look not only at the estimated values of $\mathbf{C}$ but also at those of its inverse, which expresses the structural shocks $\boldsymbol{\varepsilon}_t^{*}$ as linear combinations of the reduced form prediction errors $\mathbf{u}_t$. Given that both $\mathbf{C}$ and $\mathbf{C}^{-1}$ are almost lower triangular matrices despite the fact that they are freely estimated, we can label the first shock as a stock volatility shock. Similarly, we will refer to the second and third shocks as FX and Gold volatility shocks, respectively, in view of

the largely recursive nature of the estimated structural model.

Figures 3 displays the IRFs and FEVDs up to one-year ahead. The strong persistence implied by the SVAR(5) parameter estimates implies that all the IRFs decay rather slowly. The responses of VIX to both FX and Gold volatility shocks are hump shaped but small in magnitude. The volatility of the $/euro exchange rate seems to react mostly to its own shock, while Gold volatility is mostly affected by the other shocks and, in particular, by the FX one.

A convenient way of summarising the information in the FEVD plots is to compute the connectedness measures proposed by Diebold and Yilmaz (2014). Importantly, given that the we have identified and consistently estimated the matrix of impact multipliers $\mathbf{C}$ and the autoregressive matrices $\mathbf{A}_i$ $(i = 1, \ldots, 5)$, we can compute those measures without having to resort to the generalised FEVDs of Pesaran and Shin (1996).

Using the entire sample, we find that the one-year ahead FEVDs yield the following sample connectedness table

	Stock	FX	Gold
VIX	0.659	0.227	0.114
EVZ	0.024	0.931	0.045
GVZ	0.104	0.343	0.553

As can be seen, the historical total connectedness of the three volatility series, defined as the sum of the off-diagonal elements of this table divided by N, takes the value of 0.286, which is not very high if we take into account that the elements of each row add up to 1.

"From" connectedness, which we compute by summing the off-diagonal elements of the rows in the previous table, is

VIX	0.341
EVZ	0.070 ,
GVX	0.447

which, somewhat surprisingly, is very low for the EuroCurrency volatility index but moderately high for Gold volatility, most of which being due to FX volatility shocks.

Similarly, "To" connectedness, which is the sum of the off-diagonal column elements, yields

Stock	0.129
FX	0.570 ,
Gold	0.159

being high for the FX shock but moderately low for the other two.

In summary, we find an approximate recursive structure for the impact multipliers, which combined with the estimates of the autoregressive matrices implies that FX volatility shocks explain a non-negligible fraction of the forecast error variation of the VIX and especially the GVX index. In contrast, the converse is not true, as most of the forecast error variation in the EVZ index is explained by its own shocks.

7 Conclusions and directions for future research

We prove that maximum likelihood estimation of structural vector autoregressions with independent non-Gaussian shocks generates consistent estimators of the autoregressive coefficients and (scaled) impact multipliers under distributional misspecification, which in turn implies consistent estimation of the temporal pattern of the IRFs. In contrast, the drifts and standard deviations of the shocks are generally inconsistently estimated, and so are the FEVDs. Nevertheless, we show consistency when the non-Gaussian log-likelihood is a discrete scale mixture of normals in the symmetric case, or an unrestricted finite mixture more generally. We also confirm the validity of the consistent estimators in Fiorentini and Sentana (2019) when the shocks are assumed to follow other non-Gaussian distributions such as the Student t or the Laplace. Our detailed Monte Carlo exercises illustrate the efficiency under distributional misspecification of our proposed estimators relative to other consistent proposals.

Finally, we study the dynamic linkages between the popular volatility indices for the S&P500, the US \$/euro exchange rate and gold. Somewhat surprisingly, we find that the matrix of impact multipliers is close to lower triangular, which suggests that the structural volatility shocks that we estimate correspond to stocks, foreign exchange and gold. We also find that the historical total connectedness at the one-year ahead horizon is not very high, and that the FX volatility shocks explain a non-negligible fraction of the forecast error variation of the VIX and especially the GVX index.

A worthwhile extension of our paper would look at semiparametric estimators which use our proposed consistent estimators as initial values for a single BHHH iteration based on the semiparametric efficient score. The results in Fiorentini and Sentana (2020b) suggest that such estimators would be (partially) adaptive for the matrices of VAR coefficients $\mathbf{A}_j$ $(j = 1, \ldots, p)$ and the scaled impact multipliers in $\mathbf{J}$. Given that the structural shocks are assumed cross-sectionally independent, their densities would be estimated at univariate non-parametric rates. An alternative procedure that should in principle achieve the semiparametric efficiency bound would be a sieves-type estimator that uses discrete mixtures of normals in which the number of underlying components increases with the sample size at a suitable rate. A comparison of these two estimators with the distribution-free methods in Herwartz (2018), who exploits the proposal in Matteson and Tsay (2017), and Lanne and Luoto (2019), who employ a GMM estimator that replaces the assumption of independent shocks with analogous restrictions on a finite number of high-order cross-cumulants, would also be valuable.

Given that our theoretical results are valid not only for SVAR models with cross-sectional independent structural shocks, but also for many dynamic conditionally heteroskedastic multi-

variate regression models routinely used in empirical finance and other fields, including ARCH-M models and multivariate regressions, it would be interesting to assess the performance of discrete mixture of normals maximum likelihood estimators in those contexts. The study of the effects on our proposed estimators of structural shocks which are not serially independent because of the presence of time-varying volatility would also be worth pursuing. Finally, the empirical credibility of the identification approach that we have exploited would be enhanced if our proposed estimators would be complemented by specification tests that confirm the assumption of cross-sectionally independent shocks. We are currently exploring some of these interesting research avenues.

References

Amengual, D. and Sentana (2011): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Barigozzi, M. and Brownlees, C. (2019): "NETS: Network estimation for time series", *Journal of Applied Econometrics* 34, 347–364.

Behboodian, J. (1970): "On a mixture of normal distributions", *Biometrika* 57, 215-217.

Bekaert, G., Engstrom, E. and Ermolov, A. (2019): "Macro risks and the term structure of interest rates", Working paper, forthcoming in the *Journal of Financial Economics*.

Bekaert, G., Engstrom, E. and Ermolov, A. (2020): "Aggregate demand and aggregate supply effects of COVID-19: a real-time analysis", forthcoming in *Covid Economics*.

Bernoth, K. and Herwartz; H. (2019): "Exchange rates, foreign currency exposure and sovereign risk", DIW Discussion Paper 1792.

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

Bollerslev, T., and Wooldridge, J. M. (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.

Brunnermeier, M., Palia, D., Sastry, K.A. and Sims, C.A. (2019): "Feedbacks: financial markets and economic activity", mimeo, Princeton University.

Capasso, M. and Moneta, A. (2016): "Macroeconomic responses to an independent monetary policy shock: a (more) agnostic identification procedure", Sant'Anna School of Advanced Studies LEM Papers Series 2016/36.

Coad, A. and Grassano, N. (2019): "Firm growth and R&D investment: SVAR evidence from the world's top R&D investors", *Industry and Innovation* 26, 508-533.

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

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

Diebold, F.X. and Yılmaz, K. (2014): "On the network topology of variance decompositions: measuring the connectedness of financial firms", *Journal of Econometrics* 182, 119–134.

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

Fiorentini, G. and Sentana, E. (2919): "Consistent non-Gaussian pseudo maximum likelihood estimators", *Journal of Econometrics* 213, 321-358.

Fiorentini, G., and Sentana, E. (2020a): "New testing approaches for mean-variance predictability", forthcoming in the *Journal of Econometrics*.

Fiorentini, G., and Sentana, E. (2020b): "Specification tests for non-Gaussian maximum likelihood estimators", 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.

Gouriéroux, C., Monfort, A. and Renne, J.-P. (2020), "Identification and estimation in non-fundamental structural VARMA models", *Review of Economic Studies* 87, 1915-1953.

Guay, A. and Normandin, M. (2018): "Identification of structural vector autoregressions through higher unconditional moments", mimeo HEC Montreal.

Ha, J. and Lee, T. (2011): "NM–QELE for ARMA-GARCH models with non-Gaussian innovations", *Statistics and Probability Letters* 81, 694–703.

Hamdan, H. (2006): "Characterizing and approximating infinite scale mixtures of normals", *Communications in Statistics - Theory and Methods* 35, 407-413.

Hasselblad, V. (1966): "Estimation of parameters for a mixture of normal distributions", *Technometrics* 8, 431-444.

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

Herwartz, H. (2019): "Long-run neutrality of demand shocks: revisiting Blanchard and Quah (1989) with independent structural shocks", *Journal of Applied Econometrics* 34, 811-819.

Herwartz, H. & Plödt, M. (2016): "The macroeconomic effects of oil price shocks: evidence from a statistical identification approach", *Journal of International Money and Finance* 61, 30–44.

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

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

Kiefer, N.M. (1978): "Discrete parameter variation: efficient estimation of a switching regression model", *Econometrica* 46, 427-434.

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

Lanne, M. and Lütkephl, H. (2010): "Structural vector autoregressions with nonnormal

residuals", *Journal of Business and Economic Statistics* 28, 159-168.

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

Lee, T. and Lee, S. (2009): "Normal mixture quasi-maximum likelihood estimator for GARCH models", *Scandinavian Journal of Statistics* 36, 157–170.

Ling, S. and McAleer, M. (2003): "Asymptotic theory for a vector ARMA-GARCH model", *Econometric Theory* 19, 280-310.

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

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

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

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

Maxand, S. (2020): "Identification of independent structural shocks in the presence of multiple Gaussian components", *Econometrics and Statistics* 16, 55-68.

Mencía, J. and Sentana, E. (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105-121.

Mencía, J. and Sentana, E. (2013): "Valuation of VIX derivatives", *Journal of Financial Economics* 108, 367-391.

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

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

Nguyen, T.T., Nguyen, H.D., Chamroukhi, F. and McLachlan, G.J. (2019): "Approximation by finite mixtures of continuous density functions that vanish at infinity", available from https://arxiv.org/abs/1903.00147

Peters, B.C. and Walker, H.F. (1978): "An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions", *SIAM Journal of Applied Mathematics* 35, 362-378.

Pesaran, M.H. and Shin, Y. (1996): "Generalized impulse response analysis in linear multivariate models", *Economics Letters* 58, 17-29.

Puonti, P. (2019): "Data-driven structural BVAR analysis of unconventional monetary policy", *Journal of Macroeconomics* 61, 103131.

Sentana, E. (1999): "Econometric applications of positive rank-one modifications of the symmetric factorization of a positive semi-definite matrix", *Spanish Economic Review* 1, 79-90.

Tank, A., Fox, E.B. amd Shojaie, A. (2019): "Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series", *Biometrika* 106, 433-452.

Wolfe, J.H. (1970): "Pattern clustering by multivariate mixture analysis", *Multivariate Behavioral Research* 5, 329-350.

Appendices

A Proofs

A.1 Proposition 1

It is easy to check that the EM recursions (4a)-(4c) imply that

$$\begin{aligned}
\hat{\boldsymbol{\pi}}^{(n} &= \textstyle\sum_{k=1}^{K} \hat{\boldsymbol{\mu}}_k^{(n} \hat{\lambda}_k^{(n} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t, \\
\hat{\boldsymbol{\Omega}}^{(n} &= \textstyle\sum_{k=1}^{K} (\hat{\boldsymbol{\mu}}_k^{(n} \hat{\boldsymbol{\mu}}_k^{(n\prime} + \hat{\boldsymbol{\Sigma}}_k^{(n}) \hat{\lambda}_k^{(n} - \hat{\boldsymbol{\pi}}^{(n} \hat{\boldsymbol{\pi}}^{(n\prime} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t - \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\right)',
\end{aligned}$$

for all T regardless of the values of $\boldsymbol{\varphi}^{(n-1}$. Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}^{(\infty)}$), it follows that $\hat{\boldsymbol{\pi}}^{(n}$ and $\hat{\boldsymbol{\Omega}}^{(n}$ coincide with the Gaussian PML estimators. □

A.2 Proposition 2

It is easy to check that the EM recursions (9a)-(9b) imply that

$$\hat{\varpi}^{(n} = \textstyle\sum_{k=1}^{K} \hat{\kappa}_k^{(n} \hat{\lambda}_k^{(n} = \frac{1}{T}\sum_{t=1}^{T} \varsigma_t,$$

for all T regardless of the values of $\boldsymbol{\eta}^{(n-1}$. Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^{(\infty)}$), it follows that $\hat{\varpi}$ coincides with the sample mean of ς_t. □

A.3 Proposition 3

Proposition 12 in the supplemental appendix of Fiorentini and Sentana (2019) implies that the misspecification of the conditional distribution of the structural shocks ε^* will not affect the consistency of either the restricted or unrestricted MLEs of the elements of $\mathbf{a}$. To prove the consistency of $\mathbf{j}$, it is convenient to study the scores with respect to the different parameters, which are given by

$$\begin{aligned}
s_{\boldsymbol{\tau}}(\boldsymbol{\theta};\underline{\boldsymbol{\varrho}}) &= -\mathbf{C}^{-1\prime}\frac{\partial f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}]}{\partial \boldsymbol{\varepsilon}^*} = -\mathbf{C}^{-1\prime}\left\{\begin{array}{c} -\frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}_1]}{\partial \varepsilon^*} \\ \vdots \\ -\frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}_N]}{\partial \varepsilon^*} \end{array}\right\}, \\
s_{\mathbf{a}}(\boldsymbol{\theta};\underline{\boldsymbol{\varrho}}) &= -\left(\begin{array}{c} \mathbf{y}_{t-1}\otimes\mathbf{C}^{-1\prime} \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{C}^{-1\prime} \end{array}\right)\frac{\partial f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}]}{\partial \boldsymbol{\varepsilon}^*} = -\left(\begin{array}{c} \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \end{array}\right)\mathbf{C}^{-1\prime}\left\{\begin{array}{c} \frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}_1]}{\partial \eta_1} \\ \vdots \\ \frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\underline{\boldsymbol{\varrho}}_N]}{\partial \eta_N} \end{array}\right\},
\end{aligned}$$

$$
\begin{aligned}
s_{\mathbf{j}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= \boldsymbol{\Delta}_N'(\boldsymbol{\Psi}\otimes\mathbf{C}^{-1\prime})vec\left[\mathbf{I}_N+\frac{\partial f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right] \\
&= veco\left[\mathbf{C}^{-1\prime}\left\{\mathbf{I}_N+\begin{bmatrix}\frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*}\varepsilon_{Nt}^*(\boldsymbol{\theta})\\ \vdots & \ddots & \vdots \\ \frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*}\varepsilon_{Nt}^*(\boldsymbol{\theta})\end{bmatrix}\right\}\boldsymbol{\Psi}\right],
\end{aligned}
$$

where $\boldsymbol{\Delta}_N$ is an $N^2\times N(N-1)$ matrix such that $vec(\mathbf{J}-\mathbf{I}_N)=\boldsymbol{\Delta}_N veco(\mathbf{J}-\mathbf{I}_N)$, and

$$
s_{\boldsymbol{\psi}}(\boldsymbol{\theta};\boldsymbol{\varrho})=\boldsymbol{\Psi}^{-1}\mathbf{E}_N'vec\left[\mathbf{I}_N+\frac{\partial f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]=\boldsymbol{\Psi}^{-1}\left\{\begin{array}{c}1+\frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*}\varepsilon_{1t}^*(\boldsymbol{\theta})\\ \vdots\\ 1+\frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*}\varepsilon_{Nt}^*(\boldsymbol{\theta})\end{array}\right\}.
$$

Let us start by assuming that the shape parameters $\boldsymbol{\varrho}$ are fixed to some value $\bar{\boldsymbol{\varrho}}$. Let $\boldsymbol{\upsilon}=\mathbf{J}^{-1}\boldsymbol{\tau}$ so that $\boldsymbol{\tau}=\mathbf{J}\boldsymbol{\upsilon}$. In addition, let

$$
\boldsymbol{\varepsilon}_t^*(\boldsymbol{\tau},\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi})=\boldsymbol{\Psi}^{-1}\mathbf{J}_0^{-1}(\mathbf{y}_t-\boldsymbol{\tau}-\mathbf{A}_{10}\mathbf{y}_{t-1}-\ldots-\mathbf{A}_{p0}\mathbf{y}_{t-p})=\boldsymbol{\Psi}^{-1}[(\boldsymbol{\upsilon}_0-\boldsymbol{\upsilon})+\boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t^*].
$$

Finally, define the pseudo true values of the parameters $\boldsymbol{\upsilon}_\infty$, $\boldsymbol{\tau}_\infty=\mathbf{J}_0\boldsymbol{\upsilon}_\infty$ and $\boldsymbol{\psi}_\infty$ such that

$$
\begin{aligned}
E\left\{-\frac{\partial f[\varepsilon_{it}^*(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty);\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*}\right\}&=E\left\{-\frac{\partial f[\psi_{i\infty}^{-1}[(\upsilon_0-\upsilon_\infty)+\psi_{i0}\varepsilon_{it}^*];\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*}\right\}=0\quad\forall i,\\
E\left\{1+\frac{\partial f[\varepsilon_{it}^*(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty);\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*}\varepsilon_{it}^*(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty)\right\}&\\
=E\left\{1+\frac{\partial f[\psi_{i\infty}^{-1}[(\upsilon_0-\upsilon_\infty)+\psi_{i0}\varepsilon_{it}^*];\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*}\psi_{i\infty}^{-1}[(\upsilon_0-\upsilon_\infty)+\psi_{i0}\varepsilon_{it}^*]\right\}&=0\quad\forall i.
\end{aligned}
$$

As a result, the expected value of the scores of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ will be 0. But the cross-sectional independence of the true shocks combined with these expressions imply that the expected value of the scores of $\mathbf{a}$ and $\mathbf{j}$ and will also be 0. Consequently, all parameters except $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ will be consistently estimated.

When $\boldsymbol{\varrho}$ is simultaneously estimated, one should understand the solutions $\boldsymbol{\tau}_\infty$ and $\boldsymbol{\psi}_\infty$ to the above equations as functions of the fixed value of $\boldsymbol{\varrho}$, and add the scores for these parameters as additional model conditions, which implicitly define their pseudo true values $\boldsymbol{\varrho}_\infty$. □

A.4 Proposition 4

The proof is very similar to the proof Proposition 3. The main difference is when the assumed distributions of all the structural shocks are symmetric, the score expressions simplify to

$$
s_{\boldsymbol{\tau}}(\boldsymbol{\theta};\boldsymbol{\varrho})=\mathbf{C}^{-1\prime}\left\{\begin{array}{c}\delta[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\boldsymbol{\varepsilon}^*{}_{1t}(\boldsymbol{\theta})\\ \vdots\\ \delta[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]\boldsymbol{\varepsilon}^*{}_{Nt}(\boldsymbol{\theta})\end{array}\right\},
$$

$$s_{\mathbf{a}}(\boldsymbol{\theta};\boldsymbol{\varrho}) = \left(\begin{array}{c} \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \end{array} \right) \mathbf{C}^{-1\prime} \left\{ \begin{array}{c} \delta[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\boldsymbol{\varepsilon}^*{}_{1t}(\boldsymbol{\theta}) \\ \vdots \\ \delta[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]\boldsymbol{\varepsilon}^*{}_{Nt}(\boldsymbol{\theta}) \end{array} \right\},$$

$$\begin{aligned} s_{\mathbf{j}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= veco \left[\mathbf{C}^{-1\prime} \left\{ \begin{array}{ccc} \delta[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{1t}(\boldsymbol{\theta}) - 1 & \dots & \delta[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{1t}(\boldsymbol{\theta})\varepsilon^*_{Nt}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \delta[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{Nt}(\boldsymbol{\theta})\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & \delta[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{Nt}(\boldsymbol{\theta}) - 1 \end{array} \right\} \boldsymbol{\Psi} \right], \\ s_{\boldsymbol{\psi}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= \boldsymbol{\Psi}^{-1} \left\{ \begin{array}{c} \delta[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{1t}(\boldsymbol{\theta}) - 1 \\ \vdots \\ \delta[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{Nt}(\boldsymbol{\theta}) - 1 \end{array} \right\}, \end{aligned}$$

where $\delta[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]$ is a scalar function of the square of $\varepsilon^*_{it}(\boldsymbol{\theta})$.

In this case, it is easy to see that $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi}) = \boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}_0\boldsymbol{\varepsilon}^*_t$, so that for a fixed value of the shape parameters $\bar{\boldsymbol{\varrho}}$,

$$E\{\delta[\varepsilon^*_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi});\bar{\boldsymbol{\varrho}}_i]\varepsilon^*_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi})\} = E[\delta(\psi_i^{-1}\psi_{i0}\boldsymbol{\varepsilon}^*{}_{it};\bar{\boldsymbol{\varrho}}_i)\psi_i^{-1}\psi_{i0}\varepsilon^*_{it}] = 0 \quad \forall i \qquad \text{(A1)}$$

for any value of $\boldsymbol{\psi}$ because the integrand is an odd function of ε^*_{it}, whose true distribution is symmetric. As a result, the expected value of the scores of $\boldsymbol{\tau}$ will be 0 and the same applies to the scores of $\mathbf{a}$ because of the law of iterated expectations.

Next, let us define $\boldsymbol{\psi}_\infty$ such that

$$E\left[\delta[\varepsilon^*_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi});\bar{\boldsymbol{\varrho}}_i]\varepsilon^{*2}_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi}) - 1\right] = E[\delta(\psi_{i\infty}^{-1}\psi_{i0}\boldsymbol{\varepsilon}^*{}_{it};\bar{\boldsymbol{\varrho}}_i)\psi_{i\infty}^{-2}\psi_{i0}^2\varepsilon^{*2}_{it} - 1] = 0.$$

In this case,

$$\begin{aligned} &E\{\delta[\varepsilon^*_{it}(\boldsymbol{\tau}_\infty, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi}_\infty);\bar{\boldsymbol{\varrho}}_i]\varepsilon^*_{it}(\boldsymbol{\tau}_\infty, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi}_\infty)\} \\ &= E[(\psi_{i\infty}^{-1}[(\upsilon_0 - \upsilon_\infty) + \psi_{i0}\varepsilon^*_{it}];\bar{\boldsymbol{\varrho}}_i)\psi_{i\infty}^{-1}[(\upsilon_0 - \upsilon_\infty) + \psi_{i0}\varepsilon^*{}_{it}]] = 0 \quad \forall i, \\ &E\left[\delta[\varepsilon^*_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi});\bar{\boldsymbol{\varrho}}_i]\varepsilon^{*2}_{it}(\boldsymbol{\tau}_0, \mathbf{a}_0, \mathbf{j}_0, \boldsymbol{\psi}) - 1\right] = E[\delta(\psi_{i\infty}^{-1}\psi_{i0}\varepsilon^*_{it};\bar{\boldsymbol{\varrho}}_i)\psi_{i\infty}^{-2}\psi_{i0}^2\varepsilon^{*2}_{it} - 1] = 0, \end{aligned}$$

as desired. As a result, the expected value of the scores of $\boldsymbol{\psi}$ will be 0. But the cross-sectional independence of the true shocks combined with (A1) implies that the expected value of the scores of $\mathbf{j}$ will also be 0. Consequently, all parameters except $\boldsymbol{\psi}$ will be consistently estimated.

Once again, when $\boldsymbol{\varrho}$ is simultaneously estimated, one should understand the solution $\boldsymbol{\psi}_\infty$ to the above equation as a function of the fixed value of $\boldsymbol{\varrho}$, and add the scores for these parameters as additional model conditions, which implicitly define their pseudo true values $\boldsymbol{\varrho}_\infty$. □

B Computational details

The log-density of a univariate Student t random variable with 0 mean, unit variance and degrees of freedom $\nu_i = \eta_i^{-1}$ is given by

$$l[\varepsilon_{it}^*(\boldsymbol{\theta}); \eta_i] = c(\eta_i) - \left(\frac{\eta_i + 1}{2\eta_i}\right) \log\left[1 + \frac{\eta_i}{1 - 2\eta_i}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right],$$

with

$$c(\eta_i) = \log\left(\frac{\eta_i + 1}{2\eta_i}\right) - \log\left[\Gamma\left(\frac{1}{2\eta_i}\right)\right] - \frac{1}{2}\log\left(\frac{1 - 2\eta_i}{\eta_i}\right) - \frac{1}{2}\log\pi.$$

In contrast, it becomes

$$l[\varepsilon_{it}^*(\boldsymbol{\theta}); \lambda_i, \kappa_i] = c(\lambda_i, \kappa_i) + \log\left[\lambda_i \exp\left(-\frac{\varepsilon_{it}^{*2}(\boldsymbol{\theta})}{\kappa_i^*}\right) + (1 - \lambda_i)\kappa_i^{-1/2} \exp\left(-\frac{\varepsilon_{it}^{*2}(\boldsymbol{\theta})}{\kappa_i^* \kappa_i}\right)\right]$$

for a two-component DSMN, with

$$c(\lambda_i, \kappa_i) = -\frac{1}{2}\log\kappa_i^* - \log\Gamma\left(-\frac{1}{2}\right)$$

and

$$\kappa_i^* = \frac{1}{\lambda_i + (1 - \lambda_i)\kappa_i}.$$

Finally, it will be

$$l[\varepsilon_{it}^*(\boldsymbol{\theta})] = -\log(2) - \sqrt{2}|\varepsilon_{it}^*(\boldsymbol{\theta})|$$

under the Laplace assumption.

To apply the symmetric FS correction, one simply needs to use (15) and compute

$$\tilde{\psi}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it}^{*2},\ i = 1, \ldots, N,$$

where $\widehat{\boldsymbol{\varepsilon}}_t^* = \widehat{\mathbf{J}}^{-1}\widehat{\boldsymbol{\varepsilon}}_t$ and $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \widehat{\boldsymbol{\tau}} - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p}$. Let $\tilde{\boldsymbol{\Psi}} = diag(\tilde{\psi}_1, \ldots, \tilde{\psi}_N)$, then the consistent estimator of $\mathbf{C}$ is obtained as

$$\tilde{\mathbf{C}} = \hat{\mathbf{J}}\tilde{\boldsymbol{\Psi}}. \tag{B1}$$

In turn, to apply the asymmetric FS correction, first compute the pseudo residuals as

$$\widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p},$$

and use (18) to compute

$$\tilde{\tau}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it},\ i = 1, \ldots, N,$$

which yields the consistent estimator of $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_1, \ldots, \tilde{\tau}_N)'$. As for the standard deviations, we

can use (17) to compute

$$\tilde{\psi}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \bar{\varepsilon}_{it}^{*2},\ i = 1, \ldots, N,$$

with $\bar{\boldsymbol{\varepsilon}}_t^* = \hat{\mathbf{J}}^{-1}\bar{\boldsymbol{\varepsilon}}_t$ and $\bar{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \tilde{\boldsymbol{\tau}} - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p}$. Finally, one simply needs to use (B1).

C Standardised random variables

C.1 Univariate discrete location scale mixtures of normals

Let s_t denote an *i.i.d.* Bernoulli variate with $P(s_t = 1) = \lambda$. If $z_t|s_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^* = \frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_t - \lambda) + \frac{s_t + (1-s_t)\sqrt{\kappa}}{\sqrt{\lambda + (1-\lambda)\kappa}} z_t\right],$$

where $\delta \in \mathbb{R}$ and $\kappa > 0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t - \lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t + (1-s_t)\sqrt{\kappa}}{\sqrt{\lambda + (1-\lambda)\kappa}} z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t - \lambda)$, the sum of the two random variables will have variance $1 + \lambda(1-\lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardised random variable is as follows

$$\varepsilon_t^* = \begin{cases} N[\mu_1^*(\boldsymbol{\varrho}), \sigma_1^{*2}(\boldsymbol{\varrho})] \text{ with probability } \lambda \\ N[\mu_2^*(\boldsymbol{\varrho}), \sigma_2^{*2}(\boldsymbol{\varrho})] \text{ with probability } 1-\lambda \end{cases} \tag{C1}$$

where $\boldsymbol{\varrho} = (\delta, \kappa, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\boldsymbol{\varrho}) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\boldsymbol{\varrho}) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\boldsymbol{\varrho}), \\
\sigma_1^{*2}(\boldsymbol{\varrho}) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda + (1-\lambda)\kappa]}, \\
\sigma_2^{*2}(\boldsymbol{\varrho}) &= \frac{\kappa}{[1+\lambda(1-\lambda)\delta^2][\lambda + (1-\lambda)\kappa]} = \kappa\sigma_1^{*2}(\boldsymbol{\varrho}).
\end{aligned}$$

Therefore, we can immediately interpret κ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\boldsymbol{\varrho}) - \mu_2^*(\boldsymbol{\varrho})}{\sqrt{\lambda\sigma_1^{*2}(\boldsymbol{\varrho}) + (1-\lambda)\sigma_1^{*2}(\boldsymbol{\varrho})}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components relative to the mean of the two conditional variances.

Finally, note that we can also use the above expressions to generate a two component mixture of normals with mean π and variance ω^2 as

$$y_t = \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } \lambda \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-\lambda \end{cases}$$

with

$$\mu_1 = \pi + \omega\mu_1^*(\boldsymbol{\varrho}), \quad \mu_2 = \pi + \omega\mu_2^*(\boldsymbol{\varrho}), \quad \sigma_1^2 = \omega\sigma_1^{*2}(\boldsymbol{\varrho}), \quad \sigma_2^2 = \omega\sigma_2^{*2}(\boldsymbol{\varrho}).$$

Interestingly, the expressions for κ and δ above continue to be valid if we replace $\mu_1^*(\boldsymbol{\varrho})$, $\mu_2^*(\boldsymbol{\varrho})$, $\sigma_1^{*2}(\boldsymbol{\varrho})$ and $\sigma_2^{*2}(\boldsymbol{\varrho})$ by μ_1, μ_2, σ_1^2 and σ_2^2.

We can trivially extend this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (C1) by a $(K-1)$-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\varrho})$ and $\sigma_1^{*2}(\boldsymbol{\varrho})$, respectively, then the resulting random variable will be a K-component Gaussian mixture with zero mean and unit variance.

C.2 Standardised multivariate discrete location scale mixtures of normals

Consider the following mixture of two multivariate normals

$$\boldsymbol{\varepsilon}_t \sim \begin{cases} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \text{with probability } \lambda, \\ N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) & \text{with probability } 1-\lambda. \end{cases} \tag{C2}$$

Let s_t denote a Bernoulli variable which takes the value 1 with probability λ and 0 with probability $1-\lambda$. As is well known, the unconditional mean vector and covariance matrix of the observed variables are:

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t) &= E[E(\boldsymbol{\varepsilon}_t|s_t)] = \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2, \\ V(\boldsymbol{\varepsilon}_t) &= V[E(\boldsymbol{\varepsilon}_t|s_t)] + E[V(\boldsymbol{\varepsilon}_t|s_t)] = \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2. \end{aligned}$$

Therefore, this random vector will be standardised if and only if

$$\begin{aligned} \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2 &= \mathbf{0}, \\ \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 &= \mathbf{I}. \end{aligned}$$

Let us initially assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$ but that the mixture is not degenerate, so that $\lambda \neq 0, 1$. Let $\boldsymbol{\Sigma}_{1L}\boldsymbol{\Sigma}_{1L}'$ and $\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}_{2L}'$ denote the Cholesky decompositions of the covariance

matrices of the two components. Then, we can write

$$\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{1L}[\lambda\mathbf{I}_N + (1-\lambda)\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}_{2L}'\boldsymbol{\Sigma}_{1L}^{-1\prime}]\boldsymbol{\Sigma}_{1L}' = \boldsymbol{\Sigma}_{1L}(\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}_L')\boldsymbol{\Sigma}_{1L}',$$

where $\mathbf{K}_L = \sqrt{1-\lambda}\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}$ remains a lower triangular matrix. Given that $\mathbf{I}_N = \mathbf{e}_1\mathbf{e}_1 + \ldots + \mathbf{e}_N\mathbf{e}_N$, where $\mathbf{e}_i$ is the i^{th} vector of the canonical basis, the Cholesky decomposition of $\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}_L'$, say $\mathbf{G}_L\mathbf{G}_L'$, can be computed by means of N rank-one updates that sequentially add $\sqrt{\lambda}\mathbf{e}_i\sqrt{\lambda}\mathbf{e}_i'$ for $i = 1, \ldots, N$. The special form of those vectors can be efficiently combined with the usual rank-one update algorithms to speed up this process (see e.g. Sentana (1999) and the references therein). In any case, the elements of $\mathbf{G}_L$ will be functions of λ and the $N(N+1)/2$ elements in $\mathbf{K}_L$. If we then choose $\boldsymbol{\Sigma}_{1L} = \mathbf{G}_L^{-1}$, we will guarantee that $\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \mathbf{I}_N$. Therefore, we can achieve a standardised two-component mixture of two multivariate normals with 0 means by drawing with probability λ one random variable from a distribution with covariance matrix $\mathbf{G}_L^{-1\prime}\mathbf{G}_L^{-1}$, and with probability $1-\lambda$ from another distribution with covariance matrix $(1-\lambda)^{-1}\mathbf{K}_L\mathbf{K}_L'$.

Let us now turn to the case in which the means of the components are no longer 0. The zero unconditional mean condition is equivalent to $\boldsymbol{\mu}_1 = (1-\lambda)\boldsymbol{\delta}$ and $\boldsymbol{\mu}_2 = -\lambda\boldsymbol{\delta}$, so that $\boldsymbol{\delta}$ measures the difference between the two means. Thus, the unconditional covariance matrix will be $\lambda(1-\lambda)\boldsymbol{\delta}\boldsymbol{\delta}' + \mathbf{I}_N$ after imposing the restrictions on $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the previous paragraph. Once again, the Cholesky decomposition of this matrix is very easy to obtain because it can be regarded as a positive rank-one update of the identity matrix, whose decomposition is trivial.

Thus, we can parametrise a standardised mixture of two multivariate normals, which usually involves $2N$ mean parameters, $2N(N+1)/2$ covariance parameters and one mixing parameter, in terms of the N mean difference parameters in $\boldsymbol{\delta}$, the $N(N+1)/2$ relative variance parameters in $\mathbf{K}_L$ and the mixing parameter λ, the remaining N mean parameters and $N(N+1)/2$ covariance ones freed up to target any unconditional mean vector and covariance matrix.

Mencía and Sentana (2009) explain how to standardise Bernoulli location-scale mixtures of normals, which are a special case of the two component mixtures we have just discussed in which $\boldsymbol{\Sigma}_2 = \kappa\boldsymbol{\Sigma}_1$. Straightforward algebra confirms that the standardisation procedure described above simplifies to the one they provide in their Proposition 1.

As in the univariate case, we can trivially extended this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (C2) by a $(K-1)$-component normal mixture with mean and variance given by $\boldsymbol{\mu}_1^*(\boldsymbol{\varrho})$ and $\boldsymbol{\Sigma}_1^*(\boldsymbol{\varrho})$, respectively, then the resulting random variable will be a K-component Gaussian mixture with zero mean and unit variance.

TABLE 1: Monte Carlo results. Mean absolute bias of pooled groups of estimators.

DGP: $\varepsilon_{1t} \sim$ Student t_6 $\varepsilon_{2t} \sim$ Student t_{12} $\varepsilon_{3t} \sim$ Student t_8

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0002	0.0002	0.0002	0.0001	0.0002	0.0001	0.0002	0.0002	0.0002	0.0002
A_{ii}	0.0016	0.0016	0.0017	0.0017	0.0017	0.0015	0.0015	0.0018	0.0018	0.0018
$A_{ij,i\neq j}$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0002
C_{ii}	0.0040	0.0043	0.0047	0.0046	0.0046	0.0727	0.0061	0.0044	0.0082	0.0048
$C_{ij,i>j}$	0.0009	0.0009	0.0011	0.0010	0.0011	0.0138	0.0008	0.0009	0.0009	0.0012
$C_{ij,i<j}$	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0008	0.0007	0.0009	0.0008
$J_{ij,i>j}$	0.0009	0.0009	0.0010	0.0010	0.0010	0.0008	0.0008	0.0009	0.0007	0.0010
$J_{ij,i<j}$	0.0011	0.0011	0.0010	0.0010	0.0010	0.0021	0.0021	0.0010	0.0028	0.0010

DGP: $\varepsilon_{1t} \sim$ Laplace $\varepsilon_{2t} \sim$ Laplace $\varepsilon_{3t} \sim$ Laplace

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
A_{ii}	0.0009	0.0009	0.0008	0.0009	0.0009	0.0007	0.0007	0.0016	0.0016	0.0016
$A_{ij,i\neq j}$	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003	0.0003	0.0003
C_{ii}	0.0915	0.0014	0.0013	0.0012	0.0013	0.0016	0.0011	0.0017	0.0015	0.0016
$C_{ij,i>j}$	0.0183	0.0003	0.0002	0.0002	0.0002	0.0004	0.0003	0.0003	0.0003	0.0002
$C_{ij,i<j}$	0.0004	0.0004	0.0003	0.0003	0.0003	0.0005	0.0005	0.0004	0.0005	0.0003
$J_{ij,i>j}$	0.0001	0.0001	0.0001	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
$J_{ij,i<j}$	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0004	0.0003	0.0006	0.0003

DGP: $\varepsilon_{1t} \sim \text{DLSMN}(0.8, 0.06, 0.52)$ $\varepsilon_{2t} \sim \text{DLSMN}(1.2, 0.08, 0.4)$ $\varepsilon_{3t} \sim \text{DLSMN}(-1, 0.2, 0.2)$

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.1865	0.0001	0.0001	0.2624	0.0001	0.1998	0.0001	0.0001	0.0001	0.0001
A_{ii}	0.0009	0.0009	0.0007	0.0007	0.0007	0.0009	0.0009	0.0015	0.0015	0.0015
$A_{ij,i\neq j}$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002
C_{ii}	3.3233	0.0011	0.0007	0.0399	0.0008	0.0128	0.0011	0.0021	0.0083	0.0012
$C_{ij,i>j}$	0.7162	0.0002	0.0002	0.0108	0.0002	0.0016	0.0003	0.0005	0.0089	0.0003
$C_{ij,i<j}$	0.0005	0.0003	0.0001	0.0002	0.0002	0.0004	0.0004	0.0007	0.0099	0.0001
$J_{ij,i>j}$	0.0003	0.0003	0.0002	0.0002	0.0002	0.0003	0.0003	0.0006	0.0095	0.0003
$J_{ij,i<j}$	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0008	0.0104	0.0000

DGP: $\varepsilon_{1t} \sim$ Asy. Student $t_{12,1}$ $\varepsilon_{2t} \sim$ Asy. Student $t_{14,5}$ $\varepsilon_{3t} \sim$ Asy. Student $t_{16,100}$

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.1244	0.0001	0.0001	0.1081	0.0001	0.1692	0.0002	0.0001	0.0001	0.0001
A_{ii}	0.0011	0.0011	0.0010	0.0011	0.0011	0.0011	0.0011	0.0016	0.0016	0.0016
$A_{ij,i\neq j}$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
C_{ii}	0.0075	0.0022	0.0016	0.0046	0.0020	0.0350	0.0034	0.0027	0.0093	0.0019
$C_{ij,i>j}$	0.0005	0.0002	0.0002	0.0003	0.0002	0.0129	0.0004	0.0003	0.0007	0.0002
$C_{ij,i<j}$	0.0004	0.0004	0.0004	0.0002	0.0002	0.0004	0.0004	0.0004	0.0025	0.0004
$J_{ij,i>j}$	0.0004	0.0004	0.0003	0.0002	0.0002	0.0008	0.0008	0.0004	0.0020	0.0003
$J_{ij,i<j}$	0.0004	0.0004	0.0002	0.0002	0.0002	0.0006	0.0006	0.0006	0.0039	0.0004

Sample length=2000, Replications=5,000. S: Student-t MLE, M: DLSMN MLE, SM: DSMN MLE, L: Laplace MLE, IC: GMR two step estimator. AFS is Fiorentini and Sentana (2019) consistent PMLE.

TABLE 2: Monte Carlo results. (RMSE) of pooled groups of estimators.

DGP:	$\varepsilon_{1t} \sim$ Student t_6		$\varepsilon_{2t} \sim$ Student t_{12}			$\varepsilon_{3t} \sim$ Student t_8				
	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0221	0.0232	0.0232	0.0222	0.0232	0.0257	0.0232	0.0232	0.0232	0.0232
A_{ii}	0.0195	0.0195	0.0196	0.0196	0.0196	0.0227	0.0227	0.0204	0.0204	0.0204
$A_{ij,i\neq j}$	0.0194	0.0194	0.0195	0.0195	0.0195	0.0226	0.0226	0.0204	0.0204	0.0204
C_{ii}	0.0247	0.0250	0.0255	0.0254	0.0254	0.0789	0.0275	0.0251	0.0303	0.0255
$C_{ij,i>j}$	0.0573	0.0573	0.0604	0.0598	0.0598	0.0815	0.0746	0.0573	0.0842	0.0605
$C_{ij,i<j}$	0.0561	0.0561	0.0595	0.0588	0.0588	0.0791	0.0728	0.0562	0.0831	0.0595
$J_{ij,i>j}$	0.0576	0.0576	0.0609	0.0603	0.0603	0.0756	0.0756	0.0577	0.0859	0.0610
$J_{ij,i<j}$	0.0582	0.0582	0.0618	0.0611	0.0611	0.0763	0.0763	0.0582	0.0887	0.0618

DGP:	$\varepsilon_{1t} \sim$ Laplace		$\varepsilon_{2t} \sim$ Laplace			$\varepsilon_{3t} \sim$ Laplace				
	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0185	0.0232	0.0232	0.0185	0.0232	0.0171	0.0231	0.0233	0.0233	0.0233
A_{ii}	0.0163	0.0163	0.0162	0.0162	0.0162	0.0151	0.0151	0.0203	0.0203	0.0203
$A_{ij,i\neq j}$	0.0162	0.0162	0.0162	0.0162	0.0162	0.0151	0.0151	0.0201	0.0201	0.0201
C_{ii}	0.1042	0.0255	0.0255	0.0255	0.0255	0.0227	0.0254	0.0256	0.0254	0.0255
$C_{ij,i>j}$	0.0332	0.0243	0.0242	0.0242	0.0242	0.0207	0.0208	0.0253	0.0221	0.0249
$C_{ij,i<j}$	0.0255	0.0233	0.0232	0.0232	0.0232	0.0200	0.0200	0.0244	0.0214	0.0240
$J_{ij,i>j}$	0.0238	0.0238	0.0237	0.0237	0.0237	0.0202	0.0202	0.0248	0.0216	0.0244
$J_{ij,i<j}$	0.0231	0.0231	0.0231	0.0231	0.0231	0.0199	0.0199	0.0242	0.0213	0.0239

DGP:	$\varepsilon_{1t} \sim$ DLSMN$(0.8, 0.06, 0.52)$		$\varepsilon_{2t} \sim$ DLSMN$(1.2, 0.08, 0.4)$			$\varepsilon_{3t} \sim$ DLSMN$(-1, 0.2, 0.2)$				
	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.2342	0.0231	0.0231	0.3128	0.0231	0.2470	0.0231	0.0232	0.0232	0.0232
A_{ii}	0.0152	0.0152	0.0131	0.0138	0.0138	0.0165	0.0165	0.0202	0.0202	0.0202
$A_{ij,i\neq j}$	0.0151	0.0151	0.0130	0.0137	0.0137	0.0165	0.0165	0.0201	0.0201	0.0201
C_{ii}	5.7458	0.0227	0.0225	0.0545	0.0226	0.0282	0.0228	0.0235	0.0298	0.0226
$C_{ij,i>j}$	1.1789	0.0223	0.0187	0.0239	0.0200	0.0273	0.0271	0.0330	0.0845	0.0210
$C_{ij,i<j}$	0.1165	0.0192	0.0133	0.0152	0.0148	0.0218	0.0215	0.0333	0.0838	0.0144
$J_{ij,i>j}$	0.0219	0.0219	0.0183	0.0195	0.0195	0.0268	0.0268	0.0327	0.0854	0.0206
$J_{ij,i<j}$	0.0184	0.0184	0.0130	0.0144	0.0144	0.0204	0.0204	0.0339	0.0988	0.0150

DGP:	$\varepsilon_{1t} \sim$ Asy. Student $t_{12,1}$		$\varepsilon_{2t} \sim$ Asy. Student $t_{14,5}$			$\varepsilon_{3t} \sim$ Asy. Student $t_{16,100}$				
	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.1424	0.0231	0.0230	0.1244	0.0231	0.1861	0.0231	0.0231	0.0231	0.0231
A_{ii}	0.0180	0.0180	0.0168	0.0178	0.0178	0.0215	0.0215	0.0201	0.0201	0.0201
$A_{ij,i\neq j}$	0.0181	0.0181	0.0169	0.0179	0.0179	0.0216	0.0216	0.0202	0.0202	0.0202
C_{ii}	0.0340	0.0324	0.0321	0.0335	0.0322	0.0530	0.0338	0.0329	0.0398	0.0320
$C_{ij,i>j}$	0.0338	0.0340	0.0252	0.0319	0.0319	0.0567	0.0516	0.0392	0.0889	0.0253
$C_{ij,i<j}$	0.0366	0.0363	0.0288	0.0349	0.0346	0.0558	0.0547	0.0411	0.0887	0.0295
$J_{ij,i>j}$	0.0336	0.0336	0.0246	0.0315	0.0315	0.0517	0.0517	0.0391	0.0910	0.0247
$J_{ij,i<j}$	0.0383	0.0383	0.0303	0.0367	0.0367	0.0575	0.0575	0.0427	0.0909	0.0308

Sample length=2000, Replications=5,000. S: Student-t MLE, M: DLSMN MLE, SM: DSMN MLE, L: Laplace MLE, IC: GMR two step estimator. AFS is Fiorentini and Sentana (2019) consistent PMLE.

FIGURE 1: Volatility index series (logs)

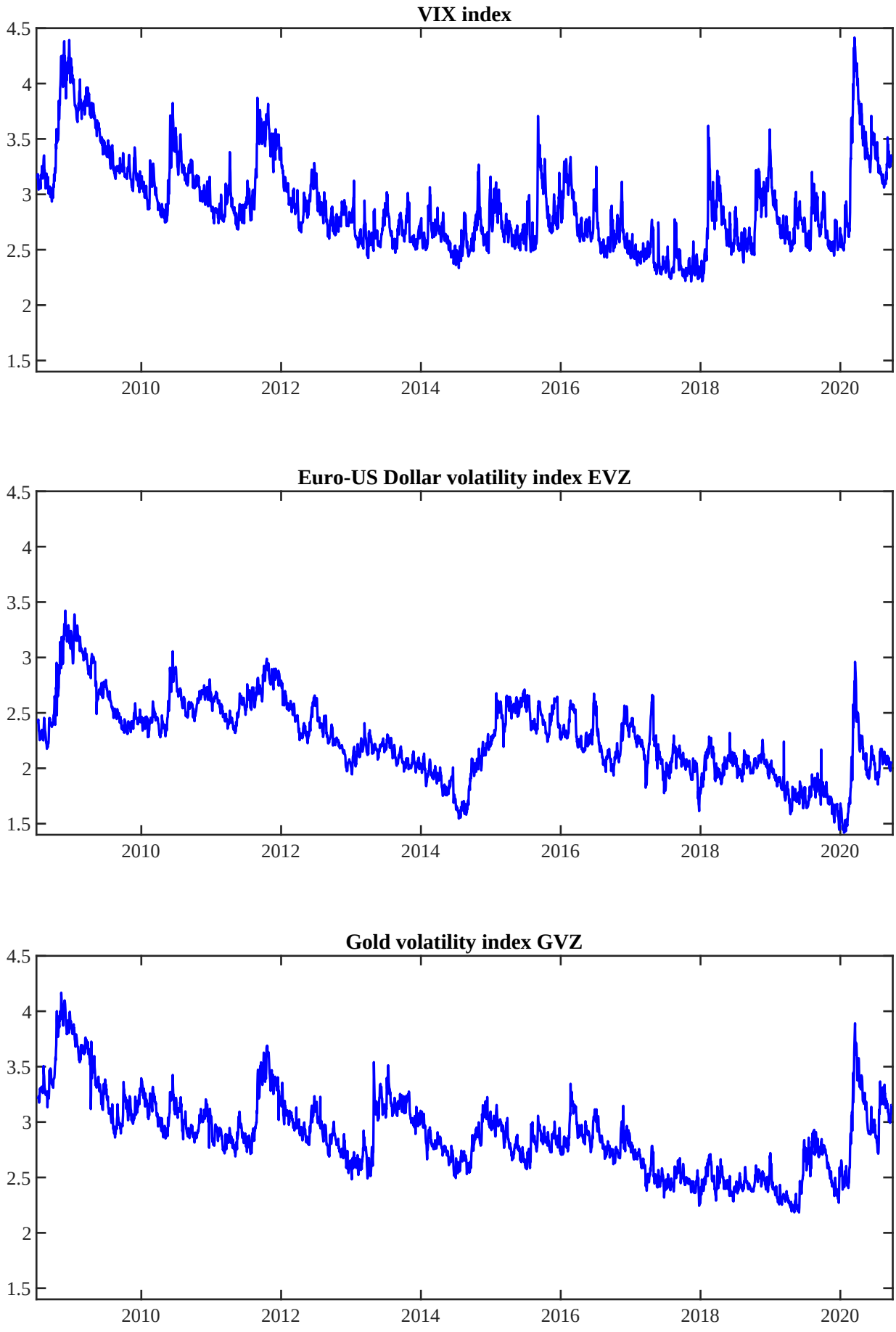

FIGURE 2: Structural Shocks

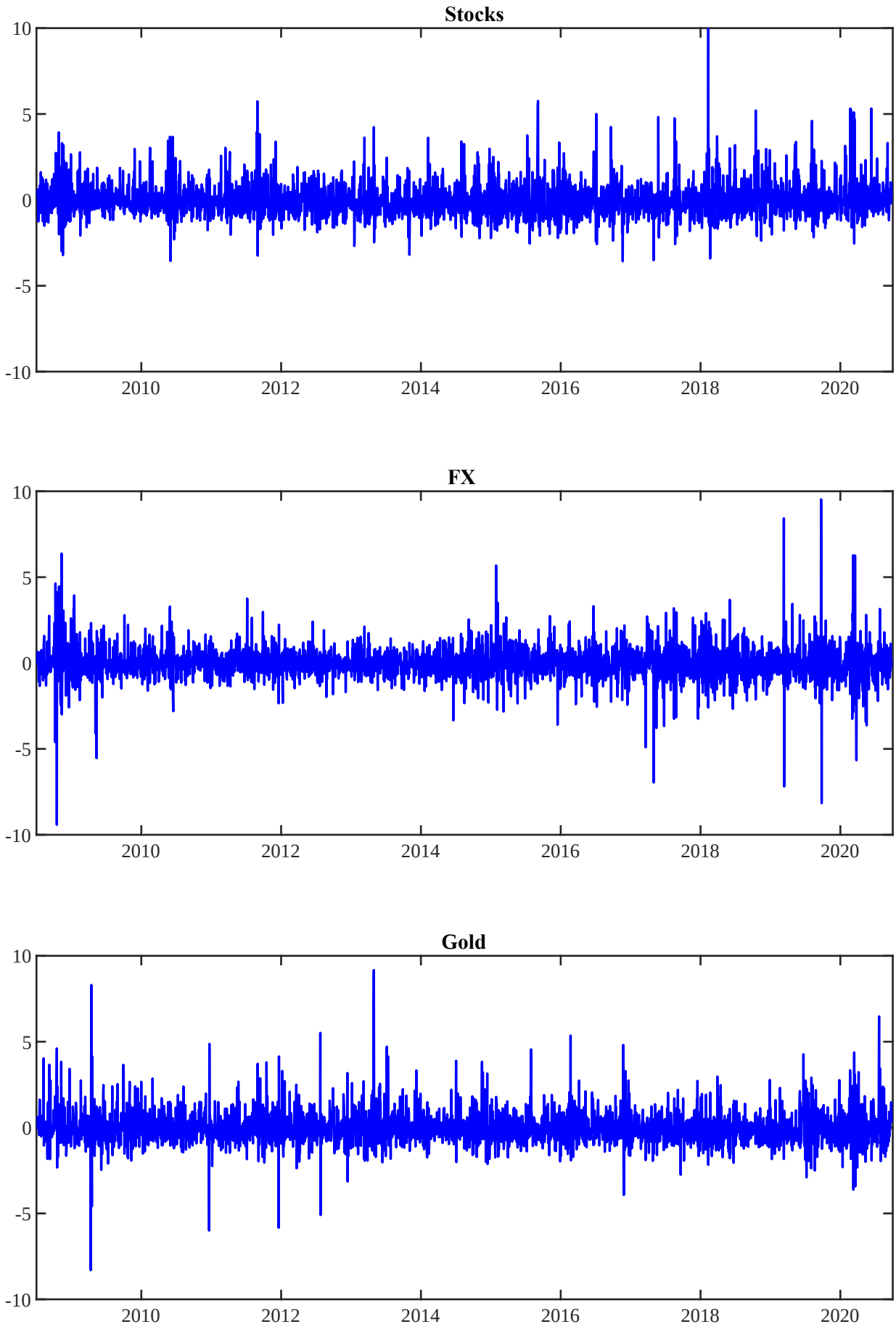

FIGURE 3 Impulse response functions and forecast error variance decompostions