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

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

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

October 2020
Revised: February 2022

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 shocks' standard deviations are generally inconsistent. Nevertheless, we show consistency when the non-Gaussian log-likelihood uses a discrete scale mixture of normals in the symmetric case, or an unrestricted finite mixture more generally, and compare the efficiency of these estimators to other consistent two-step proposals, including our own. Finally, our empirical application looks at dynamic linkages between three popular volatility indices.

Keywords: Consistency, Efficiency bound, Finite normal mixtures, Pseudo maximum likelihood estimators, Structural models, Volatility indices.

JEL: C32, C46, C51, C58

*We would like to Dante Amengual and mostly virtual audiences at ICEEE (Cagliari, 2021), IAAE Annual Conference (Rotterdam, 2021), Finance Forum (Lisbon, 2021), RCEA Money, Macro and Finance Conference (Rimini and Riverside, 2021), ESEM (Copenhagen, 2021) and SAEe (Barcelona, 2021) for useful comments and suggestions. We also thank the editor, an associate editor, and two anonymous referees for valuable feedback. Of course, the usual caveat applies. Financial support from MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic" (Fiorentini) and the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair (Sentana) is also gratefully acknowledged.

1 Introduction

Statistical identification of the parameters of a structural vector autoregression (SVAR) through independent non-Gaussian shocks is becoming increasingly popular after Lanne et al (2017) and Gouriéroux et al (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), Herwartz (2018), Coad and Grassano (2019), Herwartz (2019), Puonti (2019), Tank et al (2019), Gouriéroux et al (2020), Maxand (2020), Bekaert et al (2020, 2021), Guay (2021), Lanne and Luoto (2021), Bernoth and Herwartz (2021), Montiel Olea et al (2021), Braun (2021) and Davis and Ng (2021).

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 when the assumed univariate distributions are correctly specified. Unfortunately, while Gaussian pseudo maximum likelihood estimators (PMLE) remain consistent for the identified conditional mean and variance 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 could 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, as we show in section 2.3 below. 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,

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

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

In Fiorentini and Sentana (2007, 2019) (FS), we discussed some simple consistent estimators to replace the parameters inconsistently estimated by a distributionally misspecified log-likelihood in multivariate, conditionally heteroskedastic, dynamic regression models. In this paper, we show that the analogous consistent estimators for SVAR models with cross-sectionally independent shocks are simple transformations of the first and second sample moments of the estimated shocks obtained using the non-Gaussian PMLEs.

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, so that the second-step FS estimators are numerically identical to the first-step ones.[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 that use an alternative parametric distribution, such as the popular Student t or the Laplace. The analytical expressions for the asymptotic covariance matrices of these two estimators that we derive allow us to study this important issue both in theory and by means of Monte Carlo simulations. 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 (2020) for general ones) suggests that the flexible parametric procedure we consider has the potential to achieve the cross-sectionally independent SP efficiency bound, which we obtain in closed-form. We also compare our estimators to the two-step procedure in Gouriéroux et al (2017), which estimates all the reduced form parameters by Gaussian PML, and the orthogonal rotation matrix mapping structural shocks and reduced form innovations by

[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 (2021a)).

non-Gaussian PML.

Finally, we apply our proposed estimators to the empirical analysis of the dynamic linkages 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 SVAR models with cross-sectionally independent shocks, characterise the parameters that remain consistently estimated under distributional misspecification, and show that they can be adaptively estimated, in the sense that a suitable SP procedure would be as efficient for them as a parametric procedure that exploited knowledge of the true distribution of the shocks, including the values of its shape parameters. Next, in section 3 we present an extensive Monte Carlo exercise that combines several simulation and estimation densities, paying special attention to the coverage of the IRFs, whose confidence bands we also compute analytically. Then, we describe our empirical application to the aforementioned volatility indices in section 4, followed by our concluding remarks. Proofs and auxiliary results are gathered in appendices.

2 Structural vector autoregressions

2.1 The model

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

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

where I_{t-1} is the information set, $\mathbf{C}$ the matrix of impact multipliers and $\boldsymbol{\varepsilon}_t^*$ the "structural" shocks, which are normalised to have zero means, unit variances and zero covariances. 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 $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$, where $veco(\mathbf{J} - \mathbf{I}_N)$ stacks by columns all the elements of the zero-diagonal matrix

$\mathbf{J}-\mathbf{I}_N$ except those that appear in its diagonal, while $vecd(\mathbf{\Psi})$ places the elements in the main diagonal of $\mathbf{\Psi}$ in a column vector.[3] 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 $\boldsymbol{\varepsilon}_t|I_{t-1}\sim i.i.d.\ (\mathbf{0},\mathbf{\Sigma})$ with $\mathbf{\Sigma}=\mathbf{C}\mathbf{C}'=\mathbf{J}\mathbf{\Psi}^2\mathbf{J}'$. As is well known, a Gaussian (pseudo) log-likelihood is only able to identify $\mathbf{\Sigma}$, 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[4] to relate the matrix $\mathbf{C}$ to the Cholesky decomposition of $\mathbf{\Sigma}=\mathbf{\Sigma}_L\mathbf{\Sigma}_L'$ as $\mathbf{C}=\mathbf{\Sigma}_L\mathbf{Q}$, 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 $|\mathbf{Q}|=1$.[5,6] While $\mathbf{\Sigma}_L$ 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.[7]

Nevertheless, Lanne et al (2017) show that statistical identification of both the structural shocks and $\mathbf{C}$ (up to column permutations and sign changes) is possible assuming (i) cross-sectional independence of the N shocks, (ii) a non-Gaussian distribution for at least $N-1$ of them, and (iii) $\mathbf{C}$ has full rank. In what follows, we assume that the N structural shocks are cross-sectionally independent and non-Gaussian, such as $\varepsilon_{it}^*|I_{t-1}\sim i.i.d.\ t(0,1,\nu_i)$. 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 (2021). Other popular examples are the generalised error (or Gaussian) distribution, which includes normal, Laplace (or double exponential) and uniform as special cases, as well as symmetric and asymmetric finite normal mixtures.

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 $\boldsymbol{\varphi}=(\boldsymbol{\theta}',\boldsymbol{\varrho}')'$. We consider two ML estimators: a restricted one which fixes $\boldsymbol{\varrho}$ to its supposedly true value, and an unrestricted one,

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

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

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

[6] 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.

[7] The identifying assumption of Proposition 1 in Lanne and Lütkepohl (2010) explicitly rules out scale mixtures of normals, which are elliptical.

which simultaneously estimates all the elements of $\boldsymbol{\varphi}$ (see Fiorentini and Sentana (2021b) for further details).

2.2 Consistency results

Maximum likelihood will usually be not only consistent but also fully efficient when the assumed univariate distributions are correctly specified under standard regularity conditions (see Appendix B for detailed assumptions). Unfortunately, a misspecified non-Gaussian distribution might lead to inconsistencies. Nevertheless, this does not mean that all the parameters are inconsistently estimated. In fact, 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 under suitable regularity conditions, which we also discuss in Appendix B. More formally:

Proposition 1 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (1) is the product of N univariate densities but they are potentially different from the ones assumed for estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (1) 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 FS consistent estimators, which replace the non-Gaussian PMLEs of τ_i and ψ_i by some simple transformations of the sample mean and variance of the i^{th} pseudo-standardised residual, will work in this context too, as we explain in Appendix C.

Proposition 1 also illustrates the practical consequences of distributional misspecification. Given that the IRFs of the structural VAR model in (1) 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 1 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 2 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (1) is the product of N univariate symmetric densities but they are potentially different from the symmetric ones assumed for estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (1) remain consistent for* $\mathbf{a}$, $\mathbf{j}$ *and* $\boldsymbol{\tau}$.

Propositions 1 and 2 are reminiscent of the fact that estimators based on a Gaussian pseudo log-likelihood function are consistent for all the conditional mean and variance parameters when these two moments are correctly specified under some regularity conditions. One of those conditions, though, is that the parameters be identified from this criterion function. However, a Gaussian log-likelihood can only identify the matrix $\mathbf{C}$ up to an orthogonal rotation, so they are not useful in this case. Remarkably, we can show that we can still achieve consistency for all the conditional mean and variance parameters if we replace the Gaussian log-likelihood function by one based on discrete mixtures of normals. More formally:

Proposition 3 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (1) is the product of N univariate densities, then the unrestricted (pseudo) ML estimators of model (1) that assume discrete mixtures of normals for the shocks are consistent for $\boldsymbol{\tau}$, $\mathbf{a}$, $\mathbf{j}$ and $\boldsymbol{\psi}$.*

Not surprisingly, there is a symmetric version of this result too:

Proposition 4 *If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (1) is the product of N univariate symmetric densities, then the unrestricted (pseudo) ML estimators of model (1) that assume discrete scale mixtures of normals for the shocks are consistent for $\boldsymbol{\tau}$, $\mathbf{a}$, $\mathbf{j}$ and $\boldsymbol{\psi}$.*

These results suggest the use of log-likelihood functions based on discrete normal mixtures to estimate the parameters of model (1). Nevertheless, their efficiency is also an important consideration. In section 3, we will use our simulation experiments to compare these mixture-based PMLEs to two-step FS estimators that rely on an alternative parametric distribution, such as the popular Student t or the Laplace. In addition, in the next subsection we compare the analytical expressions for the asymptotic covariance matrix of these two estimators, which we derive in Appendices B and C, respectively. In addition, we characterise the maximum efficiency that they can achieve.

Finally, it is illustrative to compare the results in this section to the consistency results in Gouriéroux et al (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, Proposition 1 shows that a non-Gaussian PMLE can consistently estimate the $N(N-1)$ elements of $\mathbf{J}$, so there are only N left. In addition, Proposition 3 implies that the non-Gaussian PMLE of the diagonal elements of $\boldsymbol{\Psi}$ will also be consistently estimated if we assume for estimation purposes that the structural shocks follow univariate mixtures of normals. Moreover, the FS estimators, which are effectively Gaussian PMLEs based on pseudo-standardised residuals, will provide consistent estimators of the N elements of $\boldsymbol{\psi}$ too.

2.3 Efficiency results

There are situations in which some, but not all elements of $\boldsymbol{\theta}$ can be estimated as efficiently as if the true distribution of the shocks were known, a fact described in the semiparametric literature as partial adaptivity. Effectively, this requires that some elements of the score be orthogonal to the relevant tangent set after partialling out the effects of the remaining elements of score by regressing the former on the latter. In the context of the multivariate, conditionally heteroskedastic, dynamic regression models, Fiorentini and Sentana (2021b) show that the model parameters that continue to be consistently estimated under distributional misspecification of the innovations coincide with the parameters that can be adaptively estimated. In this section, we show that this result also holds for model (1), with adaptivity achieved by the class of SP estimators that impose the assumption of cross-sectionally independent shocks along the lines of Chen and Bickel (2006), which we discuss in detail in Appendix D.[8]

More formally, let $\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}, \mathbf{a}) = (\mathbf{y}_t - \boldsymbol{\tau} - \mathbf{A}_1\mathbf{y}_{t-1} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p})$ denote the estimated reduced form residuals, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\tau}, \mathbf{a}, \mathbf{j}, \boldsymbol{\psi}) = \boldsymbol{\Psi}^{-1}\mathbf{J}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}, \mathbf{a})$ their structural counterparts and ℓ_N a vector of N ones:

Proposition 5 *1. If the true joint density of the structural shocks* $\boldsymbol{\varepsilon}_t^*$ *in (1) is the product of* N *univariate densities, then:*

(a) the cross-sectionally independent SP estimators of $\mathbf{a}$ *and* $\mathbf{j}$, $\ddddot{\mathbf{a}}$ *and* $\ddddot{\mathbf{j}}$, *respectively, are* $(\boldsymbol{\tau}, \boldsymbol{\psi})$*-adaptive,*

(b) the iterated cross-sectionally independent SP estimators of $\boldsymbol{\tau}$ *and* $\boldsymbol{\psi}$ *are* $\ddddot{\boldsymbol{\tau}} = \boldsymbol{\tau}(\ddddot{\mathbf{a}})$ *and* $\ddddot{\boldsymbol{\psi}} = \boldsymbol{\psi}(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}})$, *where for* $i = 1, \ldots, N$

$$\tau_i(\mathbf{a}) = T^{-1} \textstyle\sum_{t=1}^{T} \varepsilon_{it}(\mathbf{0}, \mathbf{a}), \tag{2}$$

$$\psi_i(\boldsymbol{\tau}, \mathbf{a}, \mathbf{j}) = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}^{*2}(\boldsymbol{\tau}, \mathbf{a}, \mathbf{j}, \ell_N). \tag{3}$$

Remarkably, (2) and (3) coincide with the FS consistent estimators based on the Gaussian pseudo-scores for $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$, the only difference being the first-step estimators chosen for $\mathbf{a}$ and $\mathbf{j}$ (see Appendix C for further details).

Another important implication of this proposition is that the restricted and unrestricted parametric estimators of $\mathbf{a}$ and $\mathbf{j}$ will be equally efficient despite the simultaneous estimation of the shape parameters.[9,10]

As before, we can strengthen the results in Proposition 5 by assuming that the true univariate distributions of the structural shocks are symmetric:

[8] See Lee and Mesters (2021) for the application of the Chen and Bickel (2006) estimators in simultaneous equation models.

[9] The full efficiency of the unrestricted MLEs confirms the analysis of the rank of the difference between the two parametric efficiency bounds at the end of Appendix D.1

[10] We also show in the proof of Proposition 5 that the restricted and unrestricted estimators of $\mathbf{a}$ and $\mathbf{j}$ will be asymptotically independent of each other, but not necessarily of the estimators of $\boldsymbol{\tau}$, $\boldsymbol{\psi}$ or the shape parameters.

Proposition 6 *1. If the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (1) is the product of N univariate symmetric densities, then:*

(a) the cross-sectionally independent symmetric SP estimators of $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$, $\dddot{\boldsymbol{\tau}}$, $\dddot{\mathbf{a}}$ and $\dddot{\mathbf{j}}$, respectively, are $\boldsymbol{\psi}$-adaptive,

(b) the iterated cross-sectionally independent symmetric SP estimator of $\boldsymbol{\psi}$ is given by $\dddot{\psi}_i = \psi_i(\dddot{\boldsymbol{\tau}}, \dddot{\mathbf{a}}, \dddot{\mathbf{j}})$.

Once again, the expression for $\dddot{\psi}_i$ coincides with the FS consistent estimators that impose symmetry, the only difference being the first-step estimators for $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ (see Appendix C for further details).

As in the case of Proposition 5, this result also implies that the unrestricted parametric estimators of $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ that rely on the correct distributional specification of the shocks will be fully efficient too despite the simultaneous estimation of the shape parameters.[11]

As we mentioned in the introduction, finite mixtures of normals with an increasing number of components may approximate many other distributions. To gain some insights on whether they could constitute the basis for a proper sieves-type SP procedure, we have conducted a simple exercise in which we look at a bivariate version of model (1) with cross-sectionally independent Student t shocks with 5 degrees of freedom each whose parameters are estimated by finite mixture-based log-likelihood functions with $K = 2, \ldots, 5$ components. For comparison purposes, we consider two different benchmarks: (i) the true MLE based on the correctly specified log-likelihood function, and (ii) the FS estimator which uses those MLEs as first-step estimators, but still replaces the estimators of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$. Importantly, this FS estimator is asymptotically equivalent to the cross-sectionally independent SP estimator in view of Proposition 5.[12]

We compute the information matrix of the correctly specified MLE and the asymptotic covariance matrix of the two-step FS estimator using the expressions in Appendix D of Fiorentini and Sentana (2021b) and Appendix C of this paper, respectively, evaluated at the true values of the parameters. As for the mixture-based PMLEs, we compute the expected value of the Hessian and variance of the score using the expressions in Amengual et al (2021a), which are a special case of the expressions in Appendix B, evaluated at the true values of $\boldsymbol{\tau}$, $\mathbf{a}$, $\mathbf{j}$ and $\boldsymbol{\psi}$ and the pseudo true values of the shape parameters, which we numerically obtain from samples of 40 million simulated observations.

The results, which we report in Table 1, show that the mixture-based PMLEs of $\mathbf{a}$ and $\mathbf{j}$ quickly approach the asymptotic efficiency of the true MLEs. In fact, although panel (a)

[11] The same conclusion derives from Proposition 14 in Fiorentini and Sentana (2021b), whose proof compares the diagonal block of the inverse information matrix corresponding to $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ when $\boldsymbol{\psi}$ and $\boldsymbol{\varrho}$ are jointly estimated to the corresponding block when only $\boldsymbol{\psi}$ is simultaneously estimated while $\boldsymbol{\varrho}$ is fixed to its true value.

[12] Moreover, the FS estimator that imposes symmetry will be in this case asymptotically equivalent to the cross-sectionally independent symmetric SP estimator, which in turn is as efficient as the MLE for all the parameters except $\boldsymbol{\psi}$ in view of Proposition 6.

in Figure 3 of Gallant and Tauchen (1999) clearly illustrates that a more complex misspecified model does not necessarily lead to more efficient estimators because one is not simply adding new elements to the score, but also changing the pseudo true values of the shape parameters at which one evaluates the original components of the score, we find that the efficiency improvements occur monotonically.

In contrast, the asymptotic variances of the mixture-based PMLEs of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ reach a plateau well above the asymptotic variances of the correctly specified MLEs. Not surprisingly, their asymptotic variances effectively coincide with the asymptotic variances of the FS estimators that use the correct MLEs of $\mathbf{a}$ and $\mathbf{j}$ as first-step estimators.[13]

3 Monte Carlo evidence

In this section, we assess the small sample behaviour 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

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0.2 & 1 & 0 \\ 0.2 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon^*_{1t} \\ \varepsilon^*_{2t} \\ \varepsilon^*_{3t} \end{bmatrix}. \quad (4)$$

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 2, 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 homogeneous 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 $T = 500$, which is realistic in macro applications with monthly data such as Ludvigson et al (2021) (see Supplemental Appendix G for the corresponding results for simulated samples of length $T = 2,000$, which are representative of financial applications with daily data, such as the one in section 4). We

[13] The fact that the asymptotic variances of ψ_1 and ψ_2 for some of the mixture-based PMLEs are marginally below those of the FS estimators is likely due to the numerical approximation error in the pseudo-true values in such complex models.

then 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 et al (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 $(\boldsymbol{\tau}', \mathbf{a}', \boldsymbol{\sigma}'_L)'$, with $\boldsymbol{\sigma}_L = vech'(\boldsymbol{\Sigma}_L)$, by Gaussian PML. Then, they compute the orthogonalised 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 because the correction is unnecessary. 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 et al (2017).

Biases In Table 2, 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 1 and 2.

When the structural shocks follow independent Student t distributions with 5 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 2 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 2 indicates that many of them perform rather well in terms of finite sample biases, with the DLSMN MLE being probably the best one.

Therefore, our Monte Carlo exercises confirm the practical relevance of our consistency results for both the mixture-based PMLEs and the FS two-step estimators.

Efficiency Next, we evaluate the finite sample relative efficiency of the different consistent estimators using the Monte Carlo root mean squared errors (RMSE) in Table 3 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 (2021b) and Appendix D). 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 presumably 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 3 confirms that the relative performance of the Student t and Laplace estimators 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 3 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 **J** and the off-diagonal elements of **C**.

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

Coverage We compute pointwise confidence bands for the IRFs generated by the estimated models using the parametric procedure in Mittnik and Zadrozny (1993) (see Appendix E for further details). Since their procedure relies on the delta method, we first compute the asymptotic variance of the finite normal mixture PMLEs and the FS estimators based on the Student t MLE, as explained at the end of section 2.3. To capture what a researcher who believes in the correct specification of the distribution of the shocks would do, we also compute the asymptotic variance of the Student t MLE using the inverse information matrix. When the shocks follow independent Student t distributions, the confidence bands thus computed would be the narrowest possible, while they will generally be centred around inconsistent point estimates of the IRFs under distributional misspecification.

We first conduct Monte Carlo simulations for a bivariate SVAR(1) process obtained by simply removing the third variable from equation (4), leaving everything else unchanged. We report the results in Table 4 using $10,000$ replications to reduce sampling errors in estimating the coverage proportions. As expected, when the true distribution of the shocks is a Student t, the most accurate coverage is achieved by the Student MLE. Nevertheless, the coverages of the Mixture PMLE and the Student-AFS are reasonably close to their nominal values. In contrast, when the shocks do not follow symmetric Student ts, the Student MLE coverages are wrong. The only notable exception occurs when the shocks follow asymmetric Student innovations, the reason being that in that case the bias of the estimators is absorbed by the drift parameters, as we saw in Table 2. More importantly, the coverage of the mixture-based PMLE and the Student-AFS is reasonable in all instances.

The substantially more detailed results for the three-variate system (4) that we report in Table 5 lead to very similar conclusions.

Number of components The consistency results in Proposition 3 are valid for any finite Gaussian mixture regardless of the number of components $K \geq 2$. At the same time, at the end of section 2.3 we saw that, asymptotically at least, a larger number of components leads to a better approximation to the true distribution of the shocks, which in turn seems to generate efficiency gains. In finite samples, though, things are not so straightforward, as there is a clear trade-off between the accuracy of the approximating mixture and the number of estimated parameters. To shed some light on this trade-off, we have simulated $5,000$ samples of $T = 500$, $T = 2,000$ and $T = 20,000$ observations each from the trivariate Svar(1) process (4), with stochastically independent shocks drawn from three heterogenous asymmetric Student ts. Then, for each of those samples, we estimate the model parameters using a mixture-based log-likelihood function with both $K = 2$ and $K = 3$ components. We report the Monte Carlo root mean squared errors for the usual groups of parameters in Table 6. When the sample length is very large ($T = 20,000$), the estimator based on $K = 3$ is unsurprisingly superior to the one that relies on $K = 2$, which confirms the asymptotic results in Table 1. In contrast, it seems better to use the estimator with $K = 2$ when $T = 500$. In the intermediate case ($T = 2,000$) the more flexible estimator, $K = 3$, tends to be slightly better for most groups of parameters, but not for all of them.

Although these results suggest that practitioners should use a number of components that increases slowly with the sample size, additional research is required before making a more precise and firmer recommendation.

4 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 ν_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 previous sections, 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 our theoretical and simulation findings in previous sections.

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 ε_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.[14] 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 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.

[14] We have not included confidence bands to avoid cluttering the pictures, but they can be easily obtained using the parametric procedure in Mittnik and Zadrozny (1993) explained in Appendix E.

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.

5 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 of all the parameters 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 two-step consistent estimators à la Fiorentini and Sentana (2007, 2019) when the shocks are assumed to follow other non-Gaussian distributions such as the Student t or the Laplace.

Detailed Monte Carlo exercises confirm the practical relevance of our consistency results for both the mixture-based PMLEs and the FS two-step estimators under distributional misspecification, and illustrate the relative efficiency of the mixture-based PMLEs relative to other consistent proposals. Furthermore, they show that the closed-form expressions for the asymptotic covariance matrices of our proposed estimators that we derive in Appendices B and C in combination with the delta method proposed by Mittnik and Zadrozny (1993) lead to reliable confidence bands for the IRFs. Importantly, while we find that mixture-based PMLEs approach the SP efficiency bound as the number of components of the mixtures increases, we also observe that in finite samples practitioners should probably use more than two components only if their sample size is sufficiently large.

In fact, finding the rate at which the number of mixture components should increase as a function of the sample size in a sieves-type procedure to achieve the SP efficiency bound would be a non-trivial but worthwhile extension of our paper. Similarly, it would be interesting to study the finite sample performance of kernel-type SP estimators which use our proposed consistent estimators as initial values for a single BHHH iteration based on the cross-sectionally independent SP efficient score. Although our theoretical results indicate that such SP estimators are (partially) adaptive for the matrices of VAR coefficients $\mathbf{A}_j$ $(j = 1, \ldots, p)$ and the scaled impact multipliers in $\mathbf{J}$, and the cross-sectional independence of the shocks allows the estimation of their densities at univariate non-parametric rates regardless of N, the finite sample performance of these kernel-based estimators could be subpar. A comparison of these two SP estimators with the distribution-free methods in Herwartz (2018), who exploits the proposal in Matteson and Tsay (2017), and Lanne and Luoto (2021), 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.

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.

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 (see Hyvärinen (2013), Amengual et al (2021a,b), Montiel Olea et al (2021) and Davis and Ng (2021) for some proposals). Assessing the non-normality assumption in combination with independence is also fundamental, as is the study of the properties of the different estimators that we consider in near Gaussian situations (see Amengual et al (2021a) and Lee and Mesters (2021) for recent suggestions in those two directions). Nevertheless, the closer the true distribution of the shocks were to the normal, the better the finite Gaussian mixture approximation, and the more efficient our proposed estimators would be relative to the correctly specified parametric ones. 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.

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 (see Amengual et al (2022)). For that reason, it would be interesting to assess the performance of discrete mixture of normals maximum likelihood estimators in those contexts. We are currently exploring some of these interesting research avenues.

References

Amengual, D., Fiorentini, F. and Sentana (2021a): "Moment tests for independent components", forthcoming in *SERIEs*.

Amengual, D., Fiorentini, F. and Sentana (2021b): "Specification tests for non-Gaussian structural vector autoregressions", mimeo, CEMFI.

Amengual, D., Fiorentini, F. and Sentana (2022): "PML vs Minimum χ^2: the comeback", 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. (2020): "Aggregate demand and aggregate supply effects of Covid-19: a real-time analysis", *Covid Economics* 25, 141-168.

Bekaert, G., Engstrom, E. and Ermolov, A. (2021): "Macro risks and the term structure of interest rates", *Journal of Financial Economics* 141, 479-504.

Bernoth, K. and Herwartz; H. (2021): "Exchange rates, foreign currency exposure and sovereign risk", *Journal of International Money and Finance* 117, 102454.

Braun, R. (2021): "The importance of supply and demand for oil prices: evidence from a Svar identified by non-Gaussianity", mimeo, Bank of England.

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

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.

Chen, A. and Bickel, P.J. (2006): "Efficient independent component analysis", *Annals of Statistics* 34, 2825-2855.

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.

Davis, R. and Ng, S. (2021): "Time series estimation of the dynamic effects of disaster shocks", mimeo, Columbia University.

Diebold, F.X. and Yilmaz, 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. (2019): "Consistent non-Gaussian pseudo maximum likelihood estimators", *Journal of Econometrics* 213, 321-358.

Fiorentini, G., and Sentana, E. (2021a): "New testing approaches for mean-variance predictability", *Journal of Econometrics* 222, 516-538.

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

Gallant, A.R. and Tauchen, G. (1999): "The relative efficiency of method of moments estimators", *Journal of Econometrics* 92, 149-172.

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. (2021): "Identification of structural vector autoregressions through higher unconditional moments", *Journal of Econometrics* 225, 27-46.

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

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

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

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

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

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

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

Lee, A. and Mesters, G. (2021): "Robust non-Gaussian inference for linear simultaneous equations models", mimeo, UPF.

Ludvigson, S.C., Ma, S. and Ng, S. (2021): "Uncertainty and business cycles: exogenous impulse or endogenous response?", *American Economic Journal: Macroeconomics* 13: 369-410.

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. (2021): "The Jacobian of the exponential function", *Journal of Economic Dynamics and Control* 127, 104122.

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.

Mittnik, S. and Zadrozny, P. (1993): "Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models", *Econometrica* 61, 857-870.

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.

Montiel Olea, J.L., Plagborg-Møller, M. and Qian, E. (2021): "Svar identification from higher moments: has the simultaneous causality problem been solved?", mimeo, Princeton University.

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. (2020): "Approximation by finite mixtures of continuous density functions that vanish at infinity", *Cogent Mathematics & Statistics* 7, 1750861.

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.

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

Appendices

A Proofs

A.1 Proposition 1

Expression (B7) in Appendix B implies that

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\phi}) \end{bmatrix} = -\begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \end{pmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} \left\{ \begin{array}{c} \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*_1} \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*_N} \end{array} \right\}.$$

In turn, expression (B32) in the same appendix yields

$$s_{\mathbf{j}}(\boldsymbol{\theta};\boldsymbol{\varrho}) = -veco\left[\mathbf{J}^{-1\prime} + \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\left\{ \begin{array}{ccc} \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*_1}\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*_1}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*_N}\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*_N}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \end{array}\right\}\boldsymbol{\Psi}\right],$$

where $\boldsymbol{\Delta}_N$ is an $N^2 \times N(N-1)$ matrix such that $vec(\mathbf{J}-\mathbf{I}_N) = \boldsymbol{\Delta}_N veco(\mathbf{J}-\mathbf{I}_N)$ (see Magnus and Sentana (2020)), and

$$s_{\boldsymbol{\psi}}(\boldsymbol{\theta};\boldsymbol{\varrho}) = -\boldsymbol{\Psi}^{-1}\left\{ \begin{array}{c} 1+\frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon^*_1}\varepsilon^*_{1t}(\boldsymbol{\theta}) \\ \vdots \\ 1+\frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon^*_N}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \end{array}\right\}.$$

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

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

Next, define the pseudo true values of the parameters $\boldsymbol{\upsilon}_\infty(\bar{\boldsymbol{\varrho}})$, $\boldsymbol{\tau}_\infty(\bar{\boldsymbol{\varrho}}) = \mathbf{J}_0\boldsymbol{\upsilon}_\infty(\bar{\boldsymbol{\varrho}})$ and $\boldsymbol{\psi}_\infty(\bar{\boldsymbol{\varrho}})$ such that for $i = 1,\ldots,N$

$$E\left[\frac{\partial \ln f_i\{\varepsilon^*_{it}[\boldsymbol{\tau}_\infty(\bar{\boldsymbol{\varrho}}),\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty(\bar{\boldsymbol{\varrho}})];\bar{\boldsymbol{\varrho}}_i\}}{\partial\varepsilon^*_i}\right]$$
$$= E\left\{\frac{\partial \ln f_i[\psi^{-1}_{i\infty}(\bar{\boldsymbol{\varrho}}_i)\{[\upsilon_0 - \upsilon_\infty(\bar{\boldsymbol{\varrho}}_i)] + \psi_{i0}\varepsilon^*_{it}\};\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*_i}\right\} = 0, \quad \text{(A1)}$$

$$E\left[1+\frac{\partial \ln f\{\varepsilon^*_{it}(\boldsymbol{\tau}_\infty(\bar{\boldsymbol{\varrho}}),\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty(\bar{\boldsymbol{\varrho}}));\bar{\boldsymbol{\varrho}}_i\}}{\partial\varepsilon^*}\varepsilon^*_{it}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty)\right]$$
$$= E\left\{1+\frac{\partial \ln f[\psi^{-1}_{i\infty}(\bar{\boldsymbol{\varrho}}_i)\{[\upsilon_0-\upsilon_\infty(\bar{\boldsymbol{\varrho}}_i)]+\psi_{i0}\varepsilon^*_{it}\};\bar{\boldsymbol{\varrho}}_i]}{\partial\varepsilon^*}\psi^{-1}_{i\infty}(\bar{\boldsymbol{\varrho}}_i)\{[\upsilon_0-\upsilon_\infty(\bar{\boldsymbol{\varrho}}_i)]+\psi_{i0}\varepsilon^*_{it}\}\right\}=0, \quad \text{(A2)}$$

so that the expected value of the scores of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ are 0. 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 possibly $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ will be consistently estimated. The consistency of $\mathbf{a}$ but not $\mathbf{j}$ also follows from Proposition 12 in the supplemental appendix of Fiorentini and Sentana (2019).

When $\boldsymbol{\varrho}$ is simultaneously estimated, one should understand the solutions $\boldsymbol{\tau}_\infty$ and $\boldsymbol{\psi}_\infty$ to the equations (A1) and (A2) above as functions of $\boldsymbol{\varrho}$, and add the scores for these shape parameters as additional model conditions, which implicitly define their pseudo true values $\boldsymbol{\varrho}_\infty$. □

A.2 Proposition 2

The proof is very similar to the proof of Proposition 1. As we explain in Appendix B, the main difference is when the assumed distributions of all the structural shocks are symmetric, the score expressions simplify to

$$
\begin{aligned}
s_{\boldsymbol{\tau}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\left\{\begin{array}{c}\delta[\varepsilon^{*2}_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{1t}(\boldsymbol{\theta})\\ \vdots \\ \delta[\varepsilon^{*2}_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]\varepsilon^*_{Nt}(\boldsymbol{\theta})\end{array}\right\},\\
s_{\mathbf{a}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= \begin{pmatrix}\mathbf{y}_{t-1}\otimes\mathbf{I}_N\\ \vdots\\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N\end{pmatrix}\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\left\{\begin{array}{c}\delta[\varepsilon^{*2}_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{1t}(\boldsymbol{\theta})\\ \vdots \\ \delta[\varepsilon^{*2}_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]\varepsilon^*_{Nt}(\boldsymbol{\theta})\end{array}\right\},\\
s_{\mathbf{j}}(\boldsymbol{\theta};\boldsymbol{\varrho}) &= veco\left[\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\left\{\begin{array}{ccc}\delta[\varepsilon^{*2}_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{1t}(\boldsymbol{\theta})-1 & \dots & \delta[\varepsilon^{*2}_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{1t}(\boldsymbol{\theta})\varepsilon^*_{Nt}(\boldsymbol{\theta})\\ \vdots & \ddots & \vdots\\ \delta[\varepsilon^{*2}_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^*_{Nt}(\boldsymbol{\theta})\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & \delta[\varepsilon^{*2}_{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^{*2}_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{1t}(\boldsymbol{\theta})-1\\ \vdots \\ \delta[\varepsilon^{*2}_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]\varepsilon^{*2}_{Nt}(\boldsymbol{\theta})-1\end{array}\right\},
\end{aligned}
$$

where

$$
\delta[\varepsilon^{*2}_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i] = -2\frac{\partial\ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial\varepsilon^{*2}_i} \tag{A3}
$$

is a scalar function of $\varepsilon^{*2}_{it}(\boldsymbol{\theta})$ due to the symmetry of the assumed univariate distributions.

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 $\bar{\boldsymbol{\varrho}}$,

$$
E\{\delta[\varepsilon^{*2}_{it}(\boldsymbol{\tau}_0,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi});\bar{\boldsymbol{\varrho}}_i]\varepsilon^*_{it}(\boldsymbol{\tau}_0,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi})\} = E[\delta(\psi_i^{-1}\psi_{i0}\boldsymbol{\varepsilon}^*{}_{it};\bar{\boldsymbol{\varrho}}_i)\psi_i^{-1}\psi_{i0}\varepsilon^*_{it}] = 0 \quad \forall i \tag{A4}
$$

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 regardless of the value of $\boldsymbol{\psi}$, and the same applies to the scores of $\mathbf{a}$ because of the law of iterated expectations.

Next, let us define $\boldsymbol{\psi}_\infty(\bar{\boldsymbol{\varrho}})$ such that

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

so that the expected value of the score of $\boldsymbol{\psi}$ is 0. The cross-sectional independence of the true shocks combined with (A4) implies that the expected value of the scores of $\mathbf{j}$ will also be 0. Consequently, all parameters except possibly $\boldsymbol{\psi}$ 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 $\boldsymbol{\varrho}$, and add the scores for these shape parameters as additional model conditions, which implicitly define their pseudo true values $\boldsymbol{\varrho}_\infty$. $\square$

A.3 Proposition 3

Proposition 1 implies that $\mathbf{a}$ and $\mathbf{j}$ will be consistently estimated, so we only need to focus on $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$. Let

$$\boldsymbol{\varepsilon}_t^*(\mathbf{0}, \mathbf{a}_0, \mathbf{j}_0, \ell_N) = \mathbf{J}_0^{-1}(\mathbf{y}_t - \mathbf{A}_{10}\mathbf{y}_{t-1} - \ldots - \mathbf{A}_{p0}\mathbf{y}_{t-p}) = \boldsymbol{\upsilon}_0 + \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t^*,$$

whose mean vector and covariance matrix are $\boldsymbol{\upsilon}_0$ and $\boldsymbol{\Psi}_0$, respectively, where $\boldsymbol{\upsilon}_0 = \mathbf{J}_0^{-1}\boldsymbol{\tau}_0$. The usual numerical invariance property of ML estimators to bijective reparametrisations allows us to replace υ_i, ψ_i and $\boldsymbol{\varrho}_i$ with the natural parametrisation of finite mixtures of normals discussed in Appendix F.1. But then, Proposition F1 in that appendix applied to each shock implies that a discrete mixture log-likelihood function will consistently estimate υ_{i0} and ψ_{i0} for $i = 1, \ldots, N$ regardless of the true distribution of the shocks. $\square$

A.4 Proposition 4

Proposition 2 implies that $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ will be consistently estimated, so we only need to focus on $\boldsymbol{\psi}$. Let

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

whose mean vector and covariance matrix are $\mathbf{0}$ and $\boldsymbol{\Psi}_0$, respectively. The usual numerical invariance property of ML estimators to bijective reparametrisations allows us to replace ψ_i and $\boldsymbol{\varrho}_i$ with the natural parametrisation of finite scale mixtures of normals discussed in Appendix F.2. But then, Proposition F2 in that appendix applied to each shock implies that a discrete scale mixture log-likelihood function will consistently estimate ψ_{i0} for $i = 1, \ldots, N$ regardless of the true distribution of the shocks as long as they are symmetric. $\square$

A.5 Proposition 5

To prove the first part of the proposition, let us group the scores with respect to $\boldsymbol{\theta}$ into two sets: $\mathbf{a}$ and $\mathbf{j}$ on the one hand, and $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ on the other. Thus, we end up with

$$\begin{aligned}
\mathbf{s}_{\mathbf{a}t}(\boldsymbol{\phi}) &= \begin{bmatrix} (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} \\ \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} \end{bmatrix}\mathbf{e}_{lt}(\boldsymbol{\phi}), \\
\mathbf{s}_{\mathbf{j}t}(\boldsymbol{\phi}) &= \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\begin{pmatrix} \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix}\begin{bmatrix} \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi}) &= \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}), \\
\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\phi}) &= \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}),
\end{aligned}$$

where we have used Proposition 4 in Magnus and Sentana (2020), which says that

$$\mathbf{E}_N\mathbf{E}_N' + \boldsymbol{\Delta}_N\boldsymbol{\Delta}_N' = \begin{pmatrix} \mathbf{E}_N' \\ \boldsymbol{\Delta}_N' \end{pmatrix}\begin{pmatrix} \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix} = \mathbf{I}_{N^2}, \tag{A5}$$

with $\mathbf{E}_N$ being the $N^2\times N$ matrix such that $vec(\boldsymbol{\Psi}) = \mathbf{E}_N vecd(\boldsymbol{\Psi})$ for any diagonal matrix $\boldsymbol{\Psi}$. We can use expression (B30) in Appendix B.5.2 to show that the unconditional covariance matrix between these two sets of scores is

$$E\left\{\begin{bmatrix} \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\mathbf{j}t}(\boldsymbol{\phi}) \end{bmatrix}\begin{bmatrix} \mathbf{s}_{\boldsymbol{\tau}t}'(\boldsymbol{\phi}) & \mathbf{s}_{\boldsymbol{\psi}t}'(\boldsymbol{\phi}) \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N \end{bmatrix}\right.$$
$$\left.\begin{bmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} & \mathbf{0} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N \end{bmatrix}\begin{pmatrix} \boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right\}$$

$$= \begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N \end{bmatrix}$$
$$\times\begin{bmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} & \mathbf{0} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N \end{bmatrix}\begin{pmatrix} \boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$= \begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathcal{M}_{ll}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathrm{M}_{ls}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathcal{M}_{ll}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathcal{M}_{ll}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} \\ \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\mathrm{M}_{ls}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\mathrm{M}_{ss}\boldsymbol{\Psi}^{-1} \end{bmatrix},$$

where $\mathcal{M}_{ll}$, M_{ls} and M_{ss} are diagonal matrices with typical elements $V[\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_{i0})/\partial\varepsilon^*_i|\boldsymbol{\phi}_0]$, $cov[\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_{i0})/\partial\varepsilon^*_i,\varepsilon^*_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_{i0})/\partial\varepsilon^*_i|\boldsymbol{\phi}_0]$ and $V[\varepsilon^*_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_{i0})/\partial\varepsilon^*_i|\boldsymbol{\phi}_0]$, respectively.

In turn, the unconditional covariance matrix of the scores for $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ is

$$\begin{pmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \end{pmatrix}\begin{pmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathcal{M}_{ll}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathrm{M}_{ls}\boldsymbol{\Psi}^{-1} \\ \boldsymbol{\Psi}^{-1}\mathrm{M}_{ls}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \boldsymbol{\Psi}^{-1}\mathrm{M}_{ss}\boldsymbol{\Psi}^{-1} \end{pmatrix},$$

whose inverse is given by

$$\begin{pmatrix} \mathbf{J}\boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}\begin{pmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\Psi}\mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}.$$

As a result, the coefficients in the least squares projection of the scores of $\mathbf{a}$ and $\mathbf{j}$ onto the linear span of the scores of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ will be given by

$$\begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N \end{bmatrix}$$
$$\times\begin{bmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} & \mathbf{0} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N \end{bmatrix}\begin{pmatrix} \boldsymbol{\Psi}^{-1}\mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$
$$\times\begin{pmatrix} \mathbf{J}\boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}\begin{pmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\Psi}\mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}$$
$$=\begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N \end{bmatrix}$$
$$\times\begin{bmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{pmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls} & \mathrm{M}_{ss} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\Psi}\mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}$$
$$=\begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N \end{bmatrix}\begin{pmatrix} \boldsymbol{\Psi}\mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}$$
$$=\begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N) & \mathbf{0} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\boldsymbol{\Psi} \end{bmatrix},$$

so that the corresponding projections are

$$\begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N) & \mathbf{0} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\boldsymbol{\Psi} \end{bmatrix}\begin{pmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1} \end{pmatrix}\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix}$$

$$= \begin{bmatrix} (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N \end{bmatrix}\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix}.$$

Hence, the projection errors will be

$$\begin{bmatrix} \mathbf{s}_{\mathbf{a}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\mathbf{j}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi}) \end{bmatrix} = \left\{ \begin{array}{c} [(\mathbf{y}_{t-1}-\boldsymbol{\mu})\otimes\mathbf{I}_N]\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \vdots \\ [(\mathbf{y}_{t-p}-\boldsymbol{\mu})\otimes\mathbf{I}_N]\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{array} \right\}.$$

The last projection error is conditionally orthogonal to $\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ in view of (D44). The projection error in the first block, though, are not conditionally orthogonal to $\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})$ or $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$, but they are unconditionally orthogonal because $\boldsymbol{\mu}=E(\mathbf{y}_{t-j})$ for all j under covariance stationarity. As a result, the cross-sectionally independent SP estimators of $\mathbf{a}$ and $\mathbf{j}$ will indeed be partially adaptive.

Interestingly, note that $\mathbf{s}_{\mathbf{a}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi})$ and $\mathbf{s}_{\mathbf{j}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi})$ are conditionally orthogonal, so their unconditional covariance matrix will be block diagonal, with blocks

$$\begin{aligned} & V\left[\mathbf{s}_{\mathbf{a}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi})\right] \\ = \; & E\left\{\left[\begin{pmatrix} \mathbf{y}_{t-1}-\boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p}-\boldsymbol{\mu} \end{pmatrix}\otimes\mathbf{I}_N\right]\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathcal{M}_{ll}\boldsymbol{\Psi}^{-1}\mathbf{J}^{-1\prime}\left[\begin{pmatrix} \mathbf{y}_{t-1}-\boldsymbol{\mu} & \vdots & \mathbf{y}_{t-p}-\boldsymbol{\mu} \end{pmatrix}\otimes\mathbf{I}_N\right]\right\} \end{aligned}$$

and

$$V\left[\mathbf{s}_{\mathbf{j}|\boldsymbol{\tau},\boldsymbol{\psi}t}(\boldsymbol{\phi})\right] = \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N.$$

As a result, the restricted and unrestricted ML estimators of these parameters will be asymptotically independent. In general, though, the restricted and unrestricted ML estimators of $\mathbf{a}$ will be correlated with the corresponding estimators of $\boldsymbol{\tau}$ while the restricted and unrestricted ML estimators of $\mathbf{j}$ will be correlated with the corresponding estimators of $\boldsymbol{\psi}$.

To prove the second part, let us apply the Jacobian $\partial\mathbf{c}'/\partial\boldsymbol{\psi}=\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}')$ to the cross-sectionally independent SP efficient score for $\mathbf{c}$ that appear in (D45) in Appendix D. Specifically, if we define K_{ls} and K_{ss} as diagonal matrices of order N with typical elements $\varphi(\boldsymbol{\varrho}_i)=E(\varepsilon_{it}^{*3}|\boldsymbol{\varrho}_i)$

and $\kappa_{ii}(\boldsymbol{\varrho}_i) - 1 = E(\varepsilon_{it}^{*4}|\boldsymbol{\varrho}_i) - 1$, respectively, it is then easy to see that the cross-sectionally independent SP efficient score for $\boldsymbol{\psi}$ will be

$$\begin{aligned}
& \mathbf{E}_N'(\mathbf{I}_N \otimes \mathbf{J}')\{(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})\mathbf{E}_N[-2\mathrm{K}_{ls}(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0}) + 2(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})] \\
& +(\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\} \\
= \; & \mathbf{E}_N'(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1})\mathbf{E}_N[-2\mathrm{K}_{ls}(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0}) + 2(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})] \\
& +\mathbf{E}_N'(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\} \\
= \; & \boldsymbol{\Psi}^{-1}[-2\mathrm{K}_{ls}(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0}) + 2(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})]
\end{aligned}$$

because $\mathbf{E}_N'(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1})\mathbf{E}_N = \boldsymbol{\Psi}^{-1}$ and $\mathbf{E}_N'(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N = \mathbf{0}$ in view of Propositions 2 and 6 of Magnus and Sentana (2020).

Given that the cross-sectionally independent SP efficient score for $\boldsymbol{\tau}$ is

$$\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}[\mathrm{K}_{ss}(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0}) - \mathrm{K}_{ls}(\mathrm{K}_{ss} - \mathrm{K}_{ls}^2)^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})],$$

it is clear that the iterated version of the cross-sectionally independent SP estimators will satisfy

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_{lt}(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}}; \mathbf{0}) &= \mathbf{0}, \\
\frac{1}{T}\sum_{t=1}^{T}\mathbf{E}_N'\mathbf{e}_{st}(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}}; \mathbf{0}) &= \mathbf{0}.
\end{aligned}$$

But this means that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^*(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}}) &= \mathbf{0}, \\
\frac{1}{T}\sum_{t=1}^{T} vecd[\boldsymbol{\varepsilon}_t^*(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}})\boldsymbol{\varepsilon}_t^{*\prime}(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}}) - \mathbf{I}_N] &= \mathbf{0},
\end{aligned}$$

where

$$\boldsymbol{\varepsilon}_t^*(\ddddot{\boldsymbol{\tau}}, \ddddot{\mathbf{a}}, \ddddot{\mathbf{j}}, \ddddot{\boldsymbol{\psi}}) = \ddddot{\boldsymbol{\Psi}}^{-1}\ddddot{\mathbf{J}}^{-1}(\mathbf{y}_t - \ddddot{\boldsymbol{\tau}} - \ddddot{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \ddddot{\mathbf{A}}_p\mathbf{y}_{t-p}),$$

so they coincide with the second-step estimators of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ in the FS procedure but with the first-step estimators of $\mathbf{a}$ and $\mathbf{j}$ being the (iterated) cross-sectionally independent SP estimators $\ddddot{\mathbf{a}}$ and $\ddddot{\mathbf{j}}$. □

A.6 Proposition 6

The proof of the first part is very similar to the proof Proposition 5, but in this instance the two groups in which we split the scores correspond to $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ on the one hand, and $\boldsymbol{\psi}$ on the other.

The unconditional covariance matrix between these two sets of scores when all the shocks are symmetrically distributed is

$$E\left\{\left[\begin{array}{c}\mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi})\\ \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\phi})\\ \mathbf{s}_{\mathbf{j}t}(\boldsymbol{\phi})\end{array}\right]\mathbf{s}'_{\boldsymbol{\psi}t}(\boldsymbol{\phi})\right\}$$

$$= E\left\{\left[\begin{array}{ccc}\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots\\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\end{array}\right]\right.$$

$$\left.\left[\begin{array}{ccc}\mathcal{M}_{ll} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathrm{M}_{ss} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N\end{array}\right]\left(\begin{array}{c}\mathbf{0}\\ \boldsymbol{\Psi}^{-1}\\ \mathbf{0}\end{array}\right)\right\}$$

$$= \left[\begin{array}{ccc}\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots\\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\end{array}\right]$$

$$\times\left(\begin{array}{c}\mathbf{0}\\ \mathrm{M}_{ss}\boldsymbol{\Psi}^{-1}\\ \mathbf{0}\end{array}\right)$$

because $\mathrm{M}_{ls}=\mathbf{0}$. In turn, the covariance matrix of $\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\phi})$ is simply $\boldsymbol{\Psi}^{-1}\mathrm{M}_{ss}\boldsymbol{\Psi}^{-1}$, so the regression coefficients are

$$\left[\begin{array}{ccc}\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots\\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\end{array}\right]$$

$$\times\left(\begin{array}{c}\mathbf{0}\\ \mathrm{M}_{ss}\boldsymbol{\Psi}^{-1}\\ \mathbf{0}\end{array}\right)\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi} = \left[\begin{array}{c}\mathbf{0}\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\\ \boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\boldsymbol{\Psi}\end{array}\right].$$

Hence, the projection errors become

$$\begin{bmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N \end{bmatrix}$$
$$\times\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix}$$
$$-\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\boldsymbol{\Psi} \end{bmatrix}\boldsymbol{\Psi}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix}.$$

All these scores are conditionally orthogonal to $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ when all the shock distributions are symmetric, which proves that the cross-sectionally independent symmetric SP estimators of $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ are partially efficient in that case.

To prove the second part, we apply the Jacobian $\partial\mathbf{c}'/\partial\boldsymbol{\psi} = \mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}')$ to the cross-sectionally independent symmetric SP efficient score for $\mathbf{c}$ in (D49), which leads to

$$\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}')(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1})\mathbf{E}_N 2\mathrm{K}_{ss}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$$
$$+\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}')(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) = \boldsymbol{\Psi}^{-1}2\mathrm{K}_{ss}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0}).$$

As a result, the iterated version of the cross-sectionally independent symmetric SP efficient estimators will be such that

$$\frac{1}{T}\sum_{t=1}^{T} vecd[\boldsymbol{\varepsilon}_t^*(\ddot{\boldsymbol{\tau}},\ddot{\mathbf{a}},\dddot{\mathbf{j}},\dddot{\boldsymbol{\psi}})\boldsymbol{\varepsilon}_t^{*\prime}(\ddot{\boldsymbol{\tau}},\ddot{\mathbf{a}},\dddot{\mathbf{j}},\dddot{\boldsymbol{\psi}}) - \mathbf{I}_N] = \mathbf{0},$$

so the estimator of $\boldsymbol{\psi}$ coincides with the second-step estimators in the symmetric version of the FS procedure, but with the first-step estimators of $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ being the (iterated) cross-sectionally independent symmetric SP estimators $\ddot{\boldsymbol{\tau}}$, $\ddot{\mathbf{a}}$ and $\dddot{\mathbf{j}}$. □

TABLE 1: Asymptotic variance of estimators. Bivariate VAR(1).

DGP:	$\varepsilon_{1t} \sim$ Student t_5				$\varepsilon_{2t} \sim$ Student t_5	
	$K=2$	$K=3$	$K=4$	$K=5$	Student-AFS	Student
τ_1	1.001	1.001	1.001	1.001	1.001	0.800
τ_1	1.040	1.040	1.040	1.040	1.040	0.832
A_{11}	0.632	0.622	0.621	0.621	0.620	0.620
A_{21}	0.656	0.646	0.645	0.645	0.645	0.645
A_{12}	0.613	0.604	0.603	0.602	0.602	0.602
A_{22}	0.637	0.627	0.626	0.626	0.626	0.626
ψ_1	2.089	2.093	2.095	2.077	2.093	1.481
ψ_2	2.048	2.041	2.039	2.037	2.041	1.572
J_{21}	2.409	2.240	2.221	2.218	2.218	2.218
J_{12}	2.413	2.244	2.224	2.221	2.221	2.221

$K=k$: PMLE based on finite mixture of k normals; Student-AFS: General FS correction to the Student t MLE; Student: Student t MLE. Pseudo true values obtained with $T=40,000,000$.

TABLE 2: Monte Carlo results. Mean absolute bias of pooled groups of estimators ($T = 500$).

DGP: $\varepsilon_{1t} \sim$ Student t_5 $\varepsilon_{2t} \sim$ Student t_5 $\varepsilon_{3t} \sim$ Student t_5

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0004	0.0003	0.0003	0.0006	0.0003	0.0006	0.0003	0.0003	0.0003	0.0003
A_{ii}	0.0053	0.0053	0.0053	0.0053	0.0053	0.0048	0.0048	0.0066	0.0066	0.0066
$A_{ij,i\neq j}$	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0010	0.0010	0.0010
C_{ii}	0.0048	0.0107	0.0118	0.0112	0.0115	0.0239	0.0108	0.0114	0.0150	0.0125
$C_{ij,i>j}$	0.0006	0.0018	0.0019	0.0020	0.0020	0.0050	0.0020	0.0019	0.0029	0.0021
$C_{ij,i<j}$	0.0003	0.0003	0.0003	0.0002	0.0002	0.0006	0.0006	0.0004	0.0004	0.0004
$J_{ij,i>j}$	0.0003	0.0003	0.0004	0.0004	0.0004	0.0005	0.0005	0.0003	0.0002	0.0003
$J_{ij,i<j}$	0.0009	0.0009	0.0016	0.0012	0.0012	0.0007	0.0007	0.0010	0.0014	0.0015

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
$\boldsymbol{\tau}$	0.0003	0.0004	0.0004	0.0003	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003
A_{ii}	0.0044	0.0044	0.0041	0.0041	0.0041	0.0037	0.0037	0.0065	0.0065	0.0065
$A_{ij,i\neq j}$	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0007	0.0007	0.0007
C_{ii}	0.1109	0.0066	0.0061	0.0057	0.0061	0.0075	0.0052	0.0077	0.0073	0.0076
$C_{ij,i>j}$	0.0229	0.0009	0.0011	0.0010	0.0011	0.0013	0.0008	0.0010	0.0010	0.0014
$C_{ij,i<j}$	0.0007	0.0006	0.0003	0.0004	0.0004	0.0004	0.0003	0.0006	0.0009	0.0003
$J_{ij,i>j}$	0.0009	0.0009	0.0004	0.0005	0.0005	0.0007	0.0007	0.0010	0.0011	0.0006
$J_{ij,i<j}$	0.0014	0.0014	0.0008	0.0010	0.0010	0.0009	0.0009	0.0014	0.0016	0.0010

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
$\boldsymbol{\tau}$	0.1792	0.0003	0.0003	0.2607	0.0003	0.1970	0.0003	0.0003	0.0003	0.0003
A_{ii}	0.0040	0.0040	0.0028	0.0032	0.0032	0.0037	0.0037	0.0066	0.0066	0.0066
$A_{ij,i\neq j}$	0.0002	0.0002	0.0001	0.0003	0.0003	0.0005	0.0005	0.0005	0.0005	0.0005
C_{ii}	3.2478	0.0059	0.0043	0.0367	0.0048	0.0104	0.0058	0.0100	0.0248	0.0064
$C_{ij,i>j}$	0.8004	0.0011	0.0011	0.0100	0.0008	0.0013	0.0013	0.0018	0.0123	0.0011
$C_{ij,i<j}$	0.0060	0.0016	0.0005	0.0013	0.0012	0.0016	0.0015	0.0016	0.0147	0.0006
$J_{ij,i>j}$	0.0002	0.0002	0.0005	0.0004	0.0004	0.0003	0.0003	0.0012	0.0147	0.0001
$J_{ij,i<j}$	0.0014	0.0014	0.0005	0.0010	0.0010	0.0016	0.0016	0.0016	0.0139	0.0005

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
$\boldsymbol{\tau}$	0.1220	0.0008	0.0008	0.1091	0.0008	0.1634	0.0010	0.0008	0.0008	0.0008
A_{ii}	0.0057	0.0057	0.0050	0.0057	0.0057	0.0057	0.0057	0.0073	0.0073	0.0073
$A_{ij,i\neq j}$	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	0.0006	0.0006	0.0006
C_{ii}	0.0123	0.0103	0.0074	0.0055	0.0097	0.0290	0.0131	0.0129	0.0393	0.0085
$C_{ij,i>j}$	0.0009	0.0010	0.0009	0.0009	0.0015	0.0119	0.0013	0.0018	0.0079	0.0015
$C_{ij,i<j}$	0.0003	0.0003	0.0005	0.0008	0.0008	0.0011	0.0012	0.0007	0.0014	0.0004
$J_{ij,i>j}$	0.0010	0.0010	0.0006	0.0008	0.0008	0.0027	0.0027	0.0007	0.0022	0.0002
$J_{ij,i<j}$	0.0016	0.0016	0.0003	0.0015	0.0015	0.0026	0.0026	0.0028	0.0094	0.0005

Sample length=500, Replications=5,000. S: Student-t MLE, M: DLSMN MLE, SM: DSMN MLE, L: Laplace MLE, IC: GMR two step estimator, AFS: general FS correction.

TABLE 3: Monte Carlo results. (RMSE) of pooled groups of estimators ($T = 500$).

DGP: $\varepsilon_{1t} \sim$ Student t_5 $\varepsilon_{2t} \sim$ Student t_5 $\varepsilon_{3t} \sim$ Student t_5

	S	S_{AFS}	M	SM	SM_{AFS}	L	L_{AFS}	IC-S	IC-L	IC-M
$\boldsymbol{\tau}$	0.0433	0.0479	0.0480	0.0438	0.0480	0.0488	0.0480	0.0484	0.0484	0.0484
A_{ii}	0.0372	0.0372	0.0379	0.0377	0.0377	0.0420	0.0420	0.0413	0.0413	0.0413
$A_{ij,i\neq j}$	0.0370	0.0370	0.0376	0.0375	0.0375	0.0420	0.0420	0.0410	0.0410	0.0410
C_{ii}	0.0607	0.0617	0.0627	0.0623	0.0625	0.0528	0.0620	0.0618	0.0646	0.0629
$C_{ij,i>j}$	0.0760	0.0756	0.0820	0.0808	0.0808	0.0878	0.0847	0.0764	0.0973	0.0829
$C_{ij,i<j}$	0.0741	0.0736	0.0804	0.0790	0.0790	0.0868	0.0834	0.0744	0.0953	0.0811
$J_{ij,i>j}$	0.0772	0.0772	0.0841	0.0828	0.0828	0.0863	0.0863	0.0782	0.1008	0.0853
$J_{ij,i<j}$	0.0763	0.0763	0.0845	0.0825	0.0825	0.0860	0.0860	0.0774	0.0991	0.0857

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
$\boldsymbol{\tau}$	0.0384	0.0471	0.0470	0.0385	0.0470	0.0363	0.0468	0.0481	0.0481	0.0481
A_{ii}	0.0328	0.0328	0.0332	0.0331	0.0331	0.0317	0.0317	0.0405	0.0405	0.0405
$A_{ij,i\neq j}$	0.0330	0.0330	0.0334	0.0333	0.0333	0.0320	0.0320	0.0406	0.0406	0.0406
C_{ii}	0.2278	0.0519	0.0517	0.0517	0.0517	0.0460	0.0512	0.0520	0.0516	0.0519
$C_{ij,i>j}$	0.0776	0.0510	0.0516	0.0513	0.0513	0.0450	0.0453	0.0533	0.0497	0.0535
$C_{ij,i<j}$	0.0555	0.0497	0.0500	0.0498	0.0498	0.0439	0.0439	0.0521	0.0482	0.0518
$J_{ij,i>j}$	0.0505	0.0505	0.0511	0.0508	0.0508	0.0447	0.0447	0.0530	0.0493	0.0531
$J_{ij,i<j}$	0.0509	0.0509	0.0510	0.0507	0.0507	0.0448	0.0448	0.0533	0.0492	0.0530

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
$\boldsymbol{\tau}$	0.2317	0.0474	0.0469	0.3131	0.0470	0.2468	0.0475	0.0486	0.0486	0.0486
A_{ii}	0.0311	0.0311	0.0264	0.0281	0.0281	0.0336	0.0336	0.0405	0.0405	0.0405
$A_{ij,i\neq j}$	0.0310	0.0310	0.0264	0.0280	0.0280	0.0336	0.0336	0.0405	0.0405	0.0405
C_{ii}	5.6613	0.0456	0.0452	0.0682	0.0453	0.0492	0.0457	0.0479	0.0617	0.0453
$C_{ij,i>j}$	1.2843	0.0465	0.0388	0.0452	0.0415	0.0550	0.0547	0.0704	0.1362	0.0430
$C_{ij,i<j}$	0.2174	0.0414	0.0273	0.0319	0.0312	0.0445	0.0439	0.0712	0.1368	0.0297
$J_{ij,i>j}$	0.0462	0.0462	0.0381	0.0409	0.0409	0.0546	0.0546	0.0711	0.1441	0.0426
$J_{ij,i<j}$	0.0393	0.0393	0.0267	0.0302	0.0302	0.0421	0.0421	0.0744	0.1611	0.0309

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
$\boldsymbol{\tau}$	0.1466	0.0481	0.0478	0.1332	0.0481	0.1864	0.0484	0.0486	0.0486	0.0486
A_{ii}	0.0371	0.0371	0.0347	0.0367	0.0367	0.0438	0.0438	0.0411	0.0411	0.0411
$A_{ij,i\neq j}$	0.0370	0.0370	0.0345	0.0365	0.0365	0.0440	0.0440	0.0410	0.0410	0.0410
C_{ii}	0.0743	0.0645	0.0622	0.0653	0.0640	0.0622	0.0672	0.0658	0.0962	0.0621
$C_{ij,i>j}$	0.0729	0.0728	0.0526	0.0692	0.0690	0.1034	0.0967	0.0848	0.1811	0.0530
$C_{ij,i<j}$	0.0783	0.0771	0.0594	0.0749	0.0743	0.1043	0.1022	0.0875	0.1795	0.0599
$J_{ij,i>j}$	0.0741	0.0741	0.0521	0.0697	0.0697	0.0996	0.0996	0.0874	0.2078	0.0526
$J_{ij,i<j}$	0.0846	0.0846	0.0642	0.0825	0.0825	0.1103	0.1103	0.0961	0.2106	0.0647

Sample length=500, Replications=5,000. S: Student-t MLE, M: DLSMN MLE, SM: DSMN MLE, L: Laplace MLE, IC: GMR two step estimator, AFS: general FS correction.

TABLE 4: Coverage of Impulse Response Functions. Nominal 90%. $T = 500$.

	Mixture				**Student-AFS**				**Student**			
Lag	(1,1)	(2,1)	(1,2)	(2,2)	(1,1)	(2,1)	(1,2)	(2,2)	(1,1)	(2,1)	(1,2)	(2,2)
				DGP: $\varepsilon_{1t} \sim$ Student t_5				$\varepsilon_{2t} \sim$ Student t_5				
0	84.5	84.7	84.9	85.3	84.7	87.7	88.6	85.5	88.8	87.0	87.8	89.2
1	87.3	87.1	87.2	87.4	88.0	89.0	89.0	88.6	89.5	88.6	88.9	89.8
2	87.0	87.8	87.5	87.2	88.0	89.2	88.8	88.7	88.9	89.2	89.1	89.1
3	86.5	87.8	87.1	87.1	87.5	88.6	88.2	88.4	88.0	88.6	88.7	88.9
4	86.5	87.5	86.7	86.8	87.5	88.2	88.0	88.1	87.7	88.3	88.2	88.3
5	86.2	87.0	86.4	86.7	87.4	88.0	87.6	87.6	87.6	88.1	88.0	87.8
6	86.0	86.7	86.2	86.6	87.0	87.4	87.4	87.1	87.3	87.7	87.8	87.5
				DGP: $\varepsilon_{1t} \sim$ Laplace				$\varepsilon_{2t} \sim$ Laplace				
0	88.4	85.0	85.5	88.3	88.4	88.5	88.9	88.4	98.2	84.0	86.6	98.0
1	87.0	87.0	87.5	87.4	88.3	89.6	89.3	88.8	96.1	90.5	91.9	95.8
2	86.6	87.7	87.5	87.3	88.6	89.3	89.2	88.8	94.2	92.4	92.5	94.0
3	86.9	87.6	87.0	86.8	88.4	89.3	88.7	88.1	93.2	92.6	92.7	93.5
4	86.6	87.2	86.9	86.4	88.1	88.8	88.0	87.9	92.9	92.5	92.5	93.0
5	86.5	86.6	86.6	86.1	87.6	88.0	87.7	87.6	92.5	92.5	92.3	92.5
6	86.2	86.3	86.3	86.0	87.4	87.6	87.3	87.2	92.0	92.1	92.0	92.3
			DGP: $\varepsilon_{1t} \sim$ DLSMN$(0.8, 0.06, 0.52)$					$\varepsilon_{2t} \sim$ DLSMN$(1.2, 0.08, 0.4)$				
0	89.2	88.2	88.8	89.7	89.1	88.7	89.0	89.8	98.1	96.0	92.3	99.2
1	88.4	88.3	89.0	88.5	88.4	89.0	89.3	88.9	97.1	98.9	95.8	99.2
2	87.8	88.5	88.7	88.7	88.2	88.9	88.7	88.9	96.6	99.1	96.6	99.1
3	88.0	88.3	88.7	88.1	87.9	88.6	88.4	88.9	96.6	99.2	96.8	99.2
4	87.8	87.8	88.2	87.9	87.8	88.5	88.4	88.7	96.8	99.4	96.9	99.3
5	87.9	87.8	88.1	87.8	87.9	88.2	88.1	88.4	96.7	99.4	96.9	99.4
6	87.9	87.6	88.1	87.6	87.5	87.9	87.9	88.0	96.8	99.4	97.0	99.4
				DGP: $\varepsilon_{1t} \sim$ Asy. Student $t_{12,1}$				$\varepsilon_{2t} \sim$ Asy. Student $t_{14,5}$				
0	86.7	86.3	86.6	85.6	86.5	88.5	88.4	86.5	86.9	87.2	87.1	88.5
1	87.5	87.6	87.9	87.9	88.5	89.0	89.3	89.0	88.3	88.5	88.2	89.4
2	87.4	87.8	88.3	87.7	88.2	88.8	88.8	88.5	88.0	88.6	88.2	88.5
3	87.0	87.3	87.9	87.3	87.8	88.4	88.3	88.5	87.5	88.3	88.0	88.3
4	86.6	87.1	87.6	87.1	87.3	87.8	87.9	88.0	87.1	87.6	87.6	87.8
5	86.2	86.6	87.2	86.8	86.7	87.3	87.3	87.5	86.4	87.2	87.0	87.4
6	85.8	86.3	86.6	86.4	86.5	86.8	86.9	87.1	86.3	86.9	86.7	87.0

Sample length=500, Replications=10,000. Bivariate VAR(1). Mixture: DLSMN MLE; Student-AFS: general FS correction applied to the Student t; Student: Student t MLE.

TABLE 5: Coverage of Impulse Response Functions. Nominal 90%. $T = 500$.

	Mixture									**Student-AFS**									**Student**								
Lag	(1,1)	(2,1)	(3,1)	(1,2)	(2,2)	(3,2)	(1,3)	(2,3)	(3,3)	(1,1)	(2,1)	(3,1)	(1,2)	(2,2)	(3,2)	(1,3)	(2,3)	(3,3)	(1,1)	(2,1)	(3,1)	(1,2)	(2,2)	(3,2)	(1,3)	(2,3)	(3,3)
DGP: $\varepsilon_{1t} \sim$ Student t_5 $\varepsilon_{2t} \sim$ Student t_5 $\varepsilon_{3t} \sim$ Student t_5																											
0	83.2	83.7	83.8	83.5	83.8	82.6	84.0	83.8	84.7	83.6	87.7	87.2	87.7	84.5	87.4	87.6	87.7	85.9	87.8	87.1	86.6	87.1	88.8	86.5	87.1	87.0	89.0
1	86.4	86.1	85.7	85.6	86.3	85.3	85.6	85.9	86.3	87.5	88.5	88.7	88.1	88.2	88.4	88.6	88.5	88.1	89.0	88.4	88.3	88.1	89.4	88.0	88.4	88.4	88.9
2	86.4	86.6	86.0	86.3	86.1	85.7	85.7	85.9	85.8	88.1	88.7	88.5	88.3	88.2	88.6	87.9	88.3	88.2	88.9	88.8	88.4	88.7	88.8	88.7	88.5	88.6	88.3
3	86.0	86.3	85.3	85.9	85.9	85.4	85.6	85.6	85.3	87.4	88.5	88.0	87.5	88.0	88.2	87.5	87.9	87.9	88.2	88.8	88.1	88.0	88.5	88.5	88.1	88.1	88.0
4	85.5	85.9	84.9	85.7	85.6	85.0	85.5	85.4	84.9	86.8	87.9	87.7	87.2	87.6	87.4	87.1	87.3	87.7	87.4	88.2	88.0	87.6	87.9	87.8	87.6	87.4	87.9
5	85.0	85.3	84.6	85.3	85.1	84.6	85.0	85.0	84.7	86.4	87.2	87.0	86.8	87.1	87.1	86.6	86.8	87.0	86.9	87.4	87.3	87.2	87.4	87.5	87.2	87.1	87.3
6	84.6	84.9	84.2	84.9	84.6	84.2	84.7	84.7	84.3	86.1	86.6	86.5	86.5	86.4	86.5	86.2	86.4	86.5	86.5	87.1	86.7	87.0	86.8	86.8	86.9	86.7	86.9
DGP: $\varepsilon_{1t} \sim$ Laplace $\varepsilon_{2t} \sim$ Laplace $\varepsilon_{3t} \sim$ Laplace																											
0	88.1	84.0	83.5	84.7	87.7	84.2	84.4	84.5	87.9	87.9	88.4	87.8	88.3	87.9	87.6	88.2	88.7	87.9	98.4	83.9	83.2	86.3	98.3	82.8	86.1	86.3	98.1
1	86.8	86.3	85.2	86.1	87.2	85.8	86.1	86.2	86.0	87.9	89.0	87.9	88.5	88.3	88.4	88.4	88.3	87.9	96.3	90.2	88.7	91.7	95.4	89.6	91.9	90.9	94.6
2	86.1	86.6	85.3	86.0	86.5	85.9	86.3	86.6	85.8	88.1	88.7	87.6	88.1	88.3	88.0	88.3	88.6	87.7	94.8	92.1	90.3	93.5	93.9	90.9	93.1	92.4	92.7
3	85.7	86.3	85.5	85.9	86.5	85.5	85.9	86.2	85.4	87.6	88.2	87.4	87.6	88.3	87.4	88.0	88.1	87.7	94.1	92.7	91.0	93.5	93.5	91.2	93.6	92.8	92.0
4	85.5	86.4	85.3	85.5	86.2	85.3	85.6	85.9	85.1	87.5	87.9	87.0	87.2	88.0	86.9	87.4	87.8	87.3	93.8	92.9	91.4	93.5	93.2	91.5	93.5	92.8	91.9
5	85.1	86.1	85.2	85.1	85.9	85.0	85.4	85.8	85.0	86.9	87.4	86.7	86.9	87.3	86.5	87.1	87.4	86.8	93.3	92.8	91.1	93.3	93.0	91.5	93.2	92.7	91.7
6	84.9	85.7	84.9	85.0	85.6	84.4	84.9	85.5	84.4	86.6	87.1	86.1	86.7	87.2	86.2	86.6	87.1	86.2	92.9	92.8	91.0	92.9	92.7	91.4	93.0	92.5	91.5
DGP: $\varepsilon_{1t} \sim \text{DLSMN}(0.8, 0.06, 0.52)$ $\varepsilon_{2t} \sim \text{DLSMN}(1.2, 0.08, 0.4)$ $\varepsilon_{3t} \sim \text{DLSMN}(-1, 0.2, 0.2)$																											
0	89.8	87.6	87.3	88.3	88.9	88.3	88.1	88.1	87.4	89.6	88.5	88.0	88.0	89.2	87.8	88.4	88.6	87.6	98.2	95.8	84.1	92.2	99.5	83.7	89.4	97.7	95.0
1	88.7	88.0	87.9	89.1	88.9	88.3	89.1	88.9	87.8	88.4	88.8	88.6	89.1	89.2	88.0	89.3	89.0	88.1	97.4	98.8	89.5	95.4	99.3	89.4	94.3	98.9	92.2
2	88.3	88.5	87.4	89.0	88.6	88.0	89.1	88.6	87.3	88.3	88.7	88.1	88.9	88.9	87.8	89.2	89.1	87.7	96.9	99.2	90.6	96.2	99.2	90.2	95.8	99.1	90.8
3	88.3	88.4	87.0	88.7	88.4	87.5	88.7	88.4	87.3	88.0	88.1	87.6	88.5	88.4	87.4	88.7	88.4	87.3	96.7	99.2	90.5	96.7	99.2	90.4	96.3	99.2	90.1
4	88.1	88.0	86.5	88.4	87.9	87.1	88.2	88.0	86.6	87.5	87.8	87.1	88.1	87.9	86.9	88.2	88.2	86.9	96.7	99.3	90.4	96.8	99.3	90.5	96.7	99.3	90.0
5	87.9	87.8	86.2	88.2	87.7	87.0	87.9	88.0	86.4	87.2	87.2	86.6	87.6	87.6	86.3	87.7	87.7	86.6	96.7	99.3	90.0	96.8	99.3	90.2	96.8	99.3	89.7
6	87.7	87.6	86.1	87.8	87.7	86.5	87.8	87.7	86.1	86.9	86.8	86.2	87.3	87.0	86.0	87.3	87.4	86.1	96.7	99.4	89.8	96.8	99.3	90.0	96.7	99.3	89.5
DGP: $\varepsilon_{1t} \sim$ Asy. Student $t_{12,1}$ $\varepsilon_{2t} \sim$ Asy. Student $t_{14,5}$ $\varepsilon_{3t} \sim$ Asy. Student $t_{16,100}$																											
0	85.5	84.7	86.4	85.2	84.8	87.7	87.1	87.8	84.1	84.6	87.3	88.1	87.3	85.1	88.0	87.9	88.2	84.3	85.1	85.9	84.8	85.9	87.8	84.9	85.1	86.1	90.6
1	86.7	86.9	87.0	87.1	87.5	88.1	87.1	87.8	87.1	87.7	88.2	88.1	88.1	88.3	88.4	88.5	88.3	87.9	87.2	87.5	87.4	86.9	88.9	88.0	87.2	87.4	90.6
2	86.0	87.1	87.5	86.7	87.4	88.1	86.4	87.1	87.2	87.3	88.4	88.4	87.7	88.3	88.4	87.8	88.2	88.3	86.4	88.0	88.7	86.9	88.1	88.3	86.8	87.9	89.5
3	85.5	86.6	87.3	86.1	86.8	87.5	85.9	86.9	87.2	86.7	87.8	88.0	87.1	87.9	88.0	86.8	87.9	88.1	86.0	87.7	88.7	86.4	87.7	88.5	85.9	87.7	88.9
4	85.3	86.1	86.8	85.5	86.3	86.9	85.2	86.3	87.0	86.2	87.3	87.4	86.4	87.2	87.5	86.3	87.3	87.7	85.6	87.2	88.2	85.7	87.1	88.2	85.6	86.9	88.4
5	85.1	85.6	86.3	84.9	85.8	86.5	85.0	85.8	86.7	85.6	86.7	86.8	85.8	86.7	86.9	85.9	86.6	87.1	85.1	86.6	87.8	85.3	86.3	87.6	85.2	86.4	88.0
6	84.8	85.2	85.8	84.6	85.3	86.1	84.7	85.2	86.2	85.2	86.0	86.3	85.4	86.0	86.4	85.3	86.1	86.6	84.6	85.9	87.1	84.8	85.8	87.3	84.6	85.9	87.6

Sample length=500, Replications=10,000. Trivariate VAR(1). Mixture: DLSMN MLE; Student-AFS: general FS correction applied to the Student t; Student: Student t MLE.

TABLE 6: Monte Carlo results. (RMSE) of pooled groups of estimators.

DGP:	$\varepsilon_{1t} \sim$ Asy. Student $t_{12,1}$		$\varepsilon_{2t} \sim$ Asy. Student $t_{14,5}$		$\varepsilon_{3t} \sim$ Asy. Student $t_{16,100}$	
	$K=2$	$K=3$	$K=2$	$K=3$	$K=2$	$K=3$
	$T=500$		$T=2000$		$T=20000$	
$\boldsymbol{\tau}$	0.0478	0.0482	0.0230	0.0231	0.0072	0.0072
A_{ii}	0.0347	0.0355	0.0168	0.0163	0.0053	0.0051
$A_{ij,i\neq j}$	0.0345	0.0350	0.0169	0.0164	0.0053	0.0051
C_{ii}	0.0623	0.0671	0.0318	0.0329	0.0099	0.0099
$C_{ij,i>j}$	0.0528	0.0516	0.0252	0.0227	0.0079	0.0071
$C_{ij,i<j}$	0.0598	0.0614	0.0288	0.0271	0.0089	0.0083
$J_{ij,i>j}$	0.0523	0.0507	0.0246	0.0219	0.0077	0.0068
$J_{ij,i<j}$	0.0645	0.0677	0.0303	0.0285	0.0093	0.0087

Replications=5,000. K=2: PMLE estimator based on finite mixture of two normals, K=3: PMLE estimator based on finite mixture of three normals.

FIGURE 1: Volatility index series (logs)

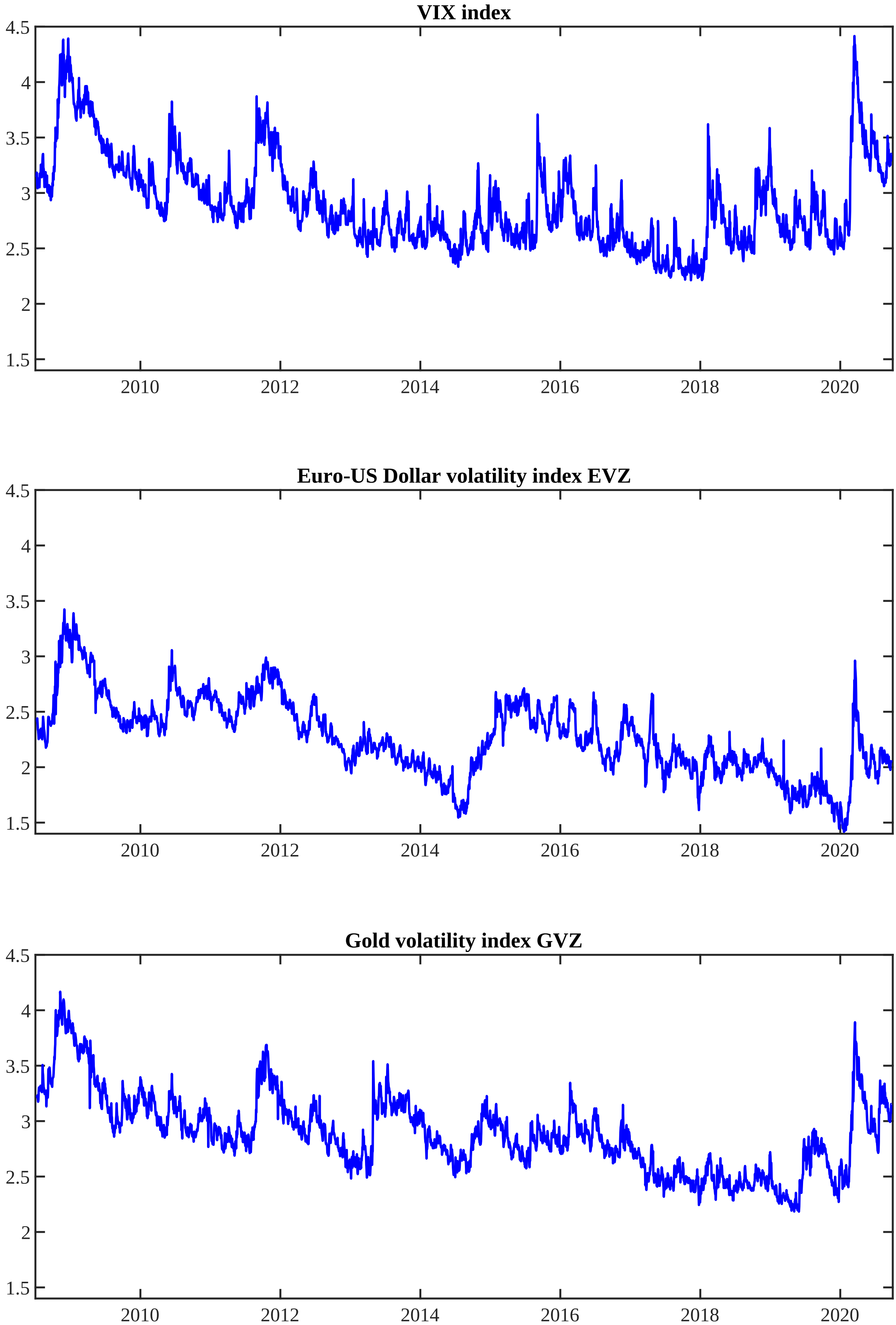

FIGURE 2: Structural Shocks

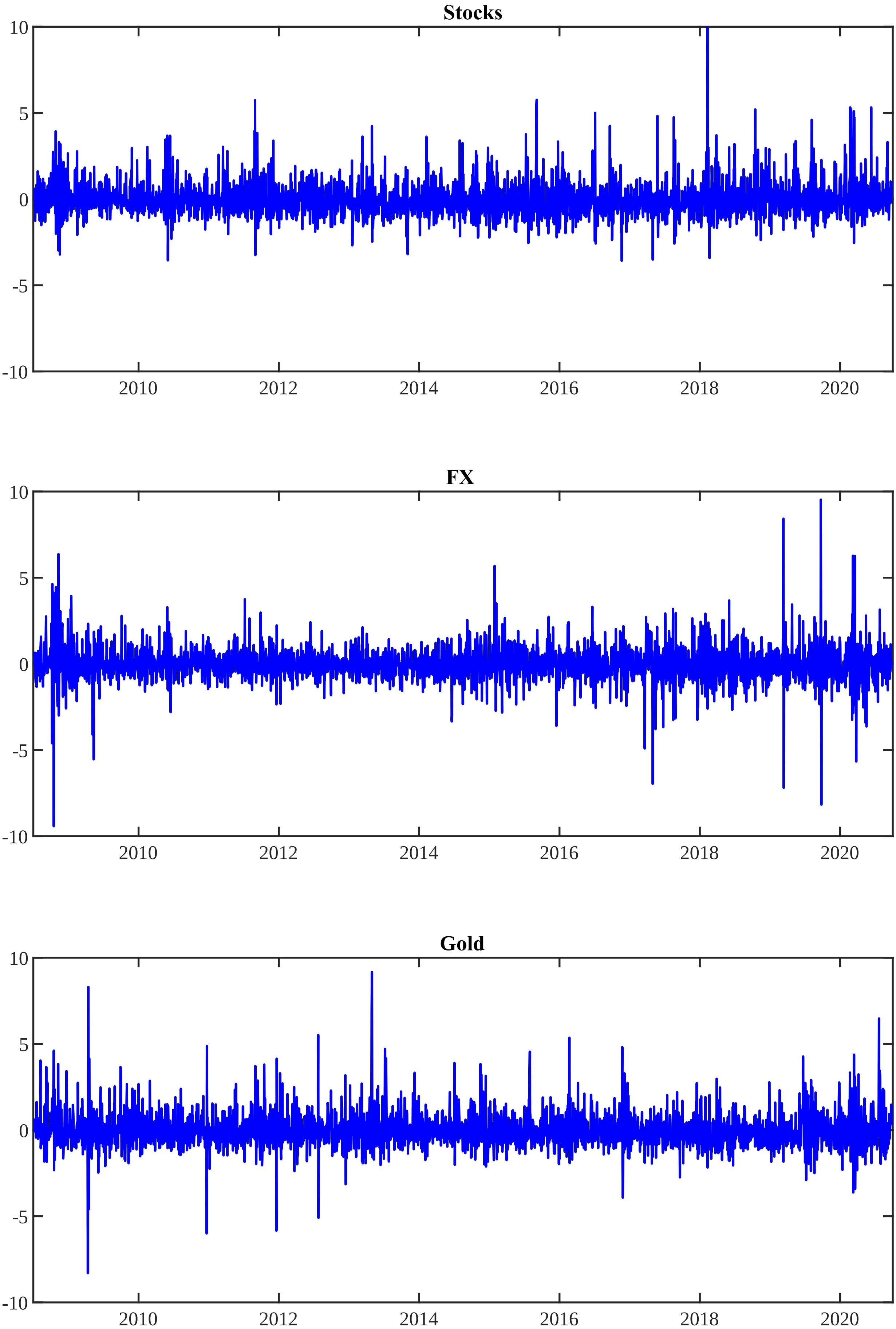

FIGURE 3 Impulse response functions and forecast error variance decompostions