

Bayesian Inference on Fully and Partially Identified Structural Vector Autoregressions

Jetto Anttonen*, Markku Lanne, and Jani Luoto

Faculty of Social Sciences, University of Helsinki

February 14, 2023

Abstract

We show that under standard assumptions the elements of the impact matrix of the structural vector autoregression (SVAR) are always at least set identified and bounded by the standard deviations of the corresponding reduced form errors. This result facilitates valid Bayesian inference without additional restrictions when only some (or none) of the columns of the impact matrix are point identified due to non-Gaussianity or heteroskedasticity. Moreover, the properties of the shocks can be assessed to find out which of them (if any) are point identified. We expand identification results put forth in the previous literature to models where all or part of the structural shocks are orthogonal but mutually dependent. To exploit deviations from Gaussianity, we propose using versatile error distributions and discuss their implementation in Bayesian analysis. Simulation results and an empirical application to U.S. fiscal policy lend support to the usefulness of efficiently accounting for non-Gaussianity.

Keywords: structural vector autoregression, Bayesian inference, statistical identification, partial identification, skewed generalized t -distribution, fiscal policy.

*Corresponding author. E-mail address: jetto.anttonen@helsinki.fi

1 Introduction

Utilisation of the statistical properties of data in identification has become increasingly popular in the recent literature on structural vector autoregressive (SVAR) models. In particular, sufficient non-Gaussianity has been shown to guarantee global and local identification of the parameters of the model and, hence, to facilitate valid statistical inference on the effects of the structural shocks. This is also the case with certain forms of heteroskedasticity. In addition to sharpening identification by restrictions based on economic theory or institutional knowledge, statistical properties of the data can be exploited to assess such restrictions and, hence, to preclude false or misleading conclusions.

Point identification of the parameters of the SVAR model (up to ordering and signs of the shocks) can be shown when at most one of the independent structural shocks is Gaussian (see, e.g., Lanne, Meitz, and Saikkonen, 2017). As we show in Section 2.1, when the structural shocks are mutually uncorrelated, irrespective of their distribution, the elements of the impact matrix of the SVAR model are always at least set identified and bounded by the standard deviations of the errors of the corresponding reduced form VAR model. This result is not surprising, but it is important in that it guarantees valid Bayesian inference even when only part or none of the structural shocks are point identified, and an improper constant (or proper but vague) prior is used on the elements of the impact matrix. This is because the marginal posterior distributions of its elements have a bounded support also when based on vaguely informative or non-informative priors. It is also worth noting that because the posterior distribution can be obtained by simulation even when none of the structural shocks are point identified, conventional sign restrictions can be imposed in a straightforward manner by restricting the domain of the posterior according to the given restrictions.

The significance of the set identification result is emphasised in the relatively common situation where only some of the structural shocks are statistically identified, or we are not even interested in the effects of all shocks. In contrast, in the frequentist setup, valid asymptotic inference calls for additional restrictions on the columns of the impact matrix corresponding to the non-identified shocks in that case (see, Maxand, 2020; Guay, 2021; and

Bertsche and Braun 2022), and if these extra restrictions are incorrect, they may adversely affect the properties of estimators. Moreover, pre-testing for the number and location of identified shocks is avoided in the Bayesian setup, as the properties of all shocks can be assessed after estimation. Frequentist estimators are also hampered by weak identification due to (some of) the shocks being nearly Gaussian or homoskedastic. Due to the fact that Bayesian analysis of a model is possible as long as the posterior distribution of its parameters is proper (see, e.g., Poirier, 1998), Bayesian methods are robust with respect to such weak identification, whereas standard asymptotic inference for frequentist estimators, in general, fails.

The assumption of mutually independent structural shocks made in a large part of the previous statistical identification literature may be restrictive and can be relaxed under certain additional assumptions (cf., Lewis, 2021; Guay, 2021; Lanne and Luoto, 2021; and Lanne, Liu, and Luoto, 2022). Lanne, Liu, and Luoto (2023) show that if the orthogonal but mutually dependent shocks have zero co-skewness, only the skewed shocks are identified. Moreover, if at most one of the shocks exhibits no skewness or persistent time-varying volatility, they are all identified. Building upon the former result, we show that, under the zero co-skewness assumption, the uncorrelated skewed shocks are point identified, while the shocks with zero skewness are at least set identified with bounds given by the standard deviations of the errors of the corresponding reduced form VAR model. This result is important because it facilitates valid Bayesian inference also in the case of dependent conditional variance processes. We also derive identification results for the case where only a subset of the structural shocks are mutually independent and independent of the rest of the shocks that are only mutually orthogonal. Hence, coupled with the set identification result mentioned above, our results cover a wide range of commonly encountered situations that can be tackled by Bayesian methods.

In order to efficiently exploit different kinds of deviations from Gaussianity as well as heteroskedasticity in identification, sufficiently versatile error distributions must be allowed for in estimation. Besides facilitating identification, accounting for these features is also important from the viewpoint of avoiding misspecification. However, complications may arise in practice due to complicated specifications and also because the ordering of the shocks is

not fixed by the statistical properties of the data. In Section 3, we discuss the implementation of the methods in detail. In particular, we recommend using the skewed t -distribution and estimating the SVAR model by means of Hamiltonian Monte Carlo methods.

To gauge the finite sample performance of the methods we report the results of two simulation experiments. First, we generate data from a trivariate SVAR model with mutually independent error terms, two of which are Gaussian. The third error term features either skewness, excess kurtosis, conditional heteroskedasticity or exhibits all of these properties, and it is hence point identified. In estimation, each of the errors is assumed to follow a skewed t -distribution that is capable of capturing the different deviations from Gaussianity. The general conclusion is that the parameters are estimated quite accurately even in relatively small samples by exploiting any of the non-Gaussian features, and accuracy is improved in the presence of all of them. Our second simulation experiment illustrates how simulation from the posterior and prior can help to assess the strength of identification. We recommend conducting similar simulation experiments following any empirical analysis, when computationally feasible.

We illustrate the methods in an empirical application to estimating U.S. fiscal multipliers. In particular, we consider a trivariate SVAR model for tax revenue, government spending and the GDP. At least two of the identified structural shocks turn out to be non-Gaussian, indicating point identification of the impact matrix. Two of the shocks can be labeled as the government spending and tax shocks by inspecting their impulse responses and forecast error variance decompositions as well as narrative records. The tax and spending multipliers are found similar to those of Lewis (2021), who also makes use of the statistical properties of the data, namely heteroskedasticity, in identification, but quite different from those in the earlier literature, mostly ignoring such information.

The rest of the paper is organized as follows. Section 2 contains the identification results. First, we show that set identification of the impact matrix is achieved even if the structural errors are Gaussian, whereas the independent shocks exhibiting non-Gaussianity are point identified, and the fact that the remaining shocks are set identified with appropriate bounds facilitates valid Bayesian inference. Next, making use of the set identification result, we expand the results for identification in the presence of orthogonal but not independent

shocks put forth in the previous literature. In Section 3, Bayesian inference is discussed. In particular, we consider estimation in the presence of complicated error distributions and heteroskedasticity that facilitate capturing versatile deviations from Gaussianity to strengthen identification. The finite sample performance of the methods is studied in Section 4, while Section 5 contains the empirical application to U.S. fiscal multipliers. Finally, Section 6 concludes.

2 Identification of SVAR Models

We build upon the literature on statistical identification of structural vector autoregressive (SVAR) models (for a survey, see Kilian and Lütkepohl, 2017, Chapter 14). Let us consider the n -variate SVAR model of order p ,

$$y_t = a + A_1 y_{t-1} + \dots + A_p y_{t-p} + B \varepsilon_t, \quad (1)$$

where y_t is the n -dimensional time series of interest, a is an $(n \times 1)$ intercept term, and A_1, \dots, A_p are $(n \times n)$ parameter matrices. The $(n \times n)$ nonsingular matrix B defines the $(n \times 1)$ vector of reduced-form errors u_t as a linear combination of the mutually uncorrelated structural errors ε_t , i.e., $u_t = B \varepsilon_t$. Thus, the (unconditional) covariance matrix of the reduced-form errors, assuming it exists, is given by $\Omega \equiv E(u_t u_t') = B \Sigma B'$, where $\Sigma \equiv E(\varepsilon_t \varepsilon_t')$. We denote the (i, j) elements of Ω and Σ by ω_{ij} and σ_{ij} , respectively.

Following the literature, we assume that the components of ε_t are not only mutually uncorrelated but independent. In particular, we make the following assumption regarding the structural errors, adapted from Assumption 1* in Lanne, Meitz, and Saikkonen (2017).

Assumption 1

- (i) *The error process $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ is a sequence of stationary random vectors with each component ε_{it} , $i = 1, \dots, n$, having zero mean and a finite positive variance σ_{ii} .*
- (ii) *The component processes ε_{it} , $i = 1, \dots, n$, are mutually independent.*

(iii) For all $i = 1, \dots, n$, the components ε_{it} are serially uncorrelated: $Cov[\varepsilon_{i,t}, \varepsilon_{i,t+k}] = 0$ for all $k \neq 0$.

Assumption 1 contains no strong distributional assumptions. In particular, in addition to Gaussian and homoskedastic structural errors, it covers a wide variety of non-Gaussian distributions and allows each component ε_{it} to follow a univariate conditionally heteroskedastic process. The independence assumption 1(ii) is standard in the related literature, and also more generally in the SVAR literature where Gaussianity and mutual orthogonality of the errors are typically assumed, resulting in independent shocks.

In the related statistical identification literature, it is assumed that either (i) at most one component of ε_t is marginally Gaussian or (ii) at least $n - 1$ components of ε_t are heteroskedastic with autocorrelated time-varying volatility. Under either of these assumptions, coupled with Assumption 1, global identification of the impact matrix B (up to sign reversals and permutation of its columns) can be established (see, Lanne, Meitz, and Saikkonen, 2017). Only the existence of the first two moments of ε_{it} , $i = 1, \dots, n$, is required, while empirical estimation of higher moments is not, in general, involved.

It is only if identification is based on conditional heteroskedasticity, or Assumption 1(ii) is relaxed to allow the structural shocks to be mutually dependent that the existence of higher moments is required to show the corresponding identification results. In the former case, finite second, third, and fourth moments of ε_{it} , $i = 1, \dots, n$ must be assumed (see, e.g., Lewis, 2021), while in the latter case the existence of their second and third moments is called for. In contrast, identification via a non-Gaussian parametric distribution need not depend on higher moments per se. This is important because some authors, including Montiel Olea, Plagborg-Møller, and Qian (2022), have been worried about potential weak identification of non-Gaussian SVAR models due to an insufficient amount of data often available to accurately estimate the third and higher moments that some approaches rely on.

Even if fewer than $n - 1$ of the shocks are non-Gaussian or heteroskedastic, partial identification is possible. In particular, as shown in Section 2.2, under the independence assumption, as stated in Assumption 1(ii), the i th column of the impact matrix B is globally

identified (up to its sign reversal), when ε_{it} is non-Gaussian and/or displays time-varying volatility, despite the remaining $n - 1$ shocks being Gaussian and homoskedastic. Partial identification can also be shown even when the independence assumption is appropriately relaxed. In particular, in Section 2.3, it is shown that the skewed structural errors are point identified (up to their sign reversals and ordering) in the presence of mutually dependent structural errors. Moreover, as shown in Section 2.1, B is always at least set identified, such that the SVAR model can be analysed by Bayesian methods.

2.1 Set Identification

In much of the SVAR literature, the components of the structural error vector ε_t are assumed to be Gaussian and homoskedastic. In light of the discussion above, it is clear that under this assumption, point identification of none of the shocks, or equivalently, columns of the impact matrix B is achieved without additional restrictions. It is, nevertheless, instructive to consider this worst-case scenario from which even the smallest deviation strengthens identification. While point identification is in this case not achieved, under Assumption 1, the SVAR model in (1) is set identified, meaning that multiple values of its parameters, collected into a $(d \times 1)$ vector $\theta = (a', \text{vec}(A_1, \dots, A_p)', \text{vec}(B)')'$,¹ are observationally equivalent, but the set of the observationally equivalent parameter values is not unconstrained.

To give a formal definition of set identification, let us first define spaces for the parameter vector θ and data sample Y , and be more specific about the conditional distribution of Y given θ , and its density (the likelihood function). Let Θ be a measurable space of $\theta \in \Theta \subseteq \mathbb{R}^d$, and \mathbf{Y} a measurable space of $Y \in \mathbf{Y} \subseteq \mathbb{R}^{T \times n}$ generated by the SVAR model (1). Let us also assume that Θ and ε_t are such that the conditional distribution of Y given θ exists and has a probability density function $f(y | \theta)$ at every $\theta \in \Theta$, where $y \in \mathbf{Y}$ is the observed data. Then set identification can be defined as follows:

Definition 1 *The model is called **set identified**, if for any $\theta_1, \theta_2 \in \Theta$ and all $y \in \mathbf{Y}$, there exists $\tilde{\Theta} \subset \Theta$, such that whenever $f(y|\theta_1) = f(y|\theta_2)$, then $\theta_1, \theta_2 \in \tilde{\Theta}$. That is, any set of*

¹We normalize the unconditional error variances σ_{ii} ($i = 1, \dots, n$) to unity to retain the elements of B unconstrained

*observationally equivalent points in Θ is a strict subset of Θ and the subset $\tilde{\Theta}$ is called the **identified set**.*

Our definition of set identification is consistent with the typical notion of set identification in the literature (see, e.g., Lewbel, 2019). Importantly, point identification is just a special case of set identification where the identified set $\tilde{\Theta}$ is a singleton and consequently $\theta_1 = \theta_2$. The latter takes place under the conditions of non-Gaussianity and heteroskedasticity, as discussed above, while if the structural errors are Gaussian and homoskedastic, point identification is not achieved without additional restrictions. However, it is possible to find lower and upper bounds for each (i, j) element B_{ij} of the impact matrix B in the identified set even in that case. These bounds, given in Proposition 1, are derived under the assumption that the components of ε_t are mutually uncorrelated, and they are obtained from the equation $\Omega = BB'$, where the unconditional variances of the elements of ε_t are normalized to unity. Together with the fact that a and A_i ($i = 1, \dots, p$) in (1) are unique (see, e.g., Lanne, Meitz, and Saikkonen, 2017), they obviously ensure boundedness of the identified set $\tilde{\Theta}$. The proof of the proposition can be found in Appendix A.

Proposition 1 *A SVAR model characterized by (1) is always at least set identified, and the bounds of the elements of B in the identified set $\tilde{\Theta}$ cannot exceed the bounds that are characterized by the following inequalities,*

$$-\omega_{ii}^{1/2} \leq B_{ij} \leq \omega_{ii}^{1/2}, \quad i, j = 1, \dots, n.$$

In particular, $B_{ij}^2 = \omega_{ii}$ ($i, j = 1, \dots, n$) if and only if $B_{ik} = 0$ for all $k = 1, \dots, n$, $k \neq j$ (i.e., the case where $\omega_{ii}^{-1/2} u_{it} = \varepsilon_{jt}$), and $B_{ij}^2 < \omega_{ii}$ ($j = 1, \dots, n$) otherwise.

It thus turns out that the identified set $\tilde{\Theta}$ is always bounded by the standard deviations of the reduced-form errors, which has interesting implications. In particular, propriety of the posterior distribution of θ can be established (i.e., it can be shown that it has a finite integral with respect to θ) even when point identification is not achieved, and an improper constant prior is used on the elements of B (however, proper priors may be needed for some

other elements of θ to ensure propriety of the posterior)². This result follows from a suitable fundamental theorem of vector calculus and the fact that within the set $\tilde{\Theta}$ the marginal posterior of B is equal to its (constant) prior. Hence, valid Bayesian inference is facilitated, as it only requires a proper posterior (see, e.g. Poirier, 1998). Moreover, the marginal posterior distributions of the set-identified parameters are expected to behave well enough for a convenient posterior analysis because they have a bounded support even if an improper constant (or proper but vague) prior is imposed on the elements of B . As a result, the posterior distribution of θ can be obtained by simulation methods even under the worst-case scenario of Gaussian Y and a constant (or proper but vague) prior.

While improper (or proper but vague) priors can be used, in most cases we recommend using proper (informative) priors that also alleviate computational problems due to the fact that statistical identification only up to the ordering of the shocks can be achieved (see Section 3.3). Nevertheless, the possibility to employ improper (or proper but vague) priors might be useful for assessing the effect of different priors on the posterior distribution of the parameters. Also, improper (or proper but vague) priors could, for instance, in some cases be used for maximum comparability to results obtained using frequentist approaches (see, e.g., Section 5).

In the Gaussian case, the strength of data-based identification corresponds to that of a traditional sign-restricted SVAR model, which is obtained by an appropriate choice of priors. Of course, identification can be strengthened (or achieved), for example, by external instruments, each of which is correlated with one of the shocks and uncorrelated with the rest. A potential problem with this approach due to Stock (2008) is that only instruments weakly correlated with the shock of interest may be found (see, e.g., Montiel Olea, Stock, and Watson, 2021), implying that only little can be learned from the data. In addition to external information, identification can be strengthened by different kinds of restrictions, including short-run restrictions on B , long-run response restrictions, or sign restrictions on impulse responses. However, in any case, we recommend using a highly flexible error distribution to capture most deviations from normality. This way, the researcher can efficiently learn

²For instance, as shown in Bauwens and Lubrano (1998), in the case of Student's t distributed errors, sufficient prior information on the degrees of freedom parameter is required to ensure the propriety of the associated posterior.

from the data Y to achieve strong identification when the structural errors exhibit non-Gaussianity. However, the restriction-based approaches should not be seen as substitutes for statistical identification but complements. Indeed, in Section 3, we provide a practical framework for the estimation of Bayesian SVAR models that allows for sign restrictions (cf., the discussion below) as well as almost any other type of a priori restrictions.

Because the posterior distribution can be obtained by simulation even in the case of Gaussian data and a constant prior, sign restrictions can be imposed in a straightforward manner by restricting the domain of the posterior distribution by any given restrictions (recall that impulse responses are functions of θ). The resulting truncated posterior distribution is easily obtained from the estimated (unrestricted) posterior distribution by discarding the values of θ that violate the sign restrictions. It is important to realize that when the data are Gaussian, the (unrestricted) posterior contains the identified set (at least asymptotically), and, hence, the latter procedure results in a posterior that is very close to that obtained by the traditional sign restriction methods.³

Robustness to weak identification in Bayesian estimation is another important implication of Proposition 1. If some (or all) of the structural errors are nearly Gaussian and homoskedastic, they may be only weakly identified without additional restrictions. Nevertheless, according to Proposition 1, even in this case, perfectly valid, albeit not very interesting (in the sense that the prior densities are not updated within the identified set), marginal posterior densities and credible intervals would be obtained. Hence, Bayesian estimation guarantees robustness to weak identification (see Drautzburg and Wright (2023) for a related frequentist approach). The significance of this feature is emphasized in the case of partial identification (see Section 2.2) because it facilitates valid inference on the effects of the point-identified shocks of interest despite the rest of the shocks being weakly (or set) identified.

In the following two subsections, we provide full and partial identification results under different sets of assumptions that cover practically all situations encountered in realistic

³As noted by Baumeister and Hamilton (2015), the traditional sign restriction methods impose implicit informational priors on the elements of B (and consequently on impulse responses), and under those specific priors, our truncated posterior coincides with the posterior obtained by traditional sign restriction methods. However, our Bayesian approach requires formulating a prior distribution of B explicitly, be it improper or not, making it impossible to impose implicit priors on B by accident.

empirical applications. They can be seen as generalizations of the identification result of Lanne, Meitz, and Saikkonen (2017) to cases where only some of the structural errors are non-Gaussian, or at least part of the components of ε_t are mutually dependent. Compared to the previous literature, the main novelty of these results is that all structural errors (and hence columns of the impact matrix B) are at least set identified with bounds given by the standard deviations of the errors of the corresponding reduced form VAR model in all conceivable scenarios, which facilitates valid Bayesian inference under partial (point) identification.

2.2 Partial Identification

Even if only one of the structural errors has a non-Gaussian distribution (and the other structural shocks are Gaussian and homoskedastic), the corresponding column of the B matrix is globally point identified (up to its sign), while the remaining $n - 1$ columns are set identified. This partial identification result in general form is stated formally as Proposition 2 below, and it provides a considerable practical improvement over Theorem 2 in Rubio-Ramírez, Waggoner, and Zha (2010), according to which partial identification of SVAR models can be achieved only with admissible restrictions on θ . Of course, the latter result is based on the assumption of Gaussian data, but in practice any conceivable empirical application is likely to involve sufficient non-Gaussianity to facilitate invoking our Proposition 2. This result is also more general than the partial identification results for non-Gaussian data in the recent literature (see, Maxand, 2020; Guay, 2021; Bertsche and Braun, 2022) in that it enables valid Bayesian statistical inference, while the asymptotic distributions of the estimators in the latter papers remain unknown unless sufficient admissible restrictions are imposed on θ .

Proposition 2 *Suppose $\varepsilon_t = B^{-1}u_t$ satisfies Assumption 1 such that r ($0 < r < n - 1$) components of ε_t are non-Gaussian, and each of the remaining $n - r$ components has a Gaussian marginal distribution. Assume that these $n - r$ components are ordered last (without loss of generality). Let $B = [B_1, B_2]$ with B_1 ($n \times r$) and B_2 ($n \times (n - r)$) matrices. Then, the ($n \times r$) matrix B_1 , corresponding to the r non-Gaussian components of ε_t , is globally point*

identified up to sign reversals and ordering of its columns, while the $(n \times (n - r))$ matrix B_2 is set identified with bounds given in Proposition 1.

The proof of Proposition 2 found in Appendix A is straightforward, and its essential parts are based on the ideas in Maxand (2020) and Proposition 1. It exploits the properties of independent Gaussian random variables (see Lemma 9 in Comon (1994)) together with the well-known Skitovich-Darmois theorem (see Lanne, Meitz, and Saikkonen (2017)). It turns out that for point identification of any exogenous shock that is mutually independent of the other shocks, only non-Gaussianity of that shock itself is required. An important implication of Proposition 2 is that because the identified set of B_2 is bounded, the posterior distribution of θ has a bounded support and, hence, it can in practice be obtained by simulation methods, as discussed in Section 2.1.

2.3 Mutual Dependence in Error Process

The global point identification results in Lanne, Meitz, and Saikkonen (2017) and in Proposition 2 are based on mutual independence (Assumption 1(ii)). However, full and partial identification can be shown even when this assumption is appropriately relaxed. In particular, we next replace Assumption 1 by the following assumption, which, among other things, allows for dependence in the volatility processes of the components of ε_t :

Assumption 2

- (i) *The error process $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ is a sequence of stationary random vectors with each component ε_{it} , $i = 1, \dots, n$, having zero mean, a finite positive variance, and a finite third moment.*
- (ii) *The component processes ε_{it} , $i = 1, \dots, n$, are orthogonal and have zero co-skewness.*
- (iii) *For all $i = 1, \dots, n$, the components ε_{it} are serially uncorrelated: $Cov[\varepsilon_{i,t}, \varepsilon_{i,t+k}] = 0$ for all $k \neq 0$.*

Parts (i) and (iii) of Assumption 2 are identical to those of Assumption 1 with the exception that Assumption 2(i) is augmented with the requirement that each component of ε_t

has a finite third moment. The latter allows for replacing Assumption 1(ii) (independence) with a clearly less restrictive assumption that the components of ε_t are only mutually uncorrelated and have zero co-skewness: $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}) = 0$ for all $i, j, k = 1, \dots, n$, excluding $i = j = k$. However, no restrictions are imposed on the fourth (co-)moments of ε_t , so the conditional variance processes of the components of ε_t can be dependent.

The partial and full identification results, based on Assumption 2, are formalized in Proposition 3 below. Part (i) complements Proposition 1 in Lanne, Liu and Luoto (2023) by the set identification result in Proposition 1, and its proof that closely follows the proof of Proposition 1 in Lanne, Liu and Luoto (2023), is found in Appendix A. Notice that Assumption 2 covers stochastic volatility (SV) and autoregressive conditional heteroskedasticity (ARCH) type models (cf. Assumption A of Lewis, 2021). Hence, any time-varying volatility captured by these models can be used to strengthen identification, provided the fourth co-moments of the components of ε_t exist (see Assumption C of Lewis, 2021).

Proposition 3 *Suppose $\varepsilon_t = B^{-1}u_t$ satisfies Assumption 2. Let h ($0 < h \leq n$) components of ε_t have non-zero skewness, and the remaining $n - h$ (when $h < n$) components of ε_t have zero skewness. The latter $n - h$ components are ordered last (without loss of generality), and $B = [B_1, B_2]$ with B_1 ($n \times h$) and B_2 ($n \times (n - h)$) matrices.*

- (i) *If $h < n - 1$, the $(n \times h)$ matrix B_1 , corresponding to the h components of ε_t with nonzero skewness, is globally point identified up to sign reversals and ordering of its columns, while the $(n \times (n - h))$ matrix B_2 is set identified and the bounds for this set cannot exceed the bounds given in Proposition 1.*
- (ii) *If at least $n - 1$ ($h \in \{n - 1, n\}$) elements of ε_t have non-zero skewness and/or display time-varying volatility with non-zero autocovariance, the full matrix B is globally point identified up to sign reversals and ordering of its columns.*

According to part (i) of Proposition 3, the skewed structural errors (and the corresponding columns of B) are globally point identified up to sign reversals and ordering even in the presence of mutually dependent structural errors, while the remaining errors are at least set identified. When at most one of the components of the structural error vector exhibits

zero skewness, the entire impact matrix B is identified, as stated in part (ii). This is also the case when at most one component of ε_t is homoskedastic. Then, a parametric volatility process may be entertained to enhance identification under the additional assumption that $E_t [\text{vec}(\varepsilon_t \varepsilon_t') \text{vec}(\varepsilon_t \varepsilon_t)'] < \infty$ for every t . Dependent volatility processes are the likeliest source of dependence among the components of the structural error vector, so this is indeed a viable option.⁴

It is, of course, possible that only some components of the structural error vector are mutually dependent. For instance, the volatility processes of closely related financial shocks may easily be dependent, but it is unlikely that the volatility processes of all shocks in a SVAR model are dependent. In these cases, the dependent and independent components can be treated separately by partitioning the B matrix to columns corresponding to each group. In particular, building upon the identification result in Lanne, Meitz, and Saikkonen (2017) and Propositions 1–3, it is stated in Proposition 4 below that if at most one of the independent errors is Gaussian, they are all identified (up to sign reversals and ordering); otherwise only the independent non-Gaussian errors are point identified, while the independent Gaussian errors are set identified. Likewise, analogously to Proposition 3, the skewed dependent shocks are identified, whereas the remaining shocks are set identified. In each case, the bounds of the identified set are those given in Proposition 1. The proof of Proposition 4 is provided in Appendix A, and it is based on the fact that independent random variables cannot generally be obtained as linear transformations of uncorrelated but dependent random variables.

Proposition 4 *Suppose $\varepsilon_t = B^{-1}u_t$ satisfies Assumption 2 such that the last $n - g$ ($0 < g < n$) components of ε_t are dependent, and the first g components of ε_t are independent of each other and of the last $n - g$ dependent components of ε_t (the ordering of the shocks is without loss of generality). Let $B = [B_1, B_2]$ with B_1 ($n \times g$) and B_2 ($n \times (n - g)$) matrices.*

(i) *If at most one of the independent components of ε_t has a Gaussian marginal distribution, the $(n \times g)$ matrix B_1 corresponding to the g independent components of ε_t , is*

⁴Skewness and time-varying volatility are not mutually exclusive. For instance, if stochastic volatility is present in the structural errors, their marginal distributions are non-Gaussian even when their conditional distributions are Gaussian. Specifically, consider the case where $\varepsilon_t = \Sigma_t^{1/2} \eta_t$, where $\Sigma_t = \text{diag}(\exp(\sigma_{1t}^2), \dots, \sigma_{nt}^2)$, $\sigma_{it}^2 = \phi_i \sigma_{i,t-1}^2 + \epsilon_{it}$, $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})' \sim N(0, \Sigma_\epsilon)$, and $\eta_t \sim N(0, I_n)$. If the innovations η_{it} and ϵ_{it} are correlated, the marginal distribution of ε_{it} is skewed and leptokurtic (see, e.g., Yang, 2008).

globally point identified up to sign reversals and ordering of its columns.

- (ii) If at most $g - r$ ($0 < r < g - 1; g > 3$) of the independent components of ε_t have Gaussian marginal distributions, the r columns of B_1 corresponding to the non-Gaussian independent components of ε_t , are globally point identified up to sign reversals and ordering, whereas the columns of B_1 , corresponding to the Gaussian components of ε_t , are set identified and the bounds of this set are given in Proposition 1.*
- (iii) The columns of the $(n \times (n - g))$ matrix B_2 , corresponding to the skewed dependent components of ε_t are globally point identified up to sign reversals and ordering, while the remaining columns of B_2 , corresponding to the symmetric dependent components of ε_t , are set identified with bounds given in Proposition 1.*

Finally, it is worth noting that the strength of identification can be assessed (ex post) by inspecting properties of the structural shocks using tests of normality and independence (see Maxand, 2020, and Montiel Olea, Plagborg-Møller, and Qian, 2022 for a discussion). For instance, if only part of the estimated structural shocks are deemed independent and only one of these independent shocks is non-Gaussian, based on Proposition 4, this non-Gaussian shock is point identified. Alternatively, identification can be assessed by inspecting the marginal posterior distributions of the parameters controlling their shape, and this is the approach used in our empirical application in Section 5.

3 Bayesian Inference

While much of the literature on non-Gaussian SVAR models employs frequentist estimators, we estimate θ comprising the parameters of (1) by Bayesian methods, not least because a Bayesian analysis of a model is possible even if the model is not identified (or, in other words, is set identified), as long as the posterior distribution of its parameters is proper. As discussed in Section 2.1, a proper posterior distribution of θ can be obtained also when an improper constant prior on the elements of B is used and only set identification is achieved, but informative priors for some other parameters may be required. Obviously, also, the

distribution of ε_t must be such that the conditional distribution of data Y given θ exists and has a probability density function $f(y | \theta)$ for every $\theta \in \Theta$. In Subsection 3.1, we discuss the distribution of ε_t , and in Subsection 3.2 we provide the likelihood function of (1).

Although a proper posterior distribution of θ can be obtained also when the SVAR model in (1) is non-identified (or set identified), identifying restrictions stemming from economic theory and institutional knowledge can be imposed on θ to strengthen the identification, as already discussed. However, such restrictions are rarely indisputable and the Bayesian approach allows for incorporating soft identifying restrictions in the form of a prior distribution $p(\theta)$ (see, e.g., Baumeister and Hamilton (2015) for a discussion). Such an informative prior distribution is recommendable also because it very likely enhances the performance of the estimation algorithm, especially when more than one of the components of ε_t are Gaussian and homoskedastic (i.e., when some columns of B are only set identified) and because it may help in providing the shocks with economic labels. The labelling problem is discussed in detail in Subsection 3.3, while the description of our prior distribution is deferred to Appendix B to save space. Finally, Subsection 3.4 contains a short discussion on the posterior sampler.

3.1 Distribution of structural errors

To be able to apply the identification result in Proposition 2, Assumption 1 must be satisfied, while for Propositions 3 and 4, Assumption 2 must hold. Hence, to facilitate strong identification, the error distribution should be sufficiently flexible to enable capturing the salient features of the data. To allow for mutually dependent errors with time-varying volatility (covered by Assumption 2), let us reparameterize ε_t as follows: $\varepsilon_t = \Sigma_t^{1/2} \eta_t$, where each element of $\eta_t = (\eta_{1t}, \dots, \eta_{nt})'$ has zero mean and unit variance, and $Var[\varepsilon_t | \sigma_t, F_{t-1}] = \Sigma_t = \text{diag}(\sigma_t^2)$, $\sigma_t^2 = \sigma_t \odot \sigma_t$, $\sigma_t = (\sigma_{1t}, \dots, \sigma_{nt})'$, $F_{t-1} = \{\varepsilon_1, \dots, \varepsilon_{t-1}, \sigma_1, \dots, \sigma_{t-1}\}$. While the conditional variances σ_{it}^2 of the components of ε_t may be mutually dependent, the elements of η_t are assumed mutually and temporally independent. Also, σ_t and η_t are assumed independent of each other. In case Σ_t is constant in time, the components of ε_t are then independent.

The error term ε_t can be heteroskedastic, and to avoid misspecification, its volatility should be appropriately accounted for. Besides following a wide variety of volatility processes (cf. Lewis, 2021), including stochastic volatility (SV) and autoregressive conditional

heteroskedasticity (ARCH) type processes, ε_t can be unconditionally heteroskedastic in different ways. For instance, there may be fixed volatility regimes in time, akin to Brunnermeier et al. (2021), each with their own value of σ_t estimated from the data. While ARCH and SV processes facilitate capturing mutual dependence in the components of ε_t , it may also be imposed in another way. For example, one particularly parsimonious alternative to that end would be to set the elements of σ_t equal and then let a single predefined equation control their law of motion, so that the volatility of all the shocks follows the same process.

The distribution of η_t should ideally be such that $\varepsilon_t = \Sigma_t^{1/2} \eta_t$ covers the shape of the 'true' error distribution of (1), but, in practice, we need to strike a fine balance between parsimony, flexibility and computational feasibility. For these reasons, the space of possible error distributions needs to be restricted to a set of reasonable alternatives with sufficient flexibility to facilitate capturing most of the potential deviations from Gaussianity. Otherwise, identifying information would be unnecessarily left unexploited and the model would be misspecified, unless the data were actually Gaussian. One viable alternative is the extended t -distribution family, or more specifically, the skewed generalized t -distribution (sgt) of Theodossiou (1998). The sgt-distribution offers a great balance between flexibility, parsimony, and interpretability, and it has a wide variety of well known univariate probability distributions from the normal distribution to the Laplace-distribution and the t -distribution (including their skewed and/or generalized versions) as its special cases.⁵

In the empirical application of Section 5, we assume that each component of η_t individually follows a skewed t -distribution, which is a special case of the sgt-distribution with comparable flexibility (and it also nests the Gaussian distribution as a limiting case), but compared to the sgt-distribution has one parameter fixed. We choose the skewed t -distribution because it is computationally more convenient than the sgt-distribution, yet allows for both skewness and fat tails, which both are characteristics commonly observed in macroeconomic and financial time series.

Finally, while we find the skewed t -distribution quite capable of capturing the salient features of economic data, other distributions could be entertained for η_t . In particular, even

⁵We refer the reader to Appendix B for further discussion on the details of the sgt-distribution, such as its point density function.

more flexibility could be afforded by non- or semi-parametric approaches (see, e.g., Braun, 2021, and Hoesch, Lee, and Mesters, 2022), but our approach is obviously more efficient than these alternatives, if the skewed t -distribution encompasses the true error distribution of a SVAR model, which may easily be the case. Furthermore, based on the arguments in Gouriéroux, Monfort, and Renne (2017) concerning pseudo maximum likelihood (PML) estimator, we expect our method to be more efficient than non- or semi-parametric alternatives also when the skewed t -distribution is sufficiently close to the true error distribution.

3.2 Likelihood function

We derive the likelihood function in the case of the sgt-distribution, which has the skewed t distribution as a special case. Let us collect the parameters controlling the law of motion of σ_t and the shape of η_{it} ($i = 1, \dots, n$) into the vectors δ and γ_i ($i = 1, \dots, n$), respectively. Then, substituting ε_t for $\Sigma_t^{1/2}\eta_t$ in (1) and recalling the mutual (and temporal) independence of the elements of η_t , we can write the density function of the distribution of the data $y = (y'_1, \dots, y'_T)'$, i.e., the likelihood function, as

$$p(y|\theta) = |\det(B)|^{-T} \prod_{i=1}^n \prod_{t=1}^T \sigma_{it}^{-1} f_i(\sigma_{it}^{-1} \iota_i' B u_t(\pi); \gamma_i), \quad (2)$$

where $\theta = (\pi', \text{vec}(B)', \delta', \gamma')'$, $\pi = \text{vec}(a, A_1, \dots, A_p)'$, $\gamma = (\gamma'_1, \dots, \gamma'_n)'$, ι_i is the i th unit vector, $u_t(\pi) = y_t - a - A_1 y_{t-1}, \dots, A_p y_{t-p}$, and $f_i(\cdot)$ ($i = 1, \dots, n$) is the density function of η_t . Notice that to retain the elements of B unconstrained, the unconditional variances of ε_{it} are normalized to unity: $E(\varepsilon_{it}^2) = E(\sigma_{it}^2) = \sigma_{ii} = 1$ ($i = 1, \dots, n$). Alternatively, the diagonal elements of B can be normalized to unity, in which case we assume that $\sigma_{ii} > 0$ ($i = 1, \dots, n$).

3.3 Shock permutations

As discussed in Section 2, the SVAR model in (1) can, in general, only be identified up to sign reversals and ordering of the elements of ε_t (or equivalently up to signed permutations of the columns of B). In other words, there are $2^n n!$ different models corresponding to

the same likelihood function. While this is not a problem (in principle at least) from the point of view of Bayesian inference (as long as the posterior distribution of θ is proper), a meaningful analysis of the SVAR model (e.g., analysis of impulse response functions) may require pinning down a specific model (i.e., a specific permutation of structural errors with fixed signs). This issue is similar to the well known label switching problem with mixture models (see, e.g., Stephens, 2000), yet there are important subtleties where the problems differ.

In point estimation (e.g. by the method of maximum likelihood or the generalized method of moments), the issue is easily dealt with. In particular, as any ordering of the structural errors is valid, the choice between the different optima is irrelevant, and the parameter space can be restricted a priori without loss of generality such that necessarily only one optimum remains (see, e.g., Lanne, Meitz, and Saikkonen, 2017). However, in Bayesian analysis, the object of interest is the whole distribution of parameters, not a single point in the parameter space, which complicates matters. If a prior distribution is exchangeable (that is, ‘permutation agnostic’), there are trivially symmetric disjoint posterior volumes of high probability mass, corresponding to different permutations. Because the shape of the posterior distribution is unknown a priori, there is no way to restrict the parameter space a priori such that one specific disjoint volume of high posterior probability mass would be necessarily preserved, while the rest of the parameter space would be discarded. Indeed, in practice, any a priori restriction would almost certainly leave out some parts of the parameter space corresponding to the permutation of interest, and, hence, contaminate the posterior sample.⁶

Nevertheless, there are various ways to alleviate the issue in practice. First, near multicollinearity in the time series should be avoided, as it might imply nearly indistinguishable or even degenerate structural errors. Second, the more flexible the chosen error distribution is, the more ways there are in which the structural errors can deviate from each other,

⁶Note that, in principle, the symmetry of the posterior volumes of high probability mass could be used to map posterior draws from the unrestricted parameter space to the a posteriori restricted space corresponding to a specific permutation of the shocks, even if the posterior sample consisted of draws from multiple volumes of high probability mass. However, such an approach has not been studied before, and jumps between different volumes would in any case most probably pose a significant computational challenge regardless of the estimation algorithm used.

and consequently the more disjoint are the posterior volumes of high probability mass (provided the ‘true’ error distributions lie sufficiently far from each other). From this point of view, a skewed t -distribution is preferable to a mere t -distribution, and a skewed generalized t -distribution may be even better.

Finally, non-exchangeable prior information related to each shock breaks down the symmetry of the different posterior volumes and may be used to the extent of suppressing the permutation issue altogether, and this is the approach used in our empirical application in Section 5. In particular, following Brunnermeier et al. (2021), we shrink B towards a suitably scaled diagonal matrix to reflect the notion that each structural shock is the main driver of unexpected changes in one specific variable (for instance, the monetary policy shock could be thought of as the main driver of a policy interest rate). This approach seems appropriate in most cases (see Appendix B for details), and we recommend the practice as a general solution to the permutation issue. However, if the structural shocks are sufficiently distinguishable for the posterior volumes to be disjoint enough (as is actually the case in the empirical application of Section 5), then non-exchangeable prior information is not necessarily needed in practice.⁷

3.4 Estimation algorithm

The model can be straightforwardly estimated by generating draws from the posterior distribution of the parameters θ : $p(\theta|y) \propto p(y|\theta)p(\theta)$, where $p(y|\theta)$ is given in (2), and $p(\theta)$ is a prior density (see Appendix B for a discussion). To this end, any typical Markov chain Monte Carlo (MCMC) algorithm could be used. However, SVAR models are not parsimonious, and if more than a few variables and lags are included, the number of free parameters is likely to be beyond the practical capabilities of most estimation algorithms available.⁸ In Anttonen,

⁷Notice also that if the volumes of high probability mass are not sufficiently disjoint, then even a non-exchangeable prior may not be able to prevent the estimation algorithm from jumping between the different volumes. Fortunately, any consequential failure of the algorithm to stay in one volume, would necessarily manifest itself as either (i) inability of separate posterior chains to converge to the same stationary distribution or (ii) mutually indistinguishable impulse responses, and, hence, such behaviour would be easily spotted by the researcher. Anttonen, Lanne, and Luoto (2022) offer a thorough discussion on diagnosing such permutation jumps.

⁸The number of free parameters in π and B alone is $n^2(p+1) + n$, where n and p are the number of variables and lags in the model, respectively.

Lanne, and Luoto (2022), a model very similar to that in this paper was estimated using the differential evolution Markov chain (DE-MC) algorithm of Ter Braak and Vrugt (2008). The DE-MC algorithm is, however, essentially an adaptive Metropolis-Hastings algorithm with random walk proposals, which are known to be inefficient in high dimensional cases, although the adaptive capabilities of the algorithm, coupled with the computational power of modern computers, make it sufficiently efficient for estimation of at least medium scale SVAR models.⁹

Nevertheless, we propose to estimate the parameters θ of the model using the No-U-Turn Sampler (NUTS) of Gelman and Hoffman (2014)¹⁰, or more specifically, the version of the algorithm as implemented in Carpenter et al. (2017). The NUTS is a state-of-the-art Hamiltonian Monte Carlo (HMC; see, e.g., Neal, 2011) algorithm for estimation of Bayesian open-ended problems, and it has proven itself capable of tackling a wide variety of models of all shapes and sizes with little to no user intervention (see, e.g., Gelman et al., 2020).

The NUTS, as any other HMC algorithm, uses information in the gradient of the log-posterior to efficiently move around the continuous probability space of the parameters. The analytic computation of the gradient would be tedious at best, but fortunately it can be avoided altogether by the use of automatic differentiation (see, e.g., Griewank and Walther, 2008). The gradient is used to simulate Hamiltonian dynamics in the parameter space, as opposed to a random walk, resulting in proposals with close to 100 percent acceptance rate and low autocorrelation. For a detailed exposition of the estimation algorithm, see Gelman and Hoffman (2014), Betancourt (2017) and Carpenter et al. (2017). In Appendix B, we also discuss some practicalities related to the estimation algorithm in the context of our paper.

4 Simulation Experiments

In this section, we conduct two simulation studies to better understand the finite sample performance and properties of the methods discussed. First, we consider a simple trivariate

⁹The model estimated in Anttonen, Lanne, and Luoto (2022), for instance, features six variables and twelve lags, the total number of free parameters in π and B alone being 468.

¹⁰For a detailed discussion on the NUTS algorithm and associated diagnostics, we refer the interested reader to Gelman and Hoffman (2014) and Gelman et al. (2020).

case with two Gaussian shocks and one non-Gaussian shock. By Proposition 2, the non-Gaussian shock is point identified, whereas the two Gaussian shocks remain set identified. We illustrate how different non-Gaussian features, together or alone, lead to fairly strong identification and precise estimates even in relatively small samples.

In the second simulation study, we take the empirical application of Section 5 as our starting point. We simulate data using draws from both the prior and posterior distributions of the parameters. Essentially, simulation studies based on draws from the prior distribution are useful in assessing whether the parameters of the model can be accurately estimated from data under the assumptions of the model. In particular, they should be useful in detecting issues related to identification and the validity of inference. Simulations based on draws from the posterior distribution of the parameters, in turn, enhance our understanding of where the data (of interest) might be especially informative and, on the other hand, which parameter estimates remain largely affected by the prior. Hence, posterior based simulations can help to assess the strength of identification in the data, provided there are no issues in the model, such as misspecification and lack of identification, or problems in the estimation procedure. For instance, if the data turn out to be informative regarding the elements of a particular column of B , we expect identification of the related shock to be strong. However, this assessment is feasible only if there are no issues in the model or inference, and this can be checked using the prior based simulation studies.

4.1 Partially Identified Trivariate System

For simplicity, we consider a trivariate system without lags or a constant. Specifically, we generate the data from the following model:

$$y_t = B\varepsilon_t, \tag{3}$$

$$B = \begin{bmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix}, \tag{4}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ is the vector of mutually independent structural shocks of which the first two are Gaussian, and the third shock is non-Gaussian. Then, by Proposition 2, the third column of B is point identified, i.e. the effect of the only non-Gaussian shock on the variables of the system is identified.

Specification 1; Only Skewness			
	$T = 200$	$T = 500$	$T = 2000$
B_{13}	-0.02 (0.27)	-0.01 (0.17)	0 (0.07)
B_{23}	-0.02 (0.27)	0 (0.17)	0.01 (0.07)
B_{33}	-0.07 (0.16)	-0.02 (0.09)	0 (0.03)
λ_3	0.01 (0.11)	0 (0.07)	0 (0.03)
$\log q_3$	$-\infty$ (0.31)	$-\infty$ (0.31)	$-\infty$ (0.29)
Specification 2; Only Excess Kurtosis			
	$T = 200$	$T = 500$	$T = 2000$
B_{13}	-0.02 (0.28)	0 (0.17)	-0.01 (0.06)
B_{23}	-0.02 (0.27)	0 (0.16)	0 (0.07)
B_{33}	-0.09 (0.16)	-0.03 (0.09)	-0.01 (0.04)
λ_3	0 (0.12)	0 (0.06)	0 (0.03)
$\log q_3$	0.37 (0.47)	0.16 (0.30)	0.04 (0.14)
Specification 3; Only Heteroskedasticity			
	$T = 200$	$T = 500$	$T = 2000$
B_{13}	-0.02 (0.25)	0 (0.13)	0 (0.05)
B_{23}	-0.01 (0.25)	0.01 (0.13)	0 (0.06)
B_{33}	-0.06 (0.16)	-0.01 (0.08)	0 (0.03)
λ_3	0 (0.13)	0 (0.07)	0 (0.03)
$\log q_3$	$-\infty$ (0.29)	$-\infty$ (0.29)	$-\infty$ (0.29)
α_3	0.11 (0.06)	0.07 (0.05)	0.02 (0.04)
β_3	-0.05 (0.06)	-0.03 (0.04)	-0.01 (0.02)
Specification 4; All of the above			
	$T = 200$	$T = 500$	$T = 2000$
B_{13}	-0.02 (0.15)	-0.01 (0.08)	0 (0.04)
B_{23}	-0.03 (0.16)	-0.01 (0.08)	0 (0.04)
B_{33}	-0.08 (0.13)	-0.03 (0.07)	-0.01 (0.04)
λ_3	0.01 (0.10)	0 (0.06)	0 (0.03)
$\log q_3$	0.47 (0.47)	0.21 (0.29)	0.06 (0.13)
α_3	0.1 (0.06)	0.06 (0.06)	0.02 (0.04)
β_3	-0.05 (0.06)	-0.03 (0.04)	-0.01 (0.02)

Table 1: Estimated bias and standard deviation (in parentheses) for the point estimates under all the specifications considered (see the text for details) over 1000 replications. For the specifications where the third shock only exhibits skewness or conditional heteroskedasticity, $\log q_3 = \infty$ and therefore the bias is obviously $-\infty$.

We consider four different forms of non-Gaussianity of the third shock, and label the resulting data generating processes (DGP) Specifications 1 to 4. In Specification 1, the third shock is assumed to be only skewed (i.e. to follow a skewed Gaussian distribution), whereas in Specification 2 it is assumed to feature only excess kurtosis (i.e. to follow a t -distribution). In Specification 3, the third shock is assumed to be conditionally Gaussian, but heteroskedastic. Finally, in Specification 4, we let the shock have all these features, which, to no surprise, results in much stronger identification than any of the features alone. This finding underlines the importance of allowing for multiple forms of non-Gaussianity.

All the specifications can be cast using the skewed t -distribution discussed in Section 3 and also employed in the empirical application in Section 5. The extent of non-Gaussianity of any specification is set to reflect plausible values in the context of macroeconomic time series (see, e.g., Figure 5 related to the empirical application of Section 5). For the skewed shock we set λ_3 , the parameter controlling the skewness of ε_{3t} , equal to 0.3 (moderate positive skewness), for the t -distributed shock we set the degree-of-freedom parameter equal to 6 (i.e., $\log q_3 = \log 3 \approx 1.1$; see Appendix B.2 for details). Finally, we impose heteroskedasticity by letting the volatility of the third shock, $\sigma_{3,t}$, follow a first order GARCH-process

$$\sigma_{3,t}^2 = \sigma_3^2 + \alpha_3 \sigma_{3,t-1}^2 + \beta_3 \varepsilon_{3,t-1}^2,$$

where $\alpha_3 = 0.7$, $\beta_3 = 0.2$ and $\varepsilon_{3,t}$ is either normally distributed (Specification 3) or skewed t -distributed (Specification 4) with mean zero and variance unity. Hence, the shock volatility process is moderately persistent.

We employ an improper constant prior on the elements of B . Our priors for λ_i , q_i , α_i and β_i are described in Figure B.1 in Appendix B.2 (see Appendix B.2 for a detailed discussion). In addition, we only concentrate on the point estimates of these parameters, which can be interpreted as the values at the posterior mode, or equivalently, as penalized maximum likelihood estimates. We thus report the bias as the average difference between the real value and the point estimate over the simulations, and the standard deviation of these point estimates over the replications.

Table 1 summarizes the results based on 1000 simulated datasets for three sample sizes

($T = 200, 500, 2000$). In general, the parameters are quite accurately estimated already in samples of 200 observations, and the biases and standard deviations get smaller as the sample size increases. As expected, the fact that q_3 (the degree-of-freedom parameter) in the DGP differs from its prior mean induces a slight finite sample bias in the scale of the estimates of B_{13} , B_{23} and B_{33} .¹¹ The bias is however not huge and seems almost non-existent already for $T = 500$.

Evidently, precise estimation of the tail behaviour of the shock is challenging with the length of time series typically available, and hence the parameter uncertainty related to $\log q_3$ remains large. This highlights the importance of integrating over all the parameter uncertainty (Bayesian approach) as opposed to point estimation only. On the other hand, the estimates of the skewness parameter λ_3 seem precise and practically unbiased regardless of the number of observations. This is quite remarkable especially in the first and fourth specifications (the ones with skewed ε_{3t}), as in those cases there is a clear disparity between the prior and the real parameter value (the prior shrinks estimates of λ_3 towards zero whereas the real value of λ_3 is 0.3). The precision of skewness estimation well matches our overall experience of significantly stronger identification of impulse responses in the presence of even slight skewness in the shocks, highlighting the huge potential benefits of allowing for (and exploiting) the skewness in the structural shocks.

Finally, as expected, the addition of more than one non-Gaussian feature in the shock (the fourth specification) results in even more precise estimates of B as judged by the lower standard deviation of the point estimates. Thus, although point identification may be obtained by exploiting only, say, heteroskedasticity, stronger identification is obtained by exploiting, not some, but *all* the non-Gaussian features in the data.

4.2 Posterior and Prior Simulations

Next we consider the baseline specification of the empirical application of Section 5. The data (with 228 observations) are thus generated from a trivariate SVAR model with each of

¹¹The variance of shock i is a function of q_i (the degree-of-freedom parameter) and the shocks are normalized to have unit variance. Hence, if the estimated q_i differs from the value of q_i in the DGP, the scale of the elements of B will be different from those of the DGP due to this normalization.

the structural errors following a skewed t -distribution and displaying time-varying volatility (a univariate GARCH(1,1)-process is considered for each individual error, see (5)). The parameter values of the data-generating process are based on either the posterior or prior distribution of the parameters.

Let us begin with simulations based on the posterior distribution of parameters. Over 1000 replications, we first draw a set of parameters from their posterior distribution and then, conditional on those parameter values, we simulate a new data set and re-estimate the model from the simulated data. That is, we apply exactly the same specification (priors included) as in the original empirical application to the new data set simulated using the posterior distribution of parameters.

Table 2 reports the estimated bias and standard deviation over the replications. The estimates are biased by design, since we have not generated the data from the prior, but from the posterior. For instance, consider a parameter for which the posterior values hover mostly around, say, 0.4, but the prior shrinks the estimates towards zero. The estimates will obviously be biased, the amount of bias depending on the strength of the prior relative to the data.

Thus, simulation of data from the posterior allows us to assess the relative importance of the data and the prior for our posterior estimates. Biases close to zero (for parameters which the prior and posterior estimates do not coincide) imply that the prior information is inconsequential for the posterior estimate, whereas larger biases indicate greater sensitivity to the prior (as well as disparity of the prior and the posterior). The posterior based simulation study thus provides an excellent tool for the assessment of prior sensitivity and strength of identification.

For instance, in our empirical application, the data seem very informative regarding the elements of the second and third columns of B as well as the off-diagonal elements of the first column of B , and they do not seem sensitive to the prior at all: there are no significant biases and the coverage of the posterior intervals matches the nominal coverage. This is especially the case for B_{13} for which the posterior simulation study suggests practically no bias at all, but for which the posterior mean, 0.42, lies far from the prior mean of zero. The first diagonal element of B , on the other hand, is clearly affected by the prior of $\log q_1$ (see the

	Simulation bias			Coverage (90%)			Posterior mean		
B									
	-0.38 (0.43)	0 (0.06)	-0.01 (0.06)	0.56	0.93	0.87	1.28 (0.49)	-0.03 (0.06)	0.42 (0.07)
	-0.02 (0.08)	0 (0.11)	0.01 (0.08)	0.85	0.86	0.94	0.07 (0.10)	0.86 (0.12)	0.05 (0.07)
	0.01 (0.08)	-0.01 (0.08)	0 (0.10)	0.86	0.94	0.85	-0.01 (0.09)	0.1 (0.08)	0.85 (0.10)
λ	0 (0.09)	-0.01 (0.10)	0 (0.10)	0.89	0.9	0.91	0.05 (0.09)	0.06 (0.10)	-0.03 (0.10)
$\log q$	0.33 (0.26)	0.52 (0.47)	0.41 (0.51)	0.44	0.8	0.86	0.26 (0.12)	1.16 (0.35)	1.58 (0.57)
α	0.08 (0.10)	-0.02 (0.07)	-0.03 (0.05)	0.78	0.92	0.94	0.65 (0.11)	0.85 (0.08)	0.85 (0.06)
β	-0.08 (0.08)	0 (0.04)	-0.01 (0.05)	0.68	0.92	0.91	0.27 (0.08)	0.09 (0.05)	0.13 (0.05)

Table 2: Results of the posterior simulation study based on 1000 replications. The leftmost panel contains the average bias and standard deviation (in parentheses) for each parameter, using the posterior mean as the point estimate. The middle panel contains the fraction of simulations with the real value of each parameter lying within the (marginal) central 90% posterior interval (i.e., the ideal value is 0.90). The rightmost panel contains the means and standard deviations (in parentheses) of the posterior used to generate the data (i.e., the posterior distribution of the *Baseline* specification from Section 5).

discussion in the previous subsection and Footnote 11). The latter point is well illustrated in Figure 5 in Section 5 depicting a clear disparity between the prior and posterior of the degree-of-freedom parameter (recall that for a skewed t -distribution the degree of freedom parameter equals $2q_i$). Specifically, the data suggest that the tails of the shock distributions are much further from those of a Gaussian distribution than our relatively conservative prior.

In sum, according to the posterior based simulations, we expect identification of at least two of the shocks, corresponding to the second and third columns of B , to be particularly strong. However, as already discussed, this assessment is sensible only if there are no issues in the model related to, for instance, identification. Also, posterior based simulations tell us relatively little about the validity of inference, but potential issues related to the model and/or estimation procedure can be detected by a prior based simulation study. Therefore, we repeat the exercise with draws from the prior instead of the posterior. Given that there are no issues in the estimation procedure or in the model regarding, say, identification of the parameters, this should result in zero bias and correct nominal coverage of all the posterior intervals.

There are some practical considerations to take into account when carrying out prior based simulation studies. The priors used in practice (e.g. in Section 5) may be impractically weak for the purposes of generating data to be used for estimation. This would, of course,

also suggest the priors to be unnecessarily weak in the first place.¹² In practice, it is, however, often best to mitigate the risk of model misspecification by erring on the side of caution with prior choices, i.e. by preferring too weak priors over too strong ones.

For these reasons, in our prior simulation study we use slightly stronger priors than in the empirical application of Section 5. Specifically, we fix the hyperparameter κ_1 (the amount of shrinkage on autoregressive parameters) to 0.2, set the prior means of the elements of a and A_i ($i = 1, \dots, p$) in (1) equal to zero to avoid explosive dynamics and fix the prior mean and standard deviation of B as reported in the rightmost panel of Table 3 (log-normal prior on diagonal elements, normal prior on non-diagonal elements). All the other priors are as in the empirical application.

	Simulation bias			Coverage (90%)			Prior mean		
B	0.01 (0.09)	0 (0.10)	-0.01 (0.10)	0.91	0.91	0.89	1.10 (0.28)	0 (0.20)	0 (0.20)
	0 (0.10)	0.01 (0.10)	0 (0.10)	0.90	0.89	0.90	0 (0.20)	1.10 (0.28)	0 (0.20)
	0.01 (0.11)	0 (0.11)	0.01 (0.10)	0.89	0.89	0.87	0 (0.20)	0 (0.20)	1.10 (0.28)
λ	0.01 (0.10)	0.01 (0.10)	0 (0.10)	0.90	0.89	0.89	0 (0.33)	0 (0.33)	0 (0.33)
$\log q$	-0.04 (0.69)	-0.02 (0.68)	-0.03 (0.69)	0.90	0.91	0.89	2.18 (0.85)	2.18 (0.85)	2.18 (0.85)
α	0 (0.09)	0 (0.09)	0 (0.09)	0.89	0.89	0.89	0.83 (0.10)	0.83 (0.10)	0.83 (0.10)
β	0 (0.05)	0 (0.05)	0 (0.04)	0.90	0.90	0.90	0.08 (0.08)	0.08 (0.08)	0.08 (0.08)

Table 3: Results of the prior simulation study based on 1000 replications. The leftmost panel contains the average bias and standard deviation (in parentheses) for each parameter, using posterior mean as the point estimate. The middle panel contains the fraction of simulations with the real value of each parameter lying within the (marginal) central 90% posterior intervals (i.e. the ideal value is 0.90). The rightmost panel contains the means and standard deviations (in parentheses) of the prior distributions used to generate the data (elaborated in the text).

The results of the prior simulation exercise in Table 3 indicate that there are no significant biases and the coverage of the posterior intervals match the nominal coverage, as they should. Nevertheless, there seems to be room for improvement in our estimation procedure. The estimates of $\log q$ have a slight negative bias, and as the marginal posteriors of $\log q$ tend to feature relatively long right tails, this suggests that the estimation algorithm fails from time to time in exploring these right tails. The bias is luckily small, but the observation of such a bias does illustrate the capabilities of prior simulation studies to unveil the kind of issues that would otherwise easily go unnoticed, including identification related issues seriously

¹²Ideally, the prior should be such that the probability of unreasonable data generated by means of prior draws is zero, or at least close to zero.

affecting the inference.

5 Empirical Application: Fiscal Multipliers

We illustrate the proposed methods by estimating and analysing a trivariate model used to study fiscal multipliers in the seminal paper of Blanchard and Perotti (2002). The model was later extended by Mertens and Ravn (2014), and more recently considered by Lewis (2021). We find our estimates to be very similar to those reported in Lewis (2021), apart from slightly stronger identification (due to our ability to exploit more information in the data). As in Lewis (2021), we conclude that the unnecessary identifying restrictions employed in Blanchard and Perotti (2002) and Mertens and Ravn (2014) lead to unrealistically high estimates for fiscal multipliers.

The trivariate fiscal policy SVAR illustrates how well data based identification works even in the presence of relatively symmetric and not especially fat-tailed structural shocks, especially when identification is strengthened by allowing for conditionally heteroskedastic shocks. Obviously, if all the shocks are exactly or even approximately Gaussian and homoskedastic, we can draw no interesting conclusions regarding the fiscal multipliers. Nevertheless, the validity of the analysis would not be violated, and we would learn that the shocks are Gaussian and homoskedastic, so no further conclusions could be drawn without additional assumptions (restrictions).

5.1 Model Specification

Our model is defined as in Equation (1) with y_t consisting of tax revenue, government spending and the GDP (in this order), and it slightly differs from that in Lewis (2021) in that we impose no deterministic trend, as is customary in the Bayesian VAR literature. As discussed in Subsection 3.1, we reparametrize ε_t as $\Sigma_t^{1/2}\eta_t$, where the structural shocks are assumed to follow a skewed t -distribution.¹³ To better understand the sensitivity of the results with respect to some of the modeling choices, we consider three specifications, two

¹³We also estimated the model using an even more flexible skewed generalized t -distribution and obtained essentially identical results.

of which exhibit conditionally heteroskedastic structural shocks. Instead of the stochastic volatility process assumed by Lewis (2021), we model the conditional heteroskedasticity by more parsimonious GARCH-processes.¹⁴ Specifically, we fix the unconditional variances of the structural shocks ε_{ti} ($i = 1, \dots, n$) to unity, and define the diagonal elements of Σ_t as

$$\sigma_{i,t}^2 = \sigma_i^2 + \alpha_i \sigma_{i,t-1}^2 + \beta_i \varepsilon_{i,t-1}^2, \quad i = 1, \dots, n, \quad (5)$$

where we restrict both α_i and β_i to be positive. Thus, the vector of the parameters controlling the conditional error variances is $\delta = (\alpha_1, \beta_1, \dots, \alpha_n, \beta_n)$. In addition, in order to keep the volatility processes stationary, we conveniently fix the unconditional shock variance to unity by setting $\sigma_i^2 = 1 - \alpha_i - \beta_i$. Moreover, as this is equivalent to $\sigma_i^2 + \alpha_i + \beta_i = 1$, we impose Dirichlet priors that are consistent with this restriction.

In all specifications, we use non-exchangeable prior on the parameter matrix B to suppress the permutation issue discussed in Section 3.3. In particular, we follow Brunnermeier et al. (2021) and shrink B towards a suitable scaled diagonal matrix (see Appendix B.3, for a detailed discussion). For our baseline specification (*Baseline*), we impose the commonly used Minnesota prior on the autoregressive parameters (see, e.g., Litterman (1986) and Doan, Litterman, and Sims (1984)). A detailed description of this prior is provided in Appendix B.4. Importantly, we do not fix the amount of shrinkage imposed (i.e. the tightness of the prior) on the autoregressive parameters (and on B), but estimate it from the data. For the elements of the vectors of the parameters controlling the law of motion of conditional volatility (δ) and the shapes of the error distributions (γ), we use the default priors explained in Appendix B.2 (see also the discussion in Section 4.1).

To gauge the sensitivity of the results to the shrinkage in the autoregressive parameters, and for maximum comparability to earlier mostly non-Bayesian literature, our second model specification (*No shrinkage*) differs from the first in that we impose an improper constant prior on the autoregressive parameters (i.e. no shrinkage). Imposing the above discussed shrinkage prior instead of an improper constant prior on B proved to be practically inconse-

¹⁴In our framework, modeling conditional heteroskedasticity via GARCH-processes is much more convenient than by latent stochastic volatility processes. The difference between these approaches is likely to be much smaller than when, say, Gaussian shock distributions are assumed because we allow the tails of the shock distributions to partly absorb any sudden changes in volatility.

quential¹⁵, and hence, the results based on the improper prior on B are not reported. Note however, that as Proposition 1 suggests, a proper posterior distribution is obtained, even under the improper prior on B , making such an assessment of prior sensitivity possible in the first place.

The third specification (*Homoskedastic*) differs from the baseline specification only in that the shocks are assumed homoskedastic (i.e. the conditional heteroskedasticity in the shocks is not modelled and hence cannot be exploited for identification).

5.2 Estimation and Labeling Shocks

We estimate the models from a quarterly U.S. data set from 1950Q1 to 2006Q4 (228 observations), consisting of federal tax revenue, federal government consumption and investment, and the GDP. The data are as provided in Lewis (2021), apart from scale transformations of the time series that are more thoroughly discussed in Appendix B.

We generate four chains, each consisting of two thousand draws, using the NUTS algorithm discussed in Section 3.4. The first half of each chain was used for automatic tuning of the algorithm, and the close to ideal values of the \hat{R} convergence diagnostic suggest nice convergence.¹⁶ The final posterior sample hence consists of four thousand only slightly autocorrelated draws. The estimated effective posterior sample size is between one and four thousand draws with respect to almost every parameter in the model, and the sampling took less than a minute on a desktop computer for all the specifications.

As statistical identification of the shocks offers no economic interpretation, the shocks of interest need to be labeled after the estimation of the model. Here we are interested in labeling two different fiscal shocks, the government spending shock and the tax shock. Ideally, this step also includes the assessment of whether the labels of the shocks of interest are supported by the data at all (for a formal approach in the context of sign restrictions see e.g. Lanne and Luoto, 2020). However, we employ a less formal procedure, and make use

¹⁵As our system is relatively small, we have enough observations for the posterior distribution of the 3×3 parameter matrix B to be mostly determined by the data, rendering the prior choice practically inconsequential. This is most likely not the case for larger models.

¹⁶For the \hat{R} convergence diagnostic, values close to unity imply convergence, see, e.g., Vehtari et al. (2021). For all of the three specifications, all values of \hat{R} turned out to be well below the threshold of 1.01.

of the impulse responses (reported in Appendix D) and the forecast error decompositions of the statistically identified shocks (reported in Appendix C). The only potential candidate for a government spending shock is the second structural shock since (for all specifications) it is the only shock having a significant immediate impact on government spending. Another distinctive feature of this shock is the slightly positive short term response in output, which lends additional support to it being the government spending shock.

For a tax shock there are two candidates as both the first and third shocks have a significant immediate effect on the tax revenue. The third shock, however, has an implausibly great effect on output to be a tax shock, as according to forecast error variance decomposition, it accounts for over 90% of the unexpected changes in output in the short to medium term (hence, we refer to it as the output shock) and would consequently also imply implausibly high fiscal multipliers. The response of output to the only remaining candidate for a tax shock, the first shock, also matches the typical characteristics of a tax shock as a negative shock induces a positive response, especially in the long term.

The results obtained under all three specifications are similar. As expected, the Minnesota prior seemed to dampen the dynamics of the system. More interestingly, even under the homoskedastic model (*Homoskedastic*), we are able to identify the same structural shocks and reach the same conclusions, although the probable misspecification of the model resulted in some differences in certainty and magnitude of the exact estimates. Thus, it seems that for the identification of the structural shocks, just letting the shocks to deviate from Gaussian might be enough. However, allowing for conditional heteroskedasticity in the shocks is likely to result in lesser misspecification and more accurate conclusions.¹⁷

As noted in Anttonen, Lanne, and Luoto (2022) narrative records can also be used to guide the labeling of the shocks, and those discussed in Lewis (2021) indeed lend overwhelming support to labeling the first shock as the tax shock. Figure 1 plots the posterior distribution of the identified tax shock in the *Baseline* specification as a function of time. The vertical dashed lines mark negative tax shocks as implied by the narrative records.¹⁸

¹⁷We also estimated a model with dependent shocks, all following the same GARCH process. If this is the correct specification, Proposition 3 guarantees identification. However, the results are similar to those based on the specifications assuming independent shocks, and therefore, to save space, they are not reported in detail, but they are available upon request.

¹⁸The narrative records considered here are the same periods listed and considered in Lewis (2021),

Evidently, the estimated posterior probability of the identified tax shock conforming with the narrative records, i.e., of its being negative at the given dates, is approximately 100 percent.

A similar assessment of the government spending shock is slightly more challenging due to a greater ambiguity in the timing of such shocks (see, e.g., Ramey, 2011). However, at least the posterior probability of a positive government spending shock corresponding to Bush’s Job Creation and Worker Assistance Act of 2002:Q1 is almost 100 percent. This is the one narrative record listed in Lewis (2021) that could be considered as primarily a government spending shock.

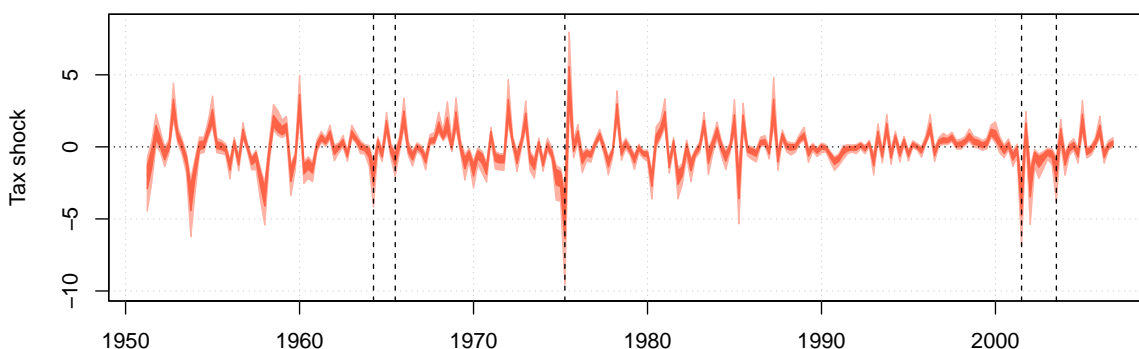


Figure 1: 68% and 90% point-wise credible sets of the identified tax shocks over time for the *Baseline* specification. Vertical dashed lines correspond to quarters for which narrative records suggest negative tax shocks (i.e. tax cuts). The narrative records have not been used in the identification and estimation of the model in any way.

5.3 Fiscal Multipliers

Figure 2 depicts the dynamic multipliers of expansionary fiscal policy shocks, i.e., a negative tax shock and a positive government spending shock, based on the three specifications. In particular, they are the estimated responses in approximate dollars to a one dollar tax cut and a one dollar spending shock (approximate since the data are defined in log-levels, not in dollar terms). Following some of the previous literature, the approximate dollars are defined

apart from Truman’s Revenue Act of 1951:Q1, which falls just outside our sample (we have a posterior distribution for the shocks starting from 1951:Q2). The five negative tax shocks considered are then, Johnson’s Revenue Act of 1964 (1964:Q2), Excise Tax Reduction Act of 1965:Q3, Ford’s Tax Reduction Act of 1975:Q2 (incorrectly timed in Lewis (2021) to Q3), Bush’s Economic Growth and Tax Relief Reconciliation Act of 2001:Q3, and Bush’s Jobs and Growth Tax Relief Reconciliation Act of 2003:Q3.

using the average ratios of the tax and government spending variables to the GDP over the entire sample. In discussing the results, we rely on comparisons to the previous literature, especially to Lewis (2021), whose results are based on a related frequentist approach. The upper and lower row of Figure 2 correspond to his Figures 3 and 4, respectively.

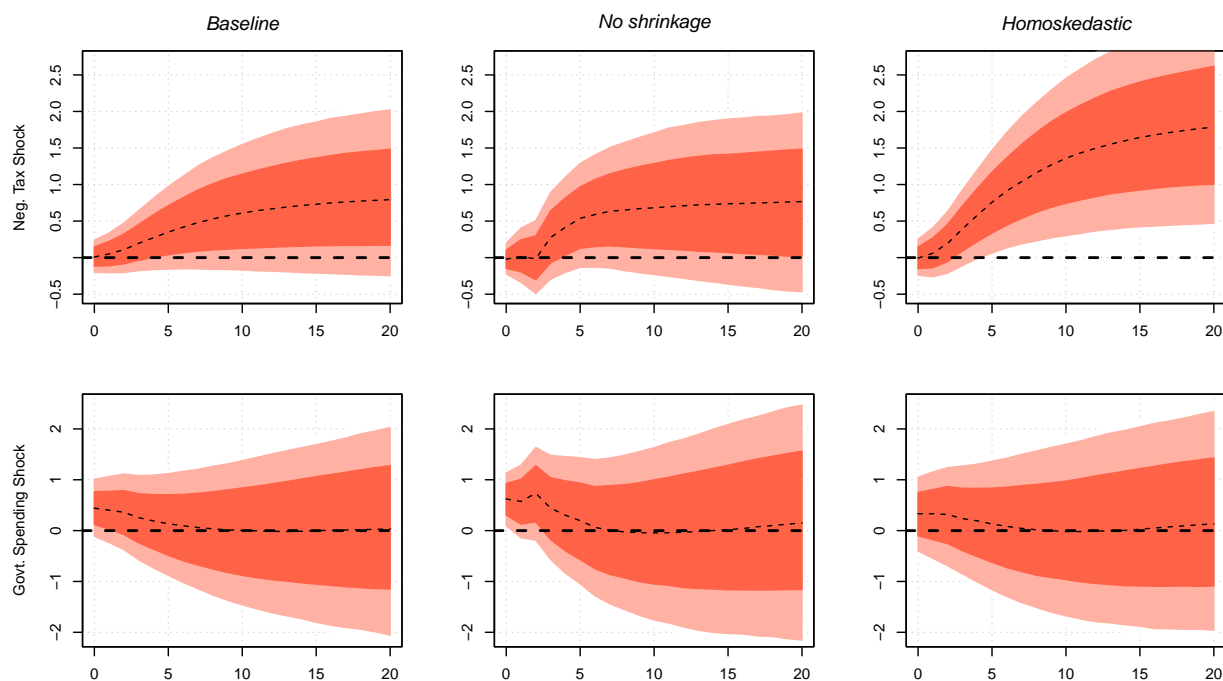


Figure 2: 68% and 90% point-wise credible sets and posterior medians of the dynamic multipliers of a negative tax shock and a positive government spending shock. Each pane depicts the response of output (in approximate dollars) to a one dollar tax cut (top row) and a one dollar government spending shock (bottom row) for one specification (columns). The vertical axis is defined in approximate dollars and the sizes of the tax and spending shocks are fixed to have a unit impact effect on the tax and government spending variables, respectively.

In line with the previous literature, we find the response of output to a negative tax shock to remain approximately zero for the first few quarters, after which the posterior probability of a positive response in any given period rises to at least 85 percent, the exact posterior probability depending on the model specification. Because our models contain no deterministic trend and we are able to accurately capture parameter uncertainty in a finite sample due to Bayesian inference, the posterior distribution of impulse responses indicates greater parameter uncertainty in the medium to long term than much of the corresponding frequentist results in the previous literature (see, e.g., Figure 3 in Lewis, 2021).¹⁹ Of course,

¹⁹Although our posterior distribution gives high posterior probability to a positive response of output to

Bayesian credible intervals and frequentist confidence intervals are not directly comparable. Our posterior means, medians and modes are very close to the point estimates of the dynamic multipliers reported in Lewis (2021), peaking after around 8 quarters (or later) and, with the exception of the most likely misspecified homoskedastic specification, they get close to but stay below one.

Our estimates of the dynamic government spending multiplier differ slightly more from the previous literature than those of the tax multiplier. Most importantly, by exploiting more efficiently the information in the data, we are able to identify a government spending shock that has a clear, yet not an implausibly large, positive contemporaneous impact on output. It is positive with posterior a probability from around 90% to 97%, depending on the specification (for a comparison, see Figure 4 in Lewis, 2021). Despite the significant positive short term response, after two quarters, any point estimates derived from our posterior, however, fall quickly close to zero, and parameter uncertainty increases to the extent that after 8 quarters the dynamic multiplier is equally likely to be positive or negative. According to our posterior distributions, government spending thus has only a short term effect on the output with the dynamic multiplier between zero and one during the first few quarters.

In addition to the dynamic multipliers, it is natural to compare the cumulative multipliers as well, as they might be more relevant in the policy analysis. We map our posterior distribution of the parameters to the posterior distribution of the present value cumulative multipliers as defined in Mountford and Uhlig (2009) and reported in Lewis (2021), and plot the associated marginal posterior distributions in Figure 3. The cumulative multipliers are qualitatively somewhat similar to those reported in Lewis (2021). In particular, while the spending shock has a greater fiscal multiplier in the short term, the cumulative multiplier of the tax shock is greater in the long term.

The cumulative multiplier of tax cuts is very close to zero in the short term, but after 3 to 5 years it has turned positive with 85 to 90 percent posterior probability, the most probable values in the long term lying between one and two. However, the distribution of the cumulative multiplier of the tax shock is highly skewed, and uncertainty is so high that

an exogenous one dollar tax cut, the posterior probability of a negative response is significant as well (10 to 15 percent for *Baseline* and *No shrinkage*).

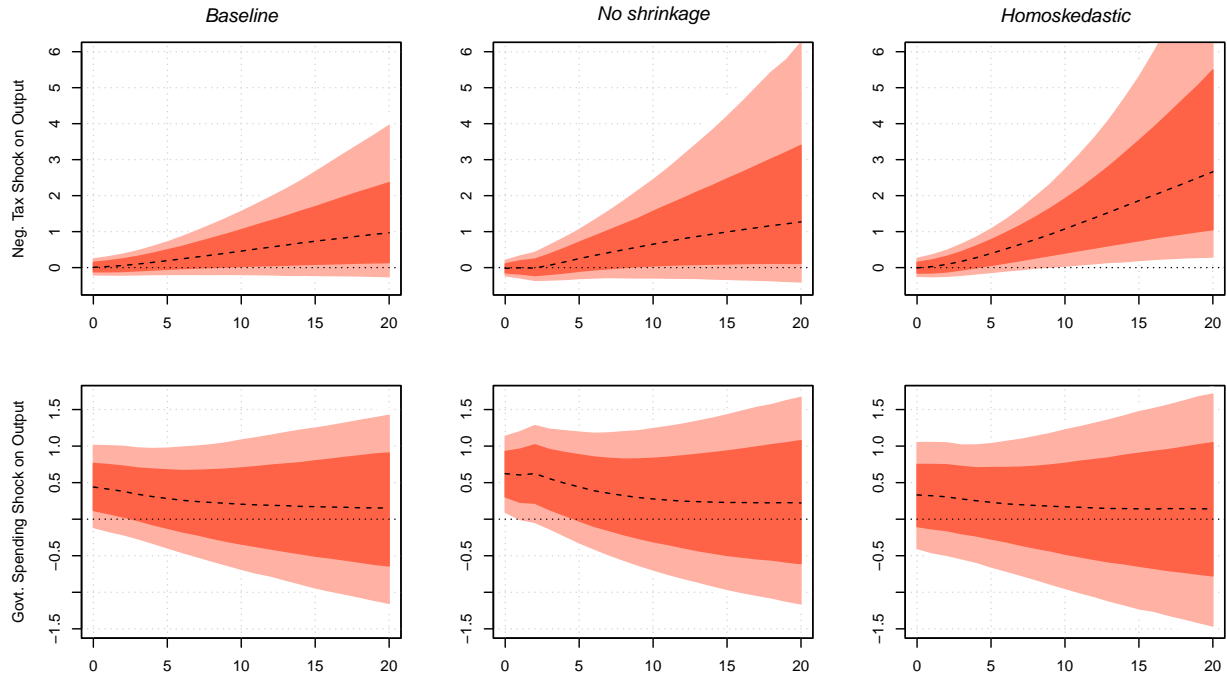


Figure 3: 68% and 90% point-wise credible sets and posterior medians for present value cumulative multipliers for negative tax shock and positive government spending shock as defined in Mountford and Uhlig (2009). The y-axis is defined as in Figure 2.

no definite statements regarding even the sign of the multiplier can be made. These features of the posterior distribution of the cumulative multiplier emphasize the fact that reporting only point estimates may be misleading.²⁰

The posterior distribution of the cumulative spending multiplier seems much more symmetric, and regardless of the high parameter uncertainty, in the short term (during the first year), the cumulative multiplier is most likely positive, the point estimates (posterior mean, median or mode) being close to but less than one. In accordance with the previous literature, we find the cumulative multiplier of the spending shock to be much smaller than that of the tax shock in the long term. However, our results slightly differ from the previous literature (see, e.g., Table 2 in Lewis, 2021) in that we find the government spending multiplier to increase more slowly in the long term. It also falls towards zero (or even turns negative) with a high probability. This is due to our government spending shock having a much more

²⁰For example, the estimated cumulative multiplier (2.06) for the horizon of 20 quarters reported in Lewis (2021) almost exactly coincides with the posterior mean of our *No shrinkage* specification (2.02) that comes closest to Lewis’s (2021) specification. Due to the high skewness of the distribution, this is, however, markedly greater than, say, the median or the mode, and reporting only this point estimate would be misleading.

persistent effect on government spending than on output. In contrast, the effect of the tax shock on the tax revenue seems temporary, whereas its effect on output is persistent.

In conclusion, apart from relatively minor details, our results are very much in accordance with those reported in Lewis (2021), where only conditional heteroskedasticity of the shocks is used for identification. On the other hand, they suggest that the estimated fiscal multipliers reported in both Blanchard and Perotti (2002) and Mertens and Ravn (2014) (where identifying restrictions on the effects of structural shocks are imposed) are unreasonable and not supported by the data, as concluded by Lewis (2021) as well.

The similarity of our results to those of Lewis (2021), and, on the other hand, the dissimilarity to those in Blanchard and Perotti (2002) and Mertens and Ravn (2014) is well depicted in Figure 4, in which the estimated marginal posterior distributions of the two key parameters (the automatic stabilizer for tax revenues, θ_Y , and the instantaneous response of output to tax revenues, ξ_T) as well as the point estimates of Lewis (2021) and Blanchard and Perotti (2002) are drawn. For ease of comparison to the previous literature, we have defined these parameters as in Lewis (2021), and for a more elaborate discussion on these parameters we refer the reader to that paper.²¹ Clearly, the posterior distributions of both θ_Y and ξ_T are very much in accordance with the point estimates reported in Lewis (2021). Also, the differences between our three specifications are minor, corroborating the robustness of our results. While the point estimates reported in Blanchard and Perotti (2002) can be found at the tails of our posterior distribution, thus being hugely improbable, yet remotely possible, the point estimates reported in Mertens and Ravn (2014) (3.13 and -0.35) fall far outside the support of our posteriors. If the identifying restrictions of Blanchard and Perotti (2002) and Mertens and Ravn (2014) were valid, their point estimates should lie well within the support of the posterior distribution obtained via statistical identification. The fact that this is clearly not the case, can be interpreted as evidence against these results. Hence, we can conclude that the fiscal multipliers in those papers are clearly overestimated.

²¹The mapping between these parameters and B as defined in this paper is obtained as follows. First, define H as a result of normalizing the columns of B such that unit diagonal remains (i.e., i 'th column of B is divided by B_{ii} , for $i = 1, 2, 3$). Then, $\theta_Y = H_{13}$ and $\xi_T = \frac{H_{31} - H_{32}H_{21}}{1 - H_{21}H_{12}}$.

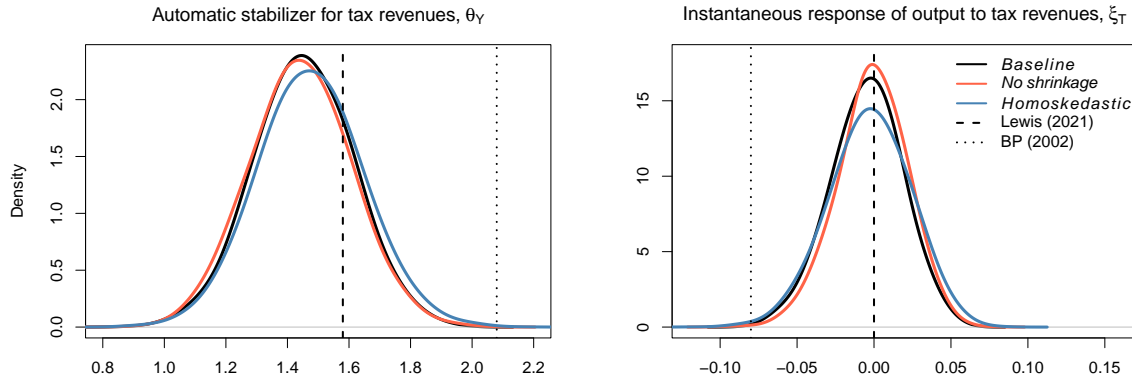


Figure 4: Marginal posterior distributions of the automatic stabilizer for tax revenues, θ_Y , and the instantaneous response of output to tax revenues, ξ_T . For the exact definition of these parameters, see Lewis (2021). Vertical lines depict the point estimates of these parameters as reported in the previous literature. The smoothed posterior densities are constructed from the posterior sample with Gaussian kernels.

5.4 Checking for Identification

As already discussed, the strength of identification depends on the properties of the structural shocks. In particular, the results in Section 2 state that identification of all or part of the shocks is achieved under deviations from normality and in the presence of dependent heteroskedasticity. Therefore, it is of special interest to assess the marginal posterior distribution of the parameters that controls for these aspects.

Figure 5 depicts the the 68% and 90% posterior intervals of the skewness and degree-of-freedom parameters for each structural shock for the *Baseline* specification (and a common prior).²² The shocks are assumed to follow a skewed t -distribution, but according to the left panel of Figure 5, they do not exhibit much skewness that would provide identifying information.

On the other hand, the marginal posterior distributions of the degree-of-freedom parameter depicted in the right panel of Figure 5 clearly suggest that the shocks are not Gaussian, even after accounting for conditional heteroskedasticity. The posterior distributions of the parameters clearly involve much smaller values of the degree-of-freedom parameter than the prior which gives a fairly large prior probability to approximately Gaussian shocks as well.²³

²²The results for the other specifications are very similar and hence not reported, but they are available upon request.

²³That is, Gaussian conditional on zero skewness. For shocks to be Gaussian, the degree-of-freedom parameter of a (skewed) t -distribution should tend to infinity, but already for values as small as 30, the tails

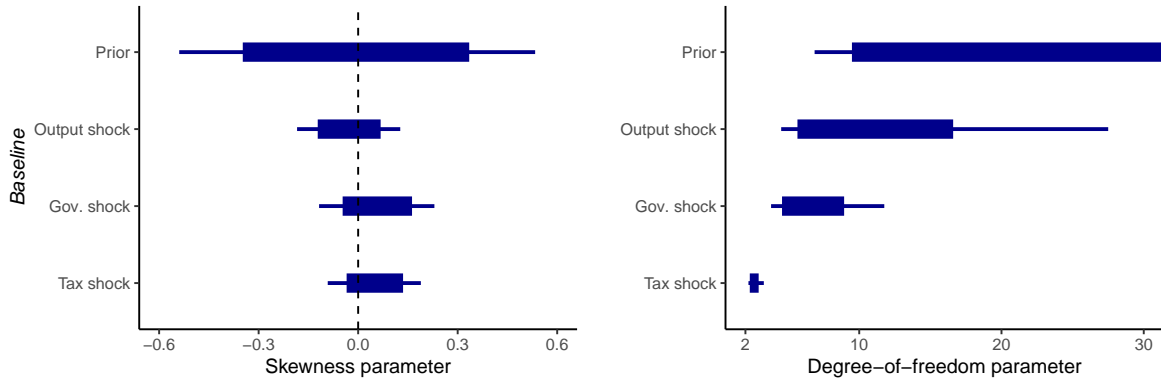


Figure 5: 68% and 90% posterior and prior intervals for the parameters controlling for skewness (left panel) and excess kurtosis (right panel; degree-of-freedom parameter) of the *t*-distributed shocks of the *Baseline* specification.

Especially the tax shock seems to feature significant excess kurtosis (i.e., a small value of the degree-of-freedom parameter) indicating strong identification of that shock. According to Proposition 2, the excess kurtosis of the tax shock alone is sufficient for identification of that shock. In addition, the government spending shock also certainly features sufficient excess kurtosis for us to conclude that, with two of the three shocks being non-Gaussian, there is enough information for the identification of the whole system.

While we have concluded excess kurtosis of the shocks to contain enough information for the identification of the system, it might be strengthened by conditional heteroskedasticity. Figure 6 depicts the posterior distributions of the conditional volatility ($\sigma_{i,t}$) of each shock over time for the baseline specification (the results for the *No Shrinkage* specification are similar).

The volatility of each shock exhibits distinct properties. Whereas the tax shock seems to feature prominent short term spikes, the changes in the conditional volatility of the output shock seem much more persistent. The government spending shock, on the other hand, seems surprisingly homoskedastic, featuring almost no changes in the conditional volatility.²⁴ By Proposition 3, dependent time-varying volatility of all but one shock is sufficient to achieve full identification, provided the shocks have finite fourth moments. Because the value of the degree-of-freedom parameter of the tax shock might very well be less than four (see Figure 5),

of the distribution can be considered approximately Gaussian for all practical purposes.

²⁴The early spike in the volatility of the government spending shock is explained by the disproportionately large movements in government spending at the very beginning of the data.

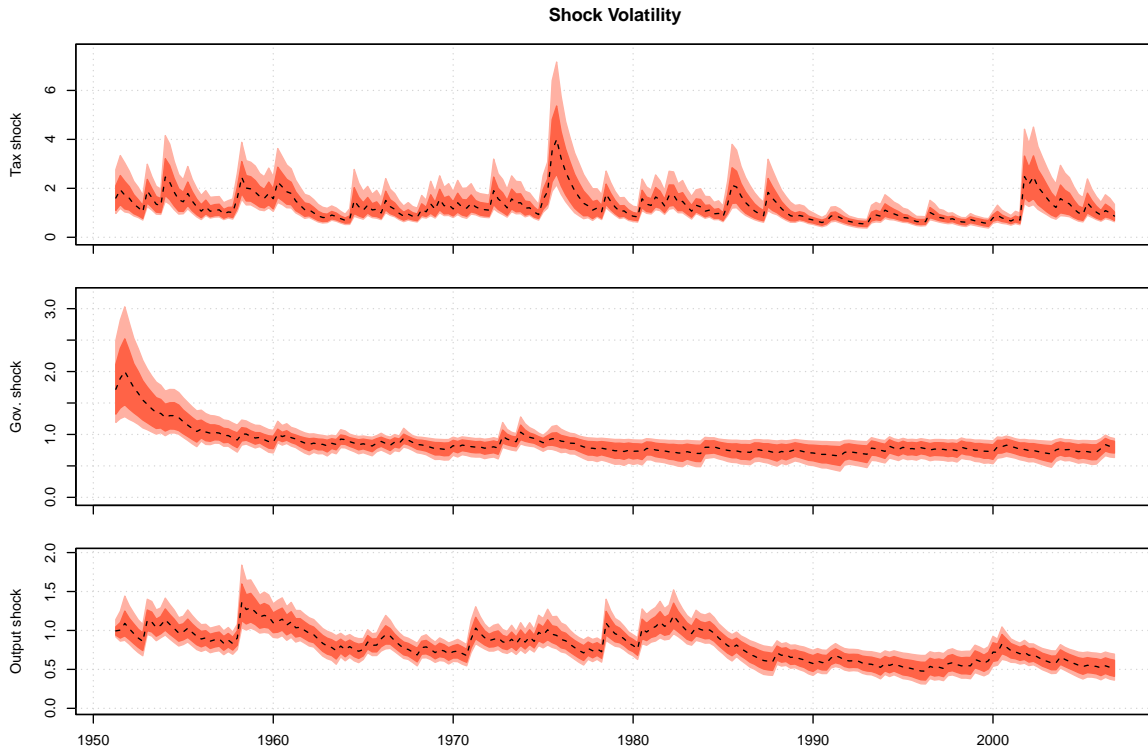


Figure 6: 68% and 90% point-wise credible sets and posterior medians (dashed lines) of the conditional shock volatilities starting from 1951Q2. *Baseline* specification.

the latter condition is hardly satisfied, and, hence, identification is not necessarily achieved via heteroskedasticity. This is not a problem as the excess kurtosis of two of the shocks guarantees full identification, but conditional volatility should still be incorporated into the SVAR model to avoid misspecification and to potentially strengthen the identification.

6 Conclusion

It is well known that if at least $n - 1$ of the independent structural shocks in an n -dimensional SVAR model are non-Gaussian or suitably heteroskedastic, the impact matrix is point identified (up to permutation and signs of its columns). However, if more than one shock is Gaussian, only the columns corresponding to the non-Gaussian shocks are point identified, and to conduct asymptotic inference in the frequentist setup, additional ad hoc restrictions are needed. Moreover, to find the (number of) identified shocks, pre-testing is required. In contrast, when Bayesian methods are used, valid inference is guaranteed by the main result

of this paper, which states that the elements of the impact matrix are always at least set identified under standard assumptions and the bounds of the set cannot exceed the standard deviations of the corresponding reduced-form errors. Importantly, this is the case even when an improper prior on the elements of the impact matrix is used. In addition, the strength of identification can be assessed in a straightforward manner by inspecting the properties of the structural shocks. We also extend the identification results for SVAR models with orthogonal but mutually dependent errors put forth in the previous literature and show that under certain additional assumptions, valid Bayesian inference can be conducted even when all or some of the shocks are dependent.

Efficient utilisation of deviations from Gaussianity requires versatile error distributions, whose implementation calls for advanced computational tools. We propose some suitable distributions and discuss the implementation of the methods in the case of SVAR models. According to simulation experiments, the methods perform well in finite samples. In particular, not surprisingly, simultaneously exploiting multiple different deviations from Gaussianity improves estimation accuracy. We also demonstrate how simulation experiments based on data generated from the prior and posterior can be helpful in the assessment of the strength of identification.

The use of the methods is illustrated in an empirical application to the effects of U.S. fiscal policy. In a trivariate SVAR model for tax revenue, government spending and the GDP, at least two of the identified structural shocks turn out to be non-Gaussian, indicating point identification of the impact matrix. By inspection of the impulse responses and forecast error variance decompositions of the shocks and by the use of narrative records, two of the shocks are labeled the government spending and tax shocks. The resulting fiscal multipliers are similar to what Lewis (2021) obtained, but differ considerably from those in the previous literature. Lewis based identification on heteroskedasticity, while we exploited (also) non-Gaussian features of the data. In contrast, statistical properties of the data were not made use of in identification in the earlier literature.

References

- Anttonen, Jetro, Markku Lanne, and Jani Luoto (2022), *Statistically identified SVAR model with potentially skewed and fat-tailed errors*, Working paper, SSRN.
- Baumeister, Christiane and James D. Hamilton (2015), “Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information”, *Econometrica* 83, 1963–1999.
- Bauwens, Luc and Michel Lubrano (1998), “Bayesian inference on GARCH models using the Gibbs sampler”, *Econometrics Journal* 1, 23–46.
- Bertsche, Dominik and Robin Braun (2022), “Identification of Structural Vector Autoregressions by Stochastic Volatility”, *Journal of Business & Economic Statistics* 40, 328–341.
- Betancourt, Michael (2017), *A Conceptual Introduction to Hamiltonian Monte Carlo*, Working paper, arXiv.
- Blanchard, Olivier and Roberto Perotti (2002), “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output”, *The Quarterly Journal of Economics* 117, 1329–1368.
- Braun, Robin (2021), *The importance of supply and demand for oil prices: evidence from non-Gaussianity*, Bank of England working paper No. 957.
- Brunnermeier, Markus, Darius Palia, Karthik A. Sastry, and Christopher A. Sims (2021), “Feedbacks: Financial Markets and Economic Activity”, *American Economic Review* 111, 1845–79.
- Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017), “Stan: A probabilistic programming language”, *Journal of statistical software* 76, 1–32.
- Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984), “Forecasting and conditional projection using realistic prior distributions”, *Econometric Reviews* 3, 1–100.
- Drautzburg, Thorsten and Jonathan H. Wright (2023), “Refining set-identification in VARs through independence”, *Journal of Econometrics*, forthcoming.

- Gelman, Andrew and Matthew D. Hoffman (2014), “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”, *Journal of Machine Learning Research* 15, 1351–1381.
- Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák (2020), *Bayesian Workflow*, Working paper, arXiv.
- Gouriéroux, Christian, Alain Monfort, and Jean-Paul Renne (2017), “Statistical inference for independent component analysis: Application to structural VAR models”, *Journal of Econometrics* 196, 111–126.
- Griewank, Andreas and Andrea Walther (2008), *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, Society for Industrial and Applied Mathematics, 2nd edition.
- Guay, Alain (2021), “Identification of structural vector autoregressions through higher unconditional moments”, *Journal of Econometrics* 225, 27–46.
- Hoesch, Lukas, Adam Lee, and Geert Mesters (2022), *Robust Inference for Non-Gaussian SVAR models*, Tinbergen Institute Discussion Papers, 22-080/III.
- Kilian, Lutz and Helmut Lütkepohl (2017), “Identification by Heteroskedasticity or Non-Gaussianity”, in: *Structural Vector Autoregressive Analysis*, 491–531.
- Lanne, Markku, Keyan Liu, and Jani Luoto (2022), “Identifying Structural Vector Autoregression via Leptokurtic Economic Shocks”, *Journal of Business & Economic Statistics*, forthcoming.
- Lanne, Markku, Keyan Liu, and Jani Luoto (2023), *Identifying Structural Vector Autoregressions via Non-Gaussianity of Potentially Dependent Structural Shocks*, Unpublished Manuscript.
- Lanne, Markku and Jani Luoto (2020), “Identification of Economic Shocks by Inequality Constraints in Bayesian Structural Vector Autoregression”, *Oxford Bulletin of Economics and Statistics* 82, 425–452.
- Lanne, Markku and Jani Luoto (2021), “GMM Estimation of Non-Gaussian Structural Vector Autoregression”, *Journal of Business & Economic Statistics* 39, 69–81.

- Lanne, Markku, Mika Meitz, and Pentti Saikkonen (2017), “Identification and estimation of non-Gaussian structural vector autoregressions”, *Journal of Econometrics* 196, 288–304.
- Lewbel, Arthur (2019), “The Identification Zoo: Meanings of Identification in Econometrics”, *Journal of Economic Literature* 57, 835–903.
- Lewis, Daniel J. (2021), “Identifying Shocks via Time-Varying Volatility”, *The Review of Economic Studies* 88, 3086–3124.
- Litterman, Robert B. (1986), “Forecasting with Bayesian Vector Autoregressions: Five Years of Experience”, *Journal of Business & Economic Statistics* 4, 25–38.
- Maxand, Simone (2020), “Identification of independent structural shocks in the presence of multiple Gaussian components”, *Econometrics and Statistics* 16, 55–68.
- Mertens, Karel and Morten O. Ravn (2014), “A reconciliation of SVAR and narrative estimates of tax multipliers”, *Journal of Monetary Economics* 68, 1–19.
- Montiel Olea, José Luis, Mikkel Plagborg-Møller, and Eric Qian (2022), “SVAR Identification from Higher Moments: Has the Simultaneous Causality Problem Been Solved?”, *AEA Papers and Proceedings* 112, 481–85.
- Montiel Olea, José Luis, James H. Stock, and Mark W. Watson (2021), “Inference in Structural Vector Autoregressions identified with an external instrument”, *Journal of Econometrics* 225, 74–87.
- Mountford, Andrew and Harald Uhlig (2009), “What are the effects of fiscal policy shocks?”, *Journal of Applied Econometrics* 24, 960–992.
- Neal, Radford (2011), “MCMC using Hamiltonian dynamics”, in: *Handbook of Markov Chain Monte Carlo*, chap. 5.
- Poirier, Dale J. (1998), “Revising Beliefs in Nonidentified Models”, *Econometric Theory* 14, 483–509.
- Ramey, Valerie A. (2011), “Can Government Purchases Stimulate the Economy?”, *Journal of Economic Literature* 49, 673–85.
- Rubio-Ramírez, Juan F., D.F. Waggoner, and Tao Zha (2010), “Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference”, *The Review of Economic Studies* 77, 665–696.

- Stephens, Matthew (2000), “Dealing with Label Switching in Mixture Models”, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62, 795–809.
- Stock, James H. (2008), “What Is New in Econometrics: Time Series, Lecture 7.”, in: *Short Course Lectures, NBER Summer Institute*.
- Ter Braak, Cajo JF and Jasper A Vrugt (2008), “Differential evolution Markov chain with snooker updater and fewer chains”, *Statistics and Computing* 18, 435–446.
- Theodossiou, Panayiotis (1998), “Financial Data and the Skewed Generalized T Distribution”, *Management Science* 44, 1650–1661.
- Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner (2021), “Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC”, *Bayesian Analysis* 16, 667–718.
- Yang, Minxian (2008), “Normal log-normal mixture, leptokurtosis and skewness”, *Applied Economics Letters* 15, 737–742.

Appendix A Proofs

Proof of Proposition 1. We begin by proving Proposition 1 under Assumption 1 with mutual independence in Assumption 1(ii) relaxed to orthogonality. Recall that the (unconditional) reduced-form covariance matrix is given by $\Omega \equiv E(u_t u_t') = B \Sigma B'$, where $\Sigma \equiv E(\varepsilon_t \varepsilon_t')$. We denote the ij -elements of Ω and Σ by ω_{ij} and σ_{ij} , respectively. Under the orthogonality of the elements of ε_t , $\sigma_{ij} = 0$ for all $i \neq j$. Furthermore, to retain the elements of B unconstrained, we normalize the unconditional variances σ_{ii} ($i = 1, \dots, n$) of the elements of ε_t to unity. These imply $\Omega = BB'$, from which, we obtain

$$\omega_{ii} = \sum_{j=1}^n B_{ij}^2 \quad i, j = 1, \dots, n. \quad (\text{A.1})$$

This means that for any $i = 1, \dots, n$, $B_{ij}^2 = \omega_{ii}$ if and only if $B_{ik} = 0$ for all $k = 1, \dots, n$, $k \neq j$, and $B_{ij}^2 < \omega_{ii}$ ($j = 1, \dots, n$) otherwise. Thus,

$$-\omega_{ii}^{1/2} \leq B_{ij} \leq \omega_{ii}^{1/2}, \quad (\text{A.2})$$

must hold for all $i, j = 1, \dots, n$. ■

Proof of Proposition 2. Let us partition ε_t as $\varepsilon_t = (\varepsilon_t^1, \varepsilon_t^2)'$, where the $(r \times 1)$ and $((n - r) \times 1)$ vectors ε_t^1 and ε_t^2 contain the non-Gaussian and Gaussian components of ε_t , respectively. Without loss of generality, we normalize each element of ε_t to have unit (unconditional) variance. Consider an observationally equivalent SVAR process defined by $B^* = BQ$ and $\varepsilon_t^* = Q^{-1}\varepsilon_t$, with Q an $(n \times n)$ orthogonal matrix, where the first r components of ε_t^* have non-Gaussian marginal distributions, and the last $n - r$ ($0 < r < n - 1$) components have Gaussian marginal distributions. We partition ε_t^* accordingly: $\varepsilon_t^* = (\varepsilon_t^{*1}, \varepsilon_t^{*2})'$ with ε_t^{*1} $(r \times 1)$ and ε_t^{*2} $((n - r) \times 1)$.

Partitioning Q as

$$Q = \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix}, \quad (\text{A.3})$$

with Q_1 ($r \times r$), Q_2 ($r \times (n - r)$), Q_3 ($(n - r) \times r$), and Q_4 ($(n - r) \times (n - r)$), we obtain

$$\begin{aligned} I_n &= QQ' \\ &= \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix} \begin{pmatrix} Q'_1 & Q'_3 \\ Q'_2 & Q'_4 \end{pmatrix} \\ &= \begin{pmatrix} Q_1Q'_1 + Q_2Q'_2 & Q_1Q'_3 + Q_2Q'_4 \\ Q_3Q'_1 + Q_4Q'_2 & Q_3Q'_3 + Q_4Q'_4 \end{pmatrix}. \end{aligned} \quad (\text{A.4})$$

Using (A.3) in $\varepsilon_t^* = Q^{-1}\varepsilon_t$, we can write $\varepsilon_t^2 = Q_3\varepsilon_t^{*1} + Q_4\varepsilon_t^{*2}$. By the mutual independence of the elements of ε_t^* , it follows from Lemma 9 in Comon (1994) that Q_3 corresponding to the non-Gaussian part ε_t^{*1} of ε_t^* must be zero. Substituting $Q_3 = 0_{n-r,r}$ into (A.4), we obtain

$$I_n = QQ' = \begin{pmatrix} Q_1Q'_1 + Q_2Q'_2 & Q_2Q'_4 \\ Q_4Q'_2 & Q_4Q'_4 \end{pmatrix}. \quad (\text{A.5})$$

Because $Q_4Q'_4 = I_{n-r}$, Q_4 is an orthogonal matrix. By the orthogonality of Q_4 , it follows that Q_4 is of full rank, and, hence, the conditions $Q_4Q'_2 = 0_{n-r \times r}$ and $Q_2Q'_4 = 0_{r \times n-r}$, provided by (A.5) above, hold if and only if $Q_2 = 0_{r \times n-r}$. Based on these results, (A.3) can be rewritten as

$$Q = \begin{pmatrix} Q_1 & 0_{r \times n-r} \\ 0_{n-r \times r} & Q_4 \end{pmatrix}. \quad (\text{A.6})$$

From $\varepsilon_t^* = Q^{-1}\varepsilon_t$, we hence obtain $\varepsilon_t^1 = Q_1\varepsilon_t^{*1}$ and $\varepsilon_t^2 = Q_4\varepsilon_t^{*2}$, where Q_4 is an $((n-r) \times (n-r))$ orthogonal matrix, as shown above. Also, Q_1 is an $(r \times r)$ orthogonal matrix, which can be seen from (A.5) using the result $Q_2 = 0_{r \times n-r}$.

Next, recall that the elements of ε_t^1 are mutually independent. Also, the elements of ε_t^{*1} are mutually independent, and all of them have non-Gaussian marginal distributions. Therefore, Lemma A.1 of Lanne, Meitz, and Saikkonen (2017) implies that each column of Q_1 contains at most one nonzero element. By the orthogonality of Q_1 , it thus follows that each column of Q_1 has exactly one nonzero element equal to ± 1 . And, for the same reason, also each row of Q_1 has exactly one nonzero element equal to ± 1 . Thus, Q_1 is a signed $(r \times r)$ permutation matrix, and hence, given the signs, the first r structural errors are identified up

to their order, whereas the Gaussian part ε_t^2 of ε_t is only set identified, as Q_4 is an orthogonal matrix.

Similarly, if we partition B as $B = [B_1, B_2]$ with B_1 ($n \times r$) and B_2 ($n \times (n - r)$), by substituting (A.6) into $B^* = BQ$, we immediately see that

$$B_1^* = B_1 Q_1, \quad (\text{A.7})$$

$$B_2^* = B_2 Q_4, \quad (\text{A.8})$$

where $B^* = [B_1^*, B_2^*]$ with B_1^* ($n \times r$) and B_2^* ($n \times (n - r)$). The fact that Q_1 is a signed permutation matrix, as shown above, implies that B_1 is identified up to permutations and sign reversals of its columns, whereas B_2 is only set identified, as Q_4 is an orthogonal matrix. However, the result in (A.2) must hold for any mutually uncorrelated structural error process, and hence it must be that $-\omega_{ii}^{1/2} \leq B_{ij} \leq \omega_{ii}^{1/2}$ ($i, j = 1, \dots, n$), as stated in Proposition 2.

■

Proof of Proposition 3. Let $B^* = BQ$ and $\varepsilon_t^* = Q^{-1}\varepsilon_t$ with Q an ($n \times n$) orthogonal matrix define observationally equivalent SVAR processes, where also ε_t^* satisfies Assumption 2 such that h components of ε_t^* have nonzero skewness. To retain the elements of B unconstrained, we normalize the unconditional variances σ_{ii} ($i = 1, \dots, n$) of the the elements of ε_t to unity.

Let us consider the quantity $E[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}] \equiv \Gamma_{ijk}$. By $\varepsilon_t^* = Q^{-1}\varepsilon_t$, it can be expressed as

$$\begin{aligned} \Gamma_{ijk} &= E \left[\left(\sum_{p=1}^n Q_{ip} \varepsilon_{p,t}^* \right) \left(\sum_{q=1}^n Q_{jq} \varepsilon_{q,t}^* \right) \left(\sum_{r=1}^n Q_{kr} \varepsilon_{r,t}^* \right) \right] \\ &= E \left[\sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n Q_{ip} Q_{jq} Q_{kr} \varepsilon_{p,t}^* \varepsilon_{q,t}^* \varepsilon_{r,t}^* \right] \\ &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n Q_{ip} Q_{jq} Q_{kr} \Gamma_{pqr}^*, \end{aligned} \quad (\text{A.9})$$

where $\Gamma_{ijk}^* \equiv E[\varepsilon_{i,t}^* \varepsilon_{j,t}^* \varepsilon_{k,t}^*]$, $\varepsilon_{i,t}^*$ is the i th, $i = 1, \dots, n$, element of ε_t^* , and Q_{ij} is the (i, j) -element, $i, j = 1, \dots, n$, of Q . Assumption 2(ii) implies that $E[\varepsilon_{it}^* \varepsilon_{jt}^* \varepsilon_{kt}^*] = E[\varepsilon_{it}^{*3}]$ when

$i = j = k$ and zero otherwise. Therefore, (A.9) above, can be written as

$$\begin{aligned}\Gamma_{ijk} &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n Q_{ip} Q_{jq} Q_{kr} \Gamma_{pqr}^* \\ &= \sum_{p=1}^n Q_{ip} Q_{jp} Q_{kp} \Gamma_{ppp}^*.\end{aligned}\tag{A.10}$$

We proceed by considering the following sum of the squared Γ_{ijk} :

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \Gamma_{ijk}^2 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left(\sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n Q_{ip} Q_{jq} Q_{kr} \Gamma_{pqr}^* \right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \sum_{x=1}^n \sum_{y=1}^n \sum_{z=1}^n Q_{ip} Q_{jq} Q_{kr} Q_{ix} Q_{jy} Q_{kz} \Gamma_{pqr}^* \Gamma_{xyz}^* \\ &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \sum_{x=1}^n \sum_{y=1}^n \sum_{z=1}^n \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n Q_{ip} Q_{jq} Q_{kr} Q_{ix} Q_{jy} Q_{kz} \Gamma_{pqr}^* \Gamma_{xyz}^*,\end{aligned}\tag{A.11}$$

where the first equality follows from (A.9). By the orthogonality of Q , we have

$$\sum_{i=1}^n Q_{ip} Q_{iq} = \delta_{pq} = \begin{cases} 1, & p = q \\ 0, & p \neq q \end{cases}\tag{A.12}$$

Using (A.12) above, (A.11) simplifies to

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \Gamma_{ijk}^2 &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \sum_{x=1}^n \sum_{y=1}^n \sum_{z=1}^n \delta_{px} \delta_{qy} \delta_{rz} \Gamma_{pqr}^* \Gamma_{xyz}^* \\ &= \sum_{p=1}^n \sum_{q=1}^n \sum_{r=1}^n \Gamma_{pqr}^{*2}\end{aligned}\tag{A.13}$$

Recall that based on Assumption 2(ii), $\Gamma_{ijk} \equiv \text{E}[\varepsilon_{it} \varepsilon_{jt} \varepsilon_{kt}] = \text{E}[\varepsilon_{it}^3]$ when $i = j = k$ and zero otherwise. Therefore, (A.13) reduces to

$$\sum_{i=1}^n \Gamma_{iii}^2 = \sum_{i=1}^n \Gamma_{iii}^{*2}\tag{A.14}$$

Using (A.10), (A.14) can be written as

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^3 \Gamma_{ppp}^* \right)^2 = \sum_{i=1}^n \Gamma_{iii}^{*2}, \quad (\text{A.15})$$

or equivalently,

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^3 \mathbb{E}[\varepsilon_{p,t}^{*3}] \right)^2 = \sum_{i=1}^n \mathbb{E}[\varepsilon_{i,t}^{*3}]^2, \quad (\text{A.16})$$

From Lemma 15 of Comon (1994), we obtain

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 \leq \sum_{i=1}^n \mathbb{E}[\varepsilon_{i,t}^{*3}]^2. \quad (\text{A.17})$$

Combining (A.16) and (A.17), we have

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 \leq \sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^3 \mathbb{E}[\varepsilon_{p,t}^{*3}] \right)^2, \quad (\text{A.18})$$

On the other hand, by the orthogonality of Q , it must be that $|Q_{ij}| \leq 1$ for all $i, j = 1, \dots, n$, and hence $Q_{ij}^2 \geq Q_{ij}^3$. This implies that $Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \geq |Q_{ip}^3| |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \geq Q_{ip}^3 \mathbb{E}[\varepsilon_{p,t}^{*3}]$ for all $i, p = 1, \dots, n$, and, hence

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 \geq \sum_{i=1}^n \left(\sum_{p=1}^n |Q_{ip}^3| |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 \geq \sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^3 \mathbb{E}[\varepsilon_{p,t}^{*3}] \right)^2. \quad (\text{A.19})$$

By (A.18) and (A.19), it must thus be that

$$\sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 = \sum_{i=1}^n \left(\sum_{p=1}^n |Q_{ip}^3| |\mathbb{E}[\varepsilon_{p,t}^{*3}]| \right)^2 = \sum_{i=1}^n \left(\sum_{p=1}^n Q_{ip}^3 \mathbb{E}[\varepsilon_{p,t}^{*3}] \right)^2, \quad (\text{A.20})$$

from which, we obtain $Q_{ip}^2 |\mathbb{E}[\varepsilon_{p,t}^{*3}]| = |Q_{ip}^3| |\mathbb{E}[\varepsilon_{p,t}^{*3}]|$, or equivalently

$$Q_{ij}^2 (|Q_{ij}| - 1) |\mathbb{E}[\varepsilon_{j,t}^{*3}]| = 0. \quad i, j = 1, \dots, n \quad (\text{A.21})$$

Now, let us assume that at least one structural error has nonzero skewness. Suppose $\varepsilon_{j,t}^{*3} \neq 0$. Then, by (A.21), Q_{ij} must be either zero or ± 1 for all $i = 1, \dots, n$. By the orthogonality of Q , it hence follows that the j th column of Q has exactly one nonzero element equal to ± 1 , and for the same reason this nonzero element ± 1 , is the only nonzero element in the corresponding row of Q . As a result, $\varepsilon_t^* = Q^{-1}\varepsilon_t$ implies that ε_{jt}^* must be equal to one of the elements of ε_t , say the k th, multiplied by ± 1 . By $B^* = BQ$, (A.21) also means that the j th column of B^* is equal to the k th column of B corresponding to the k th structural error ε_{kt} . Obviously, if the number of the skewed structural errors $h > 1$, then (A.21) and the orthogonality of Q ensure that each of the h columns of Q has exactly one nonzero element equal to ± 1 , and they also ensure that each row corresponding to these nonzero elements has exactly one nonzero element. Thus, h elements of ε_t^* are equal to the h skewed structural errors in ε_t , and also h columns of B^* are equal to the h columns of B corresponding to these h skewed structural errors in ε_t .

In other words, if the h skewed errors are ordered first in both ε_t and ε_t^* , then

$$Q = \begin{pmatrix} P & 0 \\ 0 & D \end{pmatrix} \quad (\text{A.22})$$

with some $(h \times h)$ signed permutation matrix P , and an $((n-h) \times (n-h))$ orthogonal matrix D (notice that $QQ' = I_n$ together with (A.22), implies that $DD' = I_{n-h}$). This means that if we partition B as $B = [B_1, B_2]$ with B_1 ($n \times h$) and B_2 ($n \times (n-h)$), by substituting (A.22) into $B^* = BQ$, we immediately see that

$$B_1^* = B_1 P, \quad (\text{A.23})$$

$$B_2^* = B_2 D, \quad (\text{A.24})$$

where $B^* = [B_1^*, B_2^*]$ with B_1^* ($n \times h$) and B_2^* ($n \times (n-h)$). The fact that P is a signed permutation matrix, as shown above, implies that B_1 is identified up to permutation and sign reversals of its columns, whereas B_2 is only set identified, as D is an orthogonal matrix. However, Proposition 1 must hold for any mutually uncorrelated structural error process, and hence it must be that $-\omega_{ii}^{1/2} \leq B_{ij} \leq \omega_{ii}^{1/2}$ ($i, j = 1, \dots, n$). This completes the proof of

part (i).

To prove part (ii), suppose first that only one component of ε_t , say, ε_{lt} has zero skewness, and the other components of ε_t have nonzero skewness. Then, by (A.21), Q_{ij} must be either zero or ± 1 for all $i, j = 1, \dots, n, j \neq l$, and therefore, by the orthogonality of Q , we know that each column of Q except the l th has exactly one nonzero element equal to ± 1 . Similarly, because of the orthogonality of Q , $Q'_i Q_j = 0$ ($i \neq j$), and hence the $n \times (n - 1)$ matrix Q_{-l} , obtained by dropping Q_l from Q , has exactly one zero row, and each of its remaining rows has exactly one nonzero element equal to ± 1 . Therefore, from $Q'_j Q_l = 0$ for $j = 1, \dots, n, j \neq l$, it follows that Q_l has at most one nonzero element (corresponding to the zero row of Q_{-l}), and as $Q'_l Q_l = 1$, this element must equal ± 1 . Thus, $Q = P$, an $(n \times n)$ signed permutation matrix. Obviously, if all components of ε_t have nonzero skewness, by the orthogonality of Q and (A.21), Q must be a signed permutation matrix, so B is identified by sign reversals and ordering of its columns.

Finally, based on Corollary 2 of Lewis (2021), if at least $n - 1$ components of ε_t display time-varying volatility with non-zero autocovariance, B is identified up to sign reversals and ordering of its columns.

■

Proof of Proposition 4. Let us partition ε_t as $\varepsilon_t = (\varepsilon_t^{1'}, \varepsilon_t^{2'})'$, where the $(g \times 1)$ and $((n - g) \times 1)$ vectors ε_t^1 and ε_t^2 contain the independent and dependent components of ε_t , respectively. Without loss of generality, we normalize each element of ε_t to have unit (unconditional) variance. Consider an observationally equivalent SVAR process defined by $B^* = BQ$ and $\varepsilon_t^* = Q^{-1}\varepsilon_t$, with Q an $(n \times n)$ orthogonal matrix, where the first g components of ε_t^* are independent of each other and of the remaining $n - g$ dependent components. We partition ε_t^* accordingly: $\varepsilon_t^* = (\varepsilon_t^{*1'}, \varepsilon_t^{*2'})'$ with ε_t^{*1} $(g \times 1)$ and ε_t^{*2} $((n - g) \times 1)$.

Partitioning Q as

$$Q = \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix}, \quad (\text{A.25})$$

with Q_1 ($g \times g$), Q_2 ($g \times (n - g)$), Q_3 ($(n - g) \times g$), and Q_4 ($(n - g) \times (n - g)$), we obtain

$$\begin{aligned}
 I_n &= QQ' \\
 &= \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix} \begin{pmatrix} Q'_1 & Q'_3 \\ Q'_2 & Q'_4 \end{pmatrix} \\
 &= \begin{pmatrix} Q_1Q'_1 + Q_2Q'_2 & Q_1Q'_3 + Q_2Q'_4 \\ Q_3Q'_1 + Q_4Q'_2 & Q_3Q'_3 + Q_4Q'_4 \end{pmatrix}. \tag{A.26}
 \end{aligned}$$

Notice that independent random variables cannot generally be obtained as linear transformations of uncorrelated but dependent random variables. Recall that the elements of ε_t^1 are independent of each other and of those in ε_t^2 , whereas the elements of ε_t^{*2} are dependent. Therefore, $\varepsilon_t = Q\varepsilon_t^*$ implies that either $Q_2 = 0_{g,n-g}$ or, alternatively, at most one row of Q_2 contains nonzero element(s). However, in the latter case, $Q_4 = 0_{n-g,n-g}$ because if Q_2 has even one nonzero element, the corresponding element of ε_t^1 cannot be independent of the elements of ε_t^2 as long as they contain even one of the elements of ε_t^{*2} , as its elements are dependent.

Suppose the k th ($1 \leq k \leq g$) row of Q_2 has at least one nonzero element. Then, for the corresponding element of ε_t^1 (i.e., $\varepsilon_{k,t}$) to be independent of those in ε_t^2 , it must be that $Q_4 = 0_{n-g,n-g}$. This means that one of the nonzero elements in the k th row of Q_2 is the only nonzero element in the corresponding column of Q . By the orthogonality of Q , it is equal to ± 1 , and hence it must also be the only nonzero element in the k th row of Q_2 . This together with $Q_4 = 0_{n-g,n-g}$, in turn, implies that $[Q'_2, Q'_4]'$ contains columns with all elements equal to zero, but because Q is orthogonal, this leads to a contradiction.

Thus, it must be that $Q_2 = 0_{g \times n-g}$. Substituting $Q_2 = 0_{g \times n-g}$ into (A.26), we obtain

$$I_n = QQ' = \begin{pmatrix} Q_1Q'_1 & Q_1Q'_3 \\ Q_3Q'_1 & Q_3Q'_3 + Q_4Q'_4 \end{pmatrix}. \tag{A.27}$$

Because $Q_1Q'_1 = I_g$, Q_1 is an orthogonal matrix of full rank, and, hence, the conditions $Q_3Q'_1 = 0_{n-g \times g}$ and $Q_1Q'_3 = 0_{g \times n-g}$ in (A.27) above, hold if and only if $Q_3 = 0_{n-g \times g}$.

Therefore,

$$Q = \begin{pmatrix} Q_1 & 0_{g \times n-g} \\ 0_{n-g \times g} & Q_4 \end{pmatrix}. \quad (\text{A.28})$$

From $\varepsilon_t = Q\varepsilon_t^*$, we hence obtain $\varepsilon_t^1 = Q_1\varepsilon_t^{*1}$ and $\varepsilon_t^2 = Q_4\varepsilon_t^{*2}$, where Q_1 is a $(g \times g)$ orthogonal matrix, as shown above. Also Q_4 is an $((n-g) \times (n-g))$ orthogonal matrix, which can be seen from (A.27), using the result $Q_3 = 0_{n-g \times g}$.

Next, recall that the elements of ε_t^1 are mutually independent, and Q_1 is invertible (because it is orthogonal). Also the elements of ε_t^{*1} are mutually independent. Thus, based on Proposition 1 of Lanne, Meitz, and Saikkonen (2017), if at most one element of ε_t^1 has a Gaussian marginal distribution, Q_1 must be a $(g \times g)$ permutation matrix, which implies that, given their signs, the first g independent structural errors are identified up to their order, whereas the dependent structural errors in ε_t^2 are only set identified, as Q_4 is an orthogonal matrix. This completes the proof of part (i).

To prove part (ii), notice that if more than one $(g-r)$ ($1 \leq r < g-1$) of the elements of ε_t^1 are Gaussian (when $g > 3$), Proposition 2 implies that the r non-Gaussian elements of ε_t^1 are globally point identified up to their ordering and sign reversals, and the remaining $g-r$ Gaussian components of ε_t^1 are set identified. In particular, ordering the r non-Gaussian components of ε_t^1 first, we can further partition Q_1 as

$$Q_1 = \begin{pmatrix} P & 0_{r \times g-r} \\ 0_{g-r \times r} & D \end{pmatrix}, \quad (\text{A.29})$$

with P an $(r \times r)$ permutation matrix, and D an $((g-r) \times (g-r))$ orthogonal matrix.

Thus, if we partition B as $B = [\bar{B}_1, \bar{B}_2]$ with \bar{B}_1 ($n \times r$) and \bar{B}_2 ($n \times (n-r)$), by substituting (A.28) into $B^* = BQ$ and using (A.29), we immediately see that

$$\bar{B}_1^* = \bar{B}_1 P \quad (\text{A.30})$$

$$\bar{B}_2^* = \bar{B}_2 \begin{pmatrix} D & 0_{g-r \times n-g} \\ 0_{n-g \times g-r} & Q_4 \end{pmatrix}, \quad (\text{A.31})$$

where $B^* = [\bar{B}_1^*, \bar{B}_2^*]$ with \bar{B}_1^* ($n \times r$) and \bar{B}_2^* ($n \times (n - r)$). As P is a signed permutation matrix, \bar{B}_1 is identified up to permutation and sign reversals of its columns, whereas \bar{B}_2 is only set identified since Q_4 and D are orthogonal matrices. However, Proposition 1 holds for any mutually uncorrelated structural error process, and hence $-\omega_{ii}^{1/2} \leq B_{ij} \leq \omega_{ii}^{1/2}$ ($i, j = 1, \dots, n$), as stated in Proposition 2. Also, the result in Proposition 3 implies that the columns of B_2 corresponding to the skewed components of ε_t^2 are point identified.

Part (iii) follows directly from Proposition 3. The case where g and $n - g$ components of ε_t are independent and dependent, respectively, is a special case of Assumption 2, and therefore, the skewed elements of ε_t^2 are identified. ■

Appendix B Priors and Practical Implementation

In this appendix, we describe in detail the priors used in Section 5 and recommended in general. We also touch upon the subject of practical implementation of the framework discussed in the paper.

Appendix B.1 On Priors and Hyperparameters

Given the likelihood function in (2), the posterior distribution of the parameters is proportional to the product of the likelihood and the prior,

$$p(\theta|y) \propto p(y|\theta)p(\theta), \tag{B.1}$$

where $p(\theta)$ is the density of the prior probability distribution of the parameters θ . Notice that although the notation has been suppressed in (B.1), the prior often depends on hyperparameters controlling, say, the amount of shrinkage imposed by a Minnesota type prior (see Appendix B) on the autoregressive parameters, or the variance of the prior on the elements of the matrix B . Such hyperparameters are difficult to fix a priori since the optimal amount of shrinkage is actually a feature of the data (as well as of the prior), not something known in advance. Fortunately, the hyperparameters lend themselves to a straightforward full Bayesian treatment (see, e.g., Giannone, Lenza, and Primiceri (2015) or Ghosh, Khare,

and Michailidis (2019), for a discussion), which in our framework comes with little to no additional computational cost in many cases.²⁵ In particular, instead of fixing the values of the hyperparameters, denoted by κ , a priori, implying

$$p(\theta|y, \kappa) \propto p(y|\theta)p(\theta|\kappa), \quad (\text{B.2})$$

we may consider the joint posterior distribution of θ and κ given by

$$p(\theta, \kappa|y) \propto p(y|\theta)p(\theta|\kappa)p(\kappa), \quad (\text{B.3})$$

where $p(\kappa)$ is the density of the prior distribution of the hyperparameters (commonly known as the hyperprior), which in practice may or may not be proper, depending on the model at hand. Although, ideally $p(\kappa)$ should be chosen with care, we have found in many cases the data to be extremely informative of the hyperparameters in κ (e.g., with Minnesota type shrinkage), and hence the exact hyperprior is not necessarily of much practical importance, even to the extent that it makes no difference whether it is proper or improper. Under an improper constant $p(\kappa)$, the full Bayesian treatment of the hyperparameters κ has actually an attractive interpretation as minimizing the one-step-ahead out-of-sample forecasting error of the model (in a sense), as discussed in Giannone, Lenza, and Primiceri (2015). In our framework, the treatment of the hyperparameters κ comes with no additional computational complexity, since in practice the typically low-dimensional vector of hyperparameters, κ , can merely be appended to the vector of parameters to be estimated from the data (after appending the hyperprior $p(\kappa)$ to the posterior density, of course).

Although the prior on θ can obviously be set in any way that best suits the problem at hand (as long as it ensures the propriety of the posterior), we have found some default practices to work well in most, if not all, circumstances. The specifics of the prior on γ

²⁵The computational cost of a full Bayesian treatment of hyperparameters in our framework depends on the strength of dependencies between the hyperparameters and the rest of the model parameters. More complex dependencies result in more complex posterior geometries and consequently slower exploration by the estimation algorithm. For instance, in our experience the overall shrinkage towards random walk behaviour can be estimated with practically zero computational cost, whereas the estimation of the relative importance of cross equation lags can be computationally very burdensome especially in large models.

depend, of course, on the distributional assumptions,²⁶ but in this appendix we provide some recommendations and details on the default priors used throughout Section 5. Importantly, the prior on γ should be proper, as although propriety of the posterior may be obtained with improper priors on γ in some cases, only propriety of the prior on γ guarantees the propriety of the posterior. As for the matrix B , in practice, an improper constant prior tends to work well in many cases, but if so required, a prior shrinking the elements of B towards a diagonal matrix (as discussed in the previous subsection) is a natural alternative for a default prior. Note that the amount of shrinkage of such a prior can also be treated as a hyperparameter to be estimated from the data. Moreover, it is our experience that such a data driven shrinkage of B tends to result in sufficiently informative priors for the permutation issue discussed above to be suppressed altogether.

As for the autoregressive parameters and intercept term in π , there exist well established practices regarding the use of Minnesota type priors in the literature on Bayesian vector autoregressions (see, e.g., Banbura, Giannone, and Reichlin (2010), and Giannone, Lenza, and Primiceri (2015)). Such priors typically shrink the parameters towards unit root (or white noise) processes. Further details on our recommended Minnesota type prior used throughout Section 5 are discussed below in Appendix B.4. For now, we merely note that our treatment of hyperparameters efficiently takes care of the Achilles' heel of the Minnesota type priors that the choice of the hyperparameters controls for the amount of shrinkage applied.

Appendix B.2 Shock Distribution and Prior of its Parameters

Before further discussion on priors, we first discuss the point density function of each of the structural shocks. In Section 5 we employ a skewed t -distribution as the distribution of the

²⁶Distributional assumptions, such as whether we employ a skewed t -distribution or an unrestricted sgt -distribution, or if we fix the scale or the variance of the shock distribution (and consequently assume a well defined variance, effectively discarding the possibility of a vast set of fat-tailed probability distributions), may themselves be interpreted as priors on γ . For example, employing a t -distributed shock is equivalent to imposing a degenerate prior with unit mass on the parameter controlling the skewness of a skewed t -distribution, whereas assuming the shocks to follow a normal distribution would equate to setting yet another degenerate prior on the degree-of-freedom parameter of the t -distribution. Not assuming structural shocks to be Gaussian can then merely be interpreted as applying better justified less restrictive priors on the distribution of the structural shocks.

structural shocks (that is, of the independent elements in η_t). The skewed t -distribution can be parameterized in a number of ways, but we choose to stick to the parameterization of the sgt-distribution of Theodossiou (1998), as in Anttonen, Lanne, and Luoto (2022), for flexibility. This parametrization facilitates conveniently obtaining the skewed generalized and Student's t -distributions as well as relaxing the assumption of a well-defined variance to allow for fat tails.

The point density function of the sgt-distribution can be written as

$$f_i(\eta_{it}; \gamma_i := (\lambda_i, p_i, q_i)) = \frac{p_i}{2v_i q_i^{1/p_i} B\left(\frac{1}{p_i}, q_i\right) \left(\frac{|\eta_{it} + m_i|^{p_i}}{q_i v_i^{p_i} (\lambda_i \text{sign}(\eta_{it} + m_i) + 1)^{p_i}} + 1\right)^{\frac{1}{p_i} + q_i}}, \quad (\text{B.4})$$

where $B()$ denotes the beta function and to obtain a skewed t -distribution we set $p_i = 2$. When

$$m_i = \frac{2v_i \lambda_i q_i^{1/p_i} B\left(\frac{2}{p_i}, q_i - \frac{1}{p_i}\right)}{B\left(\frac{1}{p_i}, q_i\right)},$$

the expectation of η_{it} equals zero, and when

$$v_i = q_i^{-1/p_i} \left[(3\lambda_i^2 + 1) \left(\frac{B\left(\frac{3}{p_i}, q_i - \frac{2}{p_i}\right)}{B\left(\frac{1}{p_i}, q_i\right)} \right) - 4\lambda_i^2 \left(\frac{B\left(\frac{2}{p_i}, q_i - \frac{1}{p_i}\right)}{B\left(\frac{1}{p_i}, q_i\right)} \right)^2 \right]^{-1/2},$$

its variance equals unity.

Two free parameters, λ_i and q_i control the shape of the distribution. $\lambda_i \in (-1, 1)$ controls the skewness of the distribution with negative (positive) values for negative (positive) skewness, whereas q_i controls the excess kurtosis analogously to the degree-of-freedom parameter, $d_{0,i}$, of the t -distribution. Particularly, $d_{0,i} = p_i q_i$ or as $p_i = 2$, simply $d_{0,i} = 2q_i$.

The exact prior for λ_i is not particularly important in practice since in our experience, the data seem to be very informative regarding its values. A symmetric Beta prior (with equal shape parameters) on $2\lambda_i - 1$ is an obvious choice, giving a slightly higher prior probability to symmetric rather than extremely skewed distributions. We set the value of the shape parameter of the prior at four (see Figure B.1). If it were unity, the prior would be uniform (as in Anttonen, Lanne, and Luoto (2022)).

An appropriate prior for q_i , or $d_{0,i}$, is of more importance for both practical implementation and non-biased results. The prior should both allow for an approximately Gaussian distribution, in order not to force non-Gaussianity incorrectly, but also to sufficiently tame the posterior geometry (the closer to Gaussian the posterior geometry is, the easier it is for the estimation algorithm to explore it efficiently). As the distribution of q_i is necessarily very skewed and has only a positive support and a long tail, sampling from the posterior should obviously happen in terms of $\log q_i$. More specifically, as we also require $2q_i = d_{0,i} > 2$ for well defined variances, we set a normal prior (with unit mean and standard deviation equal to 2) on $\log(q_i - 1)$ resulting in the shifted log-normal prior on the degree-of-freedom parameter $d_{0,i}$ depicted in Figure B.1. This all results in efficient posterior geometry, gives significant prior probability to an approximately Gaussian shock distribution and reflects our prior notion of reasonable values of $d_{0,i}$.

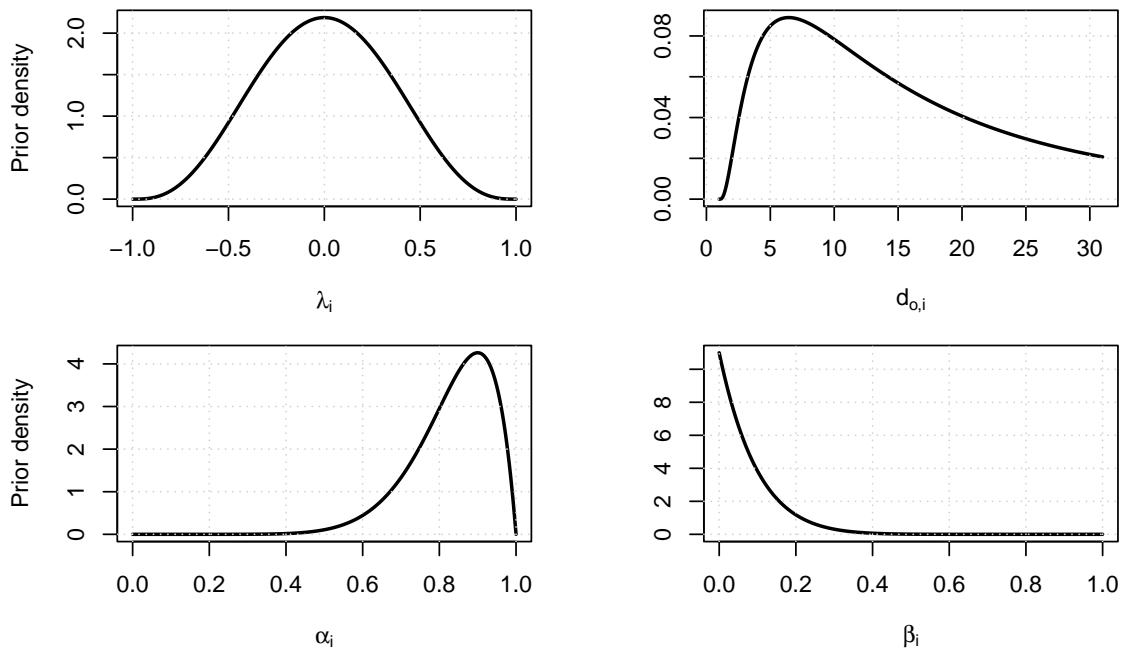


Figure B.1: Prior densities of the skewness parameter (λ_i), the degree-of-freedom parameter ($d_{0,i}$) and the parameters controlling the shock volatility processes (α_i and β_i).

As for the parameters α_i (persistence) and β_i (effect of shocks to volatility) controlling the volatility processes, we deemed it appropriate to impose priors that favour relatively persistent volatility processes. As discussed in Section 5, $\sigma_i^2 + \alpha_i + \beta_i = 1$, where σ_i^2 refers to the constant part of the volatility process. Hence, a Dirichlet prior on these parameters

is an obvious choice. Moreover, this implies marginal beta (distributed) priors on α_i and β_i , and we may tailor the parameters of the prior distribution such that the prior favours values of α_i close to unity (giving a higher prior probability to persistent volatility processes). We set the prior such that the marginal prior on α_i coincides with a beta distribution with the shape parameters equal to 10 and 2, and for β_i with them equal to 1 and 11. Both priors are depicted in Figure B.1.

Appendix B.3 Impact Matrix B

Sampling from the posterior can be done either in terms of B or B^{-1} . The latter is computationally more efficient, especially when B gets large, as likelihood evaluation requires only the knowledge of B^{-1} . On the other hand, the parameterization in terms of B has other desirable properties, such as easier interpretation. Moreover, it facilitates a straightforward implementation of exclusion and sign restrictions directly on B . As the model in Section 5 is not large, we choose to sample directly in terms of B .

The efficient implementation of the NUTS algorithm may require scaling the elements in B . As the error terms can be of different scale, in our experience, satisfactory sampling efficiency with NUTS cannot be obtained without transforming either the parameters or the data. The easiest way to achieve sufficiently uniform scales for the elements in B is to rescale the data prior to estimation. Due to linearity, the posterior distribution of the parameters is obviously invariant to such rescaling, and mapping the resulting posterior sample back to the unscaled posterior, if required, is trivial.

To be specific, in Section 5, prior to estimation, we demean the time series, after which we multiply each series by a factor that results in residual series with unit variance, conditional on our prior mean of the autoregressive parameters.²⁷ Consequently, the elements of B are of roughly similar scale.²⁸ This elementary data transformation not only allows for sufficiently efficient posterior sampling, but also makes it much easier to scale our prior distribution for the parameter matrix B . With no cross equation correlations, the data transformation

²⁷Alternatively, OLS-estimates could be used.

²⁸Although this elementary data transformation procedure results in efficiency sufficient for posterior sampling, there is without doubt room for improvement.

results in B that approximately equals an identity matrix, giving us a natural candidate for the prior mean.

We also employ an almost trivial positive sign restriction on the diagonal elements of B . Although this technically restricts the parameter space to some extent, in our experience, it rarely has any effect on the posterior distribution at all. However, it greatly alleviates the occasional practical difficulties related to trivially symmetric shock permutations (see Section 3.3). After imposing the positivity restriction on the diagonal elements, we set a log-normal prior with a sufficiently large log-variance (4) to avoid excessively restricting the scale of the diagonal elements.

As for the off-diagonal elements of B , the amount of shrinkage applied is much more important, and it has the potential to affect the identification of shocks to a large extent. However, for small models (such as that in Section 5), the data can be sufficiently informative for the prior to be of lesser importance. In such cases even improper constant priors on B might be appropriate, as suggested by Proposition 1. This does not, however, hold generally.

Especially for larger models, an appropriate amount of regularization by means of shrinkage can offer substantial benefits in finite sample performance, alleviating the danger of over-fitting and even turning computationally challenging problems from infeasible to convenient. The appropriate amount of shrinkage is, however, not only a function of the model (including the prior), but also of the data. Therefore, it is ideal to treat the hyperparameters controlling for any prior shrinkage as a set of parameters to estimate, instead of fixing them a priori. This is conveniently carried out in a hierarchical fashion, which requires so called *hyperpriors* for the shrinkage parameters (see Section Appendix B.1).

We set a standard log-normal hyperprior for the standard deviation of the prior of the off-diagonal elements of B . Such a hyperprior gives a significant prior probability to standard deviations close to, but above, zero, with the mode of the hyperprior equaling $e^{-1} \approx 0.37$. This seems sensible given that the prior on the diagonal elements of B gives the most probability mass to the diagonal elements around unity (which is also what the data transformation discussed above approximately leads to). Importantly, this does not fix the prior of the elements of B , but it can be interpreted as a sort of a suggestion for a sensible prior, and the prior itself is updated if it contradicts the data to a sufficiently large extent.

Appendix B.4 Minnesota Prior

Minnesota priors have been the standard for the autoregressive parameters of Bayesian vector autoregressive models in the literature since Litterman (1986) and Doan, Litterman, and Sims (1984). The essence of the Minnesota prior is to shrink the autoregressive parameters towards a unit root process (or some other equally parsimonious process), the shrinkage being stronger for the coefficients on higher lags. Originally, the Minnesota prior also involved stronger shrinkage on cross-equation coefficients, but most of the recent literature lacks this feature albeit its potential benefits are undeniable. This is because the computationally convenient approach of Banbura, Giannone, and Reichlin (2010) (dependent on a structure based on the Kronecker product, not allowing for this feature) has become the standard in the context of Gaussian vector autoregressions.

As already discussed in Section Appendix B.1, the exact way of setting the shrinkage has remained a challenge. We partly circumvent this issue by treating the hyperparameters controlling for the amount of shrinkage as any other parameters in the model, akin to the approach of Giannone, Lenza, and Primiceri (2015). The Minnesota prior we consider here is marginally independent, Gaussian and characterized by the following equations:

$$\mathbb{E}_0[A_1] = J \tag{B.5}$$

$$\mathbb{E}_0[A_h] = 0_{n \times n}, \quad \text{for } h \geq 2 \tag{B.6}$$

$$\text{Var}_0[A_{h,i,j}] = \begin{cases} \left(\frac{\kappa_1}{h^{\kappa_2}}\right)^2 & \text{if } i = j \\ \left(\frac{\kappa_1 \kappa_3}{h^{\kappa_2}}\right)^2 \frac{\omega_{ii}}{\omega_{jj}} & \text{if } i \neq j, \end{cases} \tag{B.7}$$

where J is the $n \times n$ identity matrix for the random walk prior used in the empirical application of Section 5, but could be set in any way seen suitable for the application at hand. Evidently $\kappa_1 \geq 0$ controls for overall tightness of the prior ($\kappa_1 = 0$ would imply a random walk process whereas the case of no shrinkage is approached as $\kappa_1 \rightarrow \infty$), $\kappa_2 \geq 0$ controls for how much tighter the prior is for higher lags (greater values imply faster decay of the coefficients towards zero) and $\kappa_3 \in [0, 1]$ controls the additional cross-equation shrinkage. For univariate dynamics $\kappa_3 = 0$ (the most parsimonious model), and for no additional shrinkage

on cross-equation coefficients $\kappa_3 = 1$ (the least parsimonious model).

The fraction, $\frac{\omega_{ii}}{\omega_{jj}}$, accounts for the different scales of the time series included, and in the previous literature, this term has been usually approximated a priori by the estimated variances of univariate autoregressive processes. Strictly speaking, this fraction is not necessarily needed at all, as we have already scaled our data such that $\omega_{ii} \approx \omega_{jj}$, for all $i, j = 1, \dots, n$ (see the discussion in Appendix B.3) and consequently $\frac{\omega_{ii}}{\omega_{jj}} \approx 1$. However, $\Omega = BB'$, and we may define the Minnesota prior conditional on B in which case the real value of $\frac{\omega_{ii}}{\omega_{jj}}$ can be used directly. To be specific, let us define $A = (A_1, \dots, A_p)$ and consider the case $\theta = (A, B)$ (a slight abuse of notation for illustrative purposes). The prior can then be written as

$$p(\theta) = p(A, B) = p(A|B)p(B), \quad (\text{B.8})$$

where $p(A|B)$ is our Minnesota prior conditional on B .

In Section 5, we fix $\kappa_2 = 1$ and $\kappa_3 = 0.5$ for simplicity and convenience and only estimate the unarguably most important hyperparameter κ_1 from the data. In our experience, this is often sufficient and already a huge improvement over most approaches that fix κ_1 a priori. The common rule-of-thumb value for $\kappa_1 = 0.2$, originally from Sims and Zha (1998), seems a sensible starting point for the initial value of the hyperparameter as well as for the hyperprior. It actually turns out, that much of the time for small to medium-dimensional models, $\kappa_1 = 0.2$ is well within the positive support of the posterior distribution (see, e.g., Figure B.2). We employ a log-normal prior for κ_1 with log-mean of 0.65 and log-variance of 1.5^2 , the hyperprior mode thus being around 0.2. As Figure B.2 depicts, in the empirical application of Section 5 most of the estimated posterior probability mass for κ_1 lies between the values from 0.1 to 0.2 for both specifications, implying on average only slightly tighter priors than the commonly used rule-of-thumb value of 0.2. Hyperparameter uncertainty is, however, very large, and ignoring such uncertainty has most certainly the potential to affect the results. On the other hand, the data seem to be very informative of the most appropriate hyperparameter values, since the marginal posteriors are much more tightly concentrated around specific values than the weakly informative log-normal hyperprior.

Something that we do not consider in this paper, are the refinements to the Minnesota

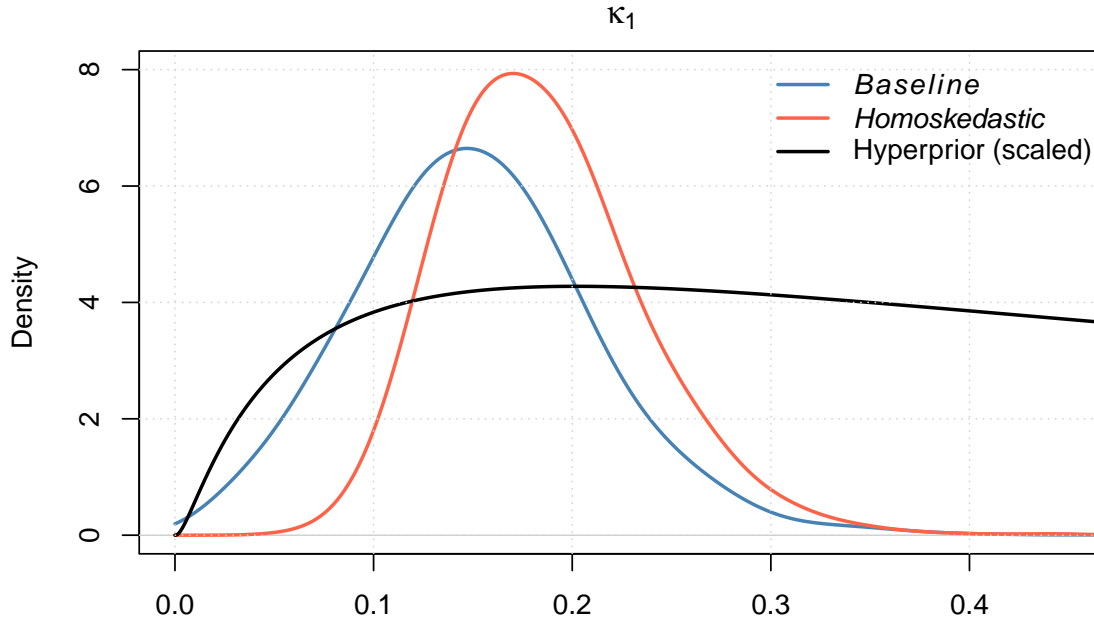


Figure B.2: Posterior and prior densities for the hyperparameter κ_1 controlling for the overall shrinkage of the Minnesota prior in Section 5. The log-normal hyperprior density is scaled (multiplied by a factor of ten) for illustrative purposes. Smoothed posterior densities are constructed from the posterior sample with Gaussian kernels, also for illustrative purposes.

prior that further "favour unit roots and cointegration" (Sims and Zha, 1998). These priors, namely, the sum-of-coefficients prior of Doan, Litterman, and Sims (1984) and the dummy-initial-observation prior of Sims (1993), have become the standard in the literature as they have been found to improve the forecasting performance of Bayesian vector autoregressive models (for a description of these priors see, e.g., Giannone, Lenza, and Primiceri, 2015). However, they are typically implemented through additional dummy observations and consequently are directly applicable only in the Gaussian case. The design and implementation of similar priors should be fairly straightforward, as our framework in principle sets no restrictions whatsoever on what kind of priors can be imposed. However, we leave such explorations for future work.

Appendix B.5 Efficient Model Evaluation and Estimation

The likelihood evaluation essentially comprises (i) a handful of matrix operations that return the structural shocks over the sample, conditional on a set of parameters, and (ii) a loop

over the point density function evaluations at those values (see Equation 2). For efficient implementation it is of utmost importance that the latter is carried out efficiently. For instance, for every shock, the bulk of the point density function evaluation only needs to be done once (i.e., the part of the point density function (see Equation (B.4)) that does not depend on η_{it} (the observed values) needs to be evaluated n times, whereas only the part that depends on η_{it} needs to be evaluated nT times). Also, parallel computing can speed up the likelihood evaluation even for the smallest of models.

Obviously, all the likelihood and prior computations need to be carried out in log-probabilities, the target function supplied to the estimation algorithm being the sum of the (natural) logarithm of the prior probability density and the log-likelihood. Note also that any hyperparameters can be treated exactly as any other parameters in the model ((B.3) can be written as (B.1) by considering κ as a subset of θ), whereas the hyperpriors can be treated exactly as any other priors, making the hierarchical treatment extremely easy in practice.

Any additional identifying restrictions can be implemented within the prior or the likelihood. For instance, exclusion restrictions should be implemented by fixing the appropriate element to zero within the likelihood, leaving the vector of parameters to be estimated shorter. Sign restrictions, in turn, are best implemented by augmenting the prior with whatever parts necessary, e.g. by just setting $p(\theta) = 0$ (zero prior probability), whenever the impulse response function $\Theta_h(\theta)$ is not in accordance with the sign restrictions for any $h \geq 0$.

The whole model is comprised of a single target function, that is the unnormalized log posterior probability of the parameters, i.e. the sum of the log-prior and the log-likelihood:

$$\log \widetilde{p(\theta | y)} = \log p(\theta | y) + \log p(y) = \log p(y | \theta) + \log p(\theta). \quad (\text{B.9})$$

Once this function includes everything deemed necessary, it can be provided together with the data to whatever estimation algorithm deemed appropriate. This effectively makes the business of model building much less dependent on the task of estimating the parameters. Because the estimation procedure stays exactly the same regardless of the changes in model

design, the researcher has much more freedom in designing and adapting the model to the problem at hand (cf. most Bayesian algorithms used in macroeconometrics, e.g. approaches based on Gibbs sampling and/or typical sign restriction algorithms). Specifically, it is this feature that allows us to switch between fixing the hyperparameters controlling for shrinkage and estimating them from the data (as discussed in Appendix B.4), without having to modify the estimation procedure (cf. the approach discussed in Giannone, Lenza, and Primiceri, 2015).

For estimation, we use in this paper (and recommend others to use as well) the actively developed state-of-the-art implementation of the NUTS algorithm in Stan (Carpenter et al., 2017), although in principle any MCMC-algorithm could, of course, suffice.

The NUTS is conveniently readily implemented in an open-source Bayesian inference package called Stan (see, Carpenter et al., 2017) with user friendly interfaces to all the commonly used computing environments, making it highly accessible to practitioners of different technical backgrounds. Moreover, accompanied with this paper, we intend to provide an R-package (see, R Core Team, 2022) that takes care of any tedious technical details regarding priors, likelihood specifications and data transformations. It contains various ready-made tools for the analysis of SVAR models, making the framework proposed in this paper both fast and easy to implement and accessible to all practitioners, regardless of their technical background.

Appendix B.6 Initial Parameter Values

Regardless of the estimation algorithm used, initial parameter values are most probably required, and initializing MCMC-chains from a point as close as possible to, or ideally within, the typical set of the posterior distribution, can significantly speed up convergence. In the context of large models, good and bad initial parameter values can even determine the kind of models that are computationally feasible.

We suggest looking for good initial parameter values by starting from the autoregressive parameters and then looking for the optimal values (in some sense) of the remaining parameters conditional on them. The prior mean (Minnesota prior) is typically a good choice for the initial values of the autoregressive parameters. In the case of weakly informative or even

improper priors, OLS-estimates can also be considered. Conditional on the initial autoregressive parameter values, an estimate of the covariance matrix of the reduced form errors, Ω , can then be computed. Because $\Omega = BB'$ (assuming unit variances for the elements of ε_t), a number of different matrix decompositions can be employed to obtain a reasonable guess for the initial parameter matrix B . The Cholesky decomposition is one obvious alternative, but because the exclusion restrictions implied by it may not be particularly plausible, we opt for a Schur decomposition based approach.²⁹

Conditioning on the initial values of both the autoregressive parameters and B , we obtain initial estimates of the values of the structural shocks, ε_t , for all t . These values can be used to fit skewed t -distribution (or any other distribution) for each shock to obtain initial values for the parameters controlling the skewness and kurtosis of the shocks.

Sensible initial parameter values for any other parameters can be obtained by following a similar procedure, or alternatively, especially if the priors are informative enough, it may suffice to use prior means, modes or medians. For instance, for the parameters controlling the conditional volatility processes we simply use the prior modes as the initial parameter values.

While the resulting values from the procedure just described can very well be used to initialize the posterior chains on most occasions, it is possible to do even better. The automatic differentiation of Stan (Carpenter et al., 2017) not only enables efficient NUTS implementation, but strikingly time efficient point estimation via (Quasi-) Newtonian optimization methods (e.g. the BFGS-algorithm). Thus, the initialization of such optimization algorithms with the initial values just discussed, results in new initial values necessarily closer to (or even at the) mode of the posterior. We have indeed found the additional optimization step to result in more time efficient estimation, the time including both the initial value optimization and posterior sampling.

Importantly, the role of the initial parameter values is just to make the estimation procedure more efficient and they have no effect whatsoever to the resulting posterior distribution per se. Hence, the exact accuracy of these approximations has no effect on the validity of the inference and the initial values should merely be within the realm of possible values.

²⁹More specifically, we have used *sqrtn* function from R-package *expm* by Goulet et al. (2021).

Appendix B.7 Convergence Diagnostics

For any chain attained by means of MCMC-algorithms to be representative of a sample from the posterior distribution of the parameters, the chain needs to have converged to the target distribution. The lack of such convergence may be diagnosed by assessing the within- and between variances of splitted chains. Although the splitting allows for such assessment for single chains as well, the convergence diagnostics are not considered especially reliable without at least two separate chains (see, e.g., Vehtari et al., 2021, who actually recommend a minimum of four chains). Generating multiple chains is also the most efficient way to exploit the parallel processing capabilities of modern computers and to build up the efficient posterior sample size.

The most commonly employed of such convergence diagnostics is the \hat{R} convergence diagnostic of Gelman et al. (2013), also known as the potential scale reduction factor. This statistic is computed for every scalar of interest (e.g. for every parameter in the model) and its values below a certain threshold are considered to indicate convergence. Although the commonly used reference, Gelman et al. (2013), recommends a threshold of 1.1 to be sufficient, Vehtari et al. (2021) point out that the statistic may very well dip below that threshold before convergence only to rise above after further sampling. Therefore, they recommend a much tighter threshold of 1.01 to be used in practice. As they emphasize, regardless of the threshold used, the potential scale reduction factor is not infallible, but it should be considered perhaps a necessary, but by no means a sufficient, indicator of convergence and additional sanity checks for the posterior chains are always recommended.

In practice, we have found the threshold of 1.01 to work well and to be practically attainable on most occasions. In the context of vector autoregressive models considered in this paper we recommend aiming for values less than 1.01 of the potential scale reduction factor with respect to every parameter in the model. However, if the computational resources are otherwise exhausted, the laxer threshold of 1.1 could be allowed for the more slowly mixing parameters (e.g. the shrinkage hyperparameters; nevertheless, especially in these cases a sufficient efficient sample size should be ensured). In our experience, values of the scale reduction factor exceeding the threshold 1.1 should never be allowed for any scalar of

interest, as they very probably indicate of a lack of sufficient convergence to the posterior distribution of the parameters.³⁰

Appendix C Forecast Error Variance Decompositions

Horizon	Tax Revenues			Govt. Spending			Output		
	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3
0	0.869	0.004	0.127	0.019	0.970	0.011	0.010	0.023	0.967
4	0.766	0.008	0.225	0.023	0.964	0.013	0.029	0.018	0.954
8	0.696	0.020	0.284	0.043	0.944	0.014	0.070	0.020	0.910
12	0.643	0.036	0.321	0.067	0.918	0.015	0.109	0.025	0.866
16	0.604	0.049	0.347	0.089	0.895	0.016	0.141	0.031	0.828
20	0.575	0.06	0.365	0.106	0.878	0.016	0.166	0.036	0.797

Table 4: Posterior means of the Forecast Error Variance Decomposition (FEVD) for the specification *Baseline* from the empirical application of Section 5.

Horizon	Tax Revenues			Govt. Spending			Output		
	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3
0	0.888	0.003	0.109	0.036	0.956	0.008	0.015	0.035	0.95
4	0.695	0.007	0.297	0.061	0.923	0.016	0.063	0.023	0.915
8	0.614	0.020	0.366	0.089	0.888	0.023	0.130	0.023	0.846
12	0.562	0.037	0.401	0.140	0.832	0.028	0.170	0.027	0.803
16	0.526	0.051	0.423	0.185	0.785	0.030	0.195	0.032	0.773
20	0.501	0.060	0.440	0.216	0.752	0.032	0.213	0.037	0.750

Table 5: Posterior means of the Forecast Error Variance Decomposition (FEVD) for the specification *No Shrinkage* from the empirical application of Section 5.

Horizon	Tax Revenues			Govt. Spending			Output		
	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3	Shock 1	Shock 2	Shock 3
0	0.787	0.012	0.201	0.019	0.960	0.021	0.006	0.020	0.974
4	0.645	0.021	0.333	0.022	0.951	0.027	0.035	0.020	0.945
8	0.558	0.035	0.407	0.034	0.937	0.028	0.110	0.025	0.865
12	0.500	0.050	0.450	0.057	0.914	0.028	0.186	0.030	0.784
16	0.466	0.060	0.474	0.081	0.891	0.029	0.246	0.035	0.719
20	0.447	0.067	0.486	0.099	0.873	0.029	0.291	0.040	0.670

Table 6: Posterior means of the Forecast Error Variance Decomposition (FEVD) for the specification *Homoskedastic* from the empirical application of Section 5.

³⁰For instance, we would be very careful with inferences based on the posterior draws used in Brunnermeier et al. (2021), as they report the scale reduction factor to be well above the threshold of 1.1 for a significant portion of the parameters in their model for all the specifications considered.

Appendix D Impulse Response Functions

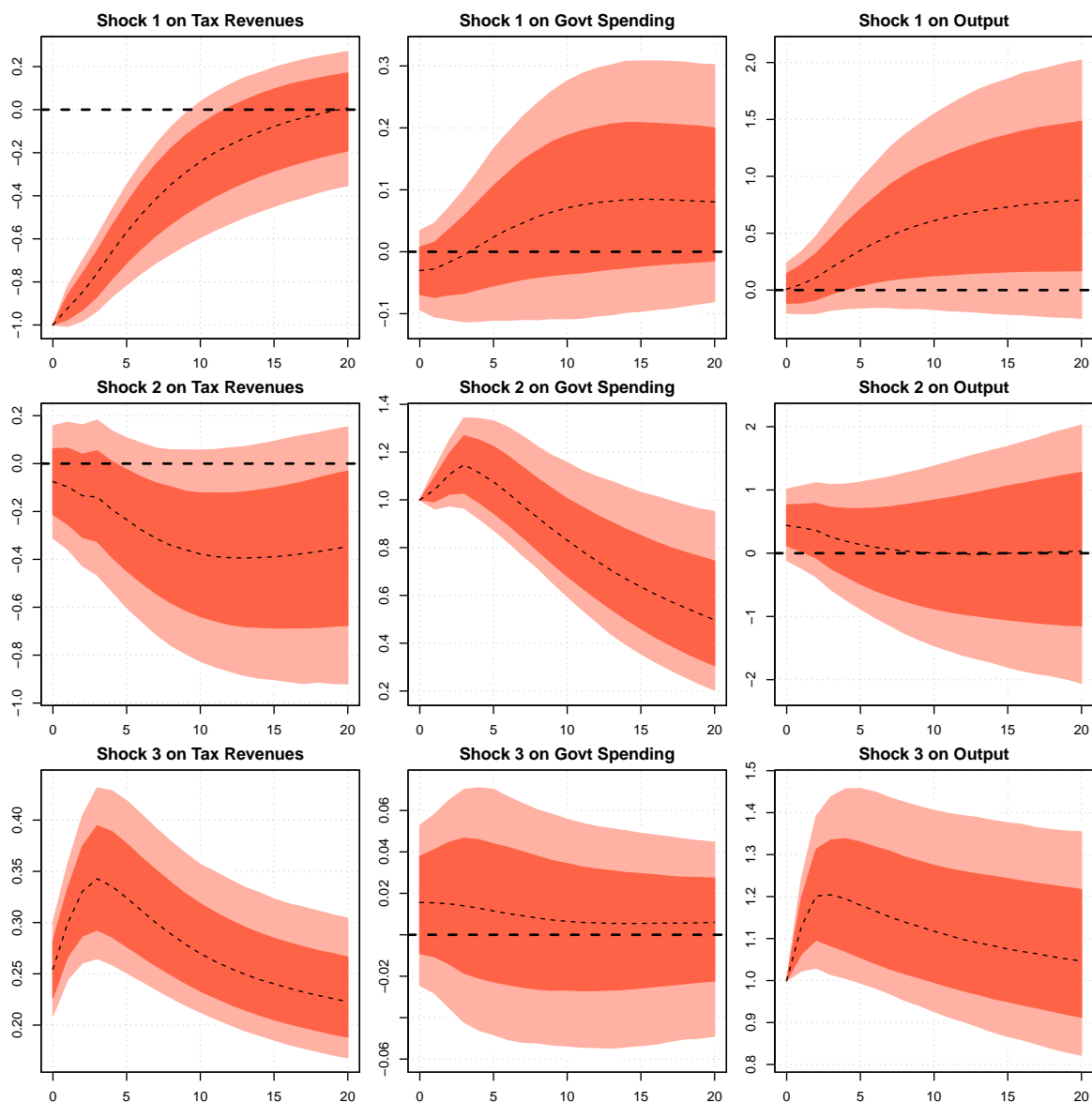


Figure D.1: 68% and 90% point-wise credible sets and posterior medians for all the impulse response functions in the empirical application of Section 5 (*Baseline* specification). The vertical axis is defined in approximate dollars (see text) and the size of the i 'th shock is normalized to have a unit impact effect on the i 'th variable, for $i = 1, 2, 3$.

References (Appendix)

- Banbura, Martha, Domenico Giannone, and Lucrezia Reichlin (2010), “Large bayesian vector auto regressions”, *Journal of Applied Econometrics* 25, 71–92.
- Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin (2013), *Bayesian data analysis*, Chapman and Hall/CRC, 3rd edition.
- Ghosh, Satyajit, Kshitij Khare, and George Michailidis (2019), “High-Dimensional Posterior Consistency in Bayesian Vector Autoregressive Models”, *Journal of the American Statistical Association* 114, 735–748.
- Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2015), “Prior Selection for Vector Autoregressions”, *The Review of Economics and Statistics* 97, 436–451.
- Goulet, Vincent, Christophe Dutang, Martin Maechler, David Firth, Marina Shapira, and Michael Stadelmann (2021), *expm: Matrix Exponential, Log, 'etc'*, R package version 0.999-6.
- R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing.
- Sims, Christopher A. and Tao Zha (1998), “Bayesian Methods for Dynamic Multivariate Models”, *International Economic Review* 39, 949–968.