Identifying Structural Vector Autoregressions via Non-Gaussianity of Potentially Dependent Structural Shocks

Markku Lanne, Keyan Liu, and Jani Luoto*

Faculty of Social Sciences, University of Helsinki

October 12, 2024

Abstract

We show that all shocks in an n-dimensional structural vector autoregression (SVAR) are globally identified up to their order and signs if they are orthogonal and either (i) have zero co-skewness and at most one of them is not skewed or (ii) exhibit no excess co-kurtosis and at least $n-1$ of them are leptokurtic. The former case covers SVAR models with errors following dependent volatility processes. Moreover, if the numbers of both skewed and leptokurtic shocks are smaller than $n-1$, the skewed and leptokurtic shocks are globally identified, while the remaining shocks are set identified. To capture the non-Gaussian features of the data, versatile error distributions are needed. We discuss the Bayesian implementation of an SVAR-GARCH model with skewed t-distributed errors, including the assessment of the strength of identification and checking the validity of exogenous instruments potentially used for identification. The methods are illustrated in an empirical application to the oil market.

Keywords: Structural vector autoregression; Non-Gaussian time series; Identification; Instrumental variable; Bayesian inference

*Corresponding author. E-mail address: jani.luoto@helsinki.fi

1 Introduction

The structural vector autoregressive (SVAR) model facilitates the analysis of economic phenomena by studying the dynamic effects of economic shocks on the variables included. The starting point of the analysis is a reduced-form vector autoregression (VAR) on which various restrictions are imposed to identify the economic shocks of interest. In addition to using restrictions derived from economic theory or institutional knowledge to identify the structural shocks, it has become increasingly common to exploit the statistical properties of the data, such as heteroskedasticity or non-Gaussianity, in identification (see, e.g. Kilian and Lütkepohl (2017, Chapter 14) for a survey of this literature). In the latter case, economic intuition is still needed in labeling the shocks, that is, in giving them an economic interpretation, as they are statistically identified only up to their ordering and signs (or equivalently, the impact matrix of the SVAR model is identified up to permutation and signs of its columns).

In the case of non-Gaussianity, the shocks are typically assumed mutually independent in the statistical identification literature (see Lanne et al. (2017), Herwartz (2018), and Keweloh (2021), among others). This assumption is more restrictive than the orthogonality assumption prevalent in most of the SVAR literature. In particular, as argued by Montiel Olea et al. (2022), it precludes dependent conditional heteroskedasticity of the shocks, while it is likely that they, to some extent, share a common source of economic volatility. Moreover, as pointed out by Kilian and Lütkepohl (2017, Chapter 14), independent structural errors cannot always even be obtained as linear transformations of the residuals of a reduced-form VAR model. Hence, assuming independence may be more restrictive than appears at first sight.

To the best of our knowledge, the independence assumption in non-Gaussian SVAR models has not been relaxed in the previous literature apart from a number of recent attempts concerning the generalized method of moments (GMM) estimation. Guay (2021) introduces a GMM estimator based on all co-skewness and/or all co-kurtosis conditions. Local, but not global, identification is achieved when at most one shock is symmetric and/or has no excess kurtosis and the shocks exhibit no excess co-kurtosis. In addition, he puts forward partial local identification results. While these conditions render the shocks nearly independent, Lanne and Luoto (2021) narrow down the set of co-kurtosis conditions needed for local identitification to a relatively small subset of the asymmetric co-kurtosis conditions. Lanne, Liu and Luoto (in press) show that if the shocks exhibit no excess co-kurtosis and all but one of them are leptokurtic, $n(n-1)$ co-kurtosis conditions, half of which are symmetric and another half asymmetric, suffice for local and global identification in GMM estimation of an n-dimensional SVAR model. Moreover, Mesters and Zwiernik (2022) establish a full identification result similar to the result of Guay (2021) using a different proof strategy, while Velasco (2023) shows global and local identification under independence up to third and/or fourth moments in the structural vector autoregressive moving average (SVARMA) model in the frequency domain. Mesters and Zwiernik (2022) also relax some of the restrictive assumptions in Guay (2021) to show identification even in cases where the errors share a common stochastic variance.

In this paper, we introduce a number of general full and partial identification results, not just pertaining to GMM estimation. In particular, we show that if the shocks are orthogonal, have zero co-skewness and at most one of them is not skewed, they are all globally identified. Moreover, if there are fewer skewed shocks, they are globally identified, while the remaining shocks are not identified (or, in other words, are only set identified). In the same vein, global identification is achieved if the shocks exhibit no excess co-kurtosis and at least all but one of the orthogonal shocks are leptokurtic (or feature persistent conditional heteroskedasticity). Again, if fewer shocks are leptokurtic, only these shocks are globally identified. The results pertaining to leptokurtic shocks equivalently apply to platykurtic shocks, which, however, are expected to be rare in economic applications.

Our full identification results parallel the corresponding results for independent non-Gaussian shocks in the previous literature (see, e.g., Lanne et al. (2017)), while, to our knowledge, similar partial identification results have not been presented previously. The partial identification results also facilitate making simultaneously use of both skewness and excess kurtosis of the structural shocks in identification. In particular, in the case of zero co-skewness and no excess co-kurtosis, the symmetric and leptokurtic as well as the skewed shocks are globally identified, while the remaining shocks are not identified.

We also contribute to the literature on identification of SVAR models using exogenous instruments, or proxy variables, for structural shocks. Even if identification is statistically achieved, it can be strengthened by instruments, and they may also be useful in labeling the structural shocks. However, instruments must be valid, and checking for this is straightforward under statistical identification. To be valid, such instruments, or proxy variables, must be relevant, i.e., correlated with the shock that they are instrumenting, and exogneous to the remaining shocks. Validity checks also help to find the shock of interest and to label it accordingly. Our approach is akin to those of Schlaak et al. (2023), and Braun and Brüggemann (2022), who consider indentification by a combination of heteroskedasticity and external instruments, and a combination of sign restrictions and external instuments, respectively. It extends the method of Schlaak et al. (2023) to the case of non-Gaussian shocks, and compared to Braun and Brüggemann (2022), it avoids using (potentially incorrect) economically (as opposed to statistically) motivated restrictions for identification prior to checking for the validity of instruments.

To efficiently capture non-Gaussian features of the data, it is important to use versatile error distributions in empirical analysis. In this paper, we consider the skewed t-distribution, which nests the Gaussian distribution in the limit, and hence, facilitates checking for identification by skewness and excess kurtosis in a straightforward manner. To avoid misspecification and to potentially strengthen identification, we recommend checking for the presence of conditional heteroskedasticity, and if detected, explicitly incorporating it into the model. Because of the complexity of the SVAR specification, we recommend using Bayesian methods because, as pointed out by Anttonen et al. (2023), Bayesian inference remains valid even if some of the shocks (or equivalently, some columns of the impact matrix) are only set identified. In contrast, frequentist methods would require imposing additional restrictions in case of partial identification (cf. Maxand (2020), or Bertsche and Braun (2022)).

As an empirical illustration, we present an application to the world oil market. In particular, we consider the four-variable "workhorse" SVAR model proposed by Kilian and Murphy (2014), also entertained by Baumeister and Hamilton (2019), Zhou (2020), Braun and Brüggemann (2022), among others, and recently extended by Cross et al. (2022). We assume that the structural shocks are skewed t-distributed and follow GARCH(1,1) processes. Because of sufficient skewness and persistent conditional volatility, all shocks are deemed identified. Kilian's (2008) measure of exogenous oil supply shocks helps in labeling one of the statistically identified shocks as the oil supply shock, but turns out to be a weak instrument. In addition, the sign restrictions used in the previous literature and the oil price uncertainty index of Cross et al. (2022) facilitate labeling the flow demand and precautionary demand shocks, respectively.

The rest of the paper is organized as follows. Section 2 introduces the n-dimensional SVAR model and our main assumption, under which full identification can be shown when at least $n-1$ of the structural shocks either are skewed or follow persistent conditional volatility processes. In Section 3, we show how the SVAR model is fully or partly identified using third moments (Subsetion 3.1) or fourth moments (Subsection 3.2), and discuss strengthening identification via external instruments (Subsection 3.3). Section 4 concentrates on Bayesian inference in the non-Gaussian SVAR model. We explicitly discuss the implementation of the model in the case of skewed t-distributed structural shocks following GARCH(1,1) processes that is expected to be quite widely applicable. Moreover, in Subsection 4.3 we consider checking for specification and identification using the Bayes factor. Section 5 contains the empirical application. Finally, Section 6 concludes.

2 Model

We consider the following structural vector autoregressive (SVAR) model of order p:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t, \tag{1}$$

where y_t is the n-dimensional time series of interest, ν is an $(n \times 1)$ intercept term, and $A_1, \ldots, A_p$ 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 $(n \times 1)$ vector of structural errors ε_t, i.e., $u_t = B\varepsilon_t$. The covariance matrix of ε_t is $E(\varepsilon_t\varepsilon_t') \equiv \Sigma$, and thus the (unconditional) covariance matrix of the reduced-form errors is $E(u_t u_t') = B\Sigma B' \equiv \Omega$. We denote the (i, j) elements of Ω and Σ by ω_{ij} and σ_{ij}, respectively.

Throughout, we assume that the components of ε_t are mutually uncorrelated but not necessarily independent. The following is our main assumption, under which we derive results related to identification by skewness of the structural shocks. It will be slightly modified later on to facilitate making use of excess kurtosis in identification.

Assumption 1

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

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

(iii) *The component processes* ε_{it}, $i = 1, \dots, n$, *are orthogonal and have zero co-skewness.*

(iv) *At least one component of* ε_t *has nonzero skewness.*

Parts (i) and (ii) of Assumption 1 are standard in the SVAR literature with the exception that we assume each component of ε_t to have a finite third moment in addition to zero mean and finite positive variance. As we will see in Section 3, this additional assumption is needed to identify matrix B under Assumption 1(iii) that the structural shocks are only mutually orthogonal instead of being independent, as typically assumed in the statistical identification literature (see, e.g., Lanne et al. (2017)). Part (iii) also requires the components of ε_t to have zero co-skewness, i.e., $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}) = 0$ for all $i, j, k = 1, \dots, n$, excluding $i = j = k$. In contrast, nothing is assumed about the fourth co-moments, so the structural shocks can follow (univariate) time-varying volatility processes that are mutually dependent. As will be discussed in detail in Section 3, if there is sufficient heteroskedasticity in the structural shocks, we can make use of Lewis's (2021) result that matrix B is fully identified, provided the fourth (co-)moments of the components of ε_t exist because Assumption 1(i)–(iii) covers Lewis's assumptions.

Even if the structural shocks are homoskedastic, B may be identified if they are sufficiently skewed, and Assumption 1(iv) is related to this. It seems quite different from the statistical identification literature, where typically at most one of the structural shocks is allowed to be Gaussian to guarantee identification of B. Also in our setup, full identification is achieved when at most one component of ε_t is symmetric, but as shown in Section 3, for partial identification it suffices that only one component has nonzero skewness.

3 Partial and full identification of structural errors

In this section, we provide full and partial identification results based on skewness and excess kurtosis of the structural errors. Our main result is that all skewed structural errors (and the corresponding columns of B) are globally point identified (up to ordering and signs) even when they are not mutually independent but only orthogonal. The remaining shocks are only set identified, but as pointed out by Anttonen et al. (2023), the bounds of the identified set of B are relatively narrow. Moreover, if at most one of the shocks has zero skewness, all structural errors are globally point identified. The latter is akin to the result in Lewis (2022) that if all but one of the structural shocks exhibit time-varying volatility with non-zero autocovariance, they are all identified. These results are useful because they allow for dependent time-varying volatility processes. While it is not necessary to assume any parametric model to capture conditional heteroskedasticity, if such a model is known, it can be used to strengthen (achieve) identification. For instance, in the empirical application in Section 5, we specify first-order GARCH processes for the volatilities of the structural shocks.

Indeed, skewness and time-varying volatility are not mutually exclusive. As an example, consider the case $\varepsilon_t = \Sigma_t^{1/2}\epsilon_t$, where $\Sigma_t = \text{diag}(exp(\sigma_{1t}^2), \ldots, exp(\sigma_{nt}^2))$, $\sigma_{it}^2 = \phi_i\sigma_{i,t-1}^2 + \xi_{it}$, $\xi_t = (\xi_{1t}, \ldots, \xi_{nt})' \sim N(0, \Sigma_\xi)$, and $\epsilon_t \sim N(0, I_n)$. Then, if the innovations ϵ_{it} and ξ_{it} are correlated, the marginal distribution of ε_{it} is skewed and leptokurtic (see, for example, Yang (2008)). Such a situation arises, for instance, when the structural error at time t is correlated with the volatility at time $t+1$, and is especially common in financial data, where negative correlation between the return at time t and the realized volatility at time $t+1$ is often observed (see, e.g., Yu (2005), Omori et al. (2007), and Chan and Grant (2017)). However, if the structural shocks exhibit conditional heteroskedasticity, it should, in general, be incorporated into the model to avoid misspecification and to sharpen identification even if the model is identified by skewness.

We also consider the case where all structural shocks exhibiting excess kurtosis are either leptokurtic or platykurtic. Then the leptokurtic (platykurtic) shocks are globally point identified, and if at most one of them is mesokurtic (i.e., has no excess kurtosis) all shocks are globally point identified. This result holds for both skewed and symmetric structural errors.

However, identification by leptokurtic shocks is somewhat restrictive in that dependent time-varying volatility processes are precluded, although the structural errors can otherwise be dependent. While identification cannot be established by excess kurtosis, when the structural shocks follow mutually dependent time-varying volatility processes, it is plausible that excess kurtosis still strengthens identification.

Finally, in Subsection 3.3 we show how identification achieved based on our results can be strengthened by external instruments, or proxies, each of which is correlated with only one of the structural shocks. External instruments can also be useful in labeling the shocks, as statistical properties of the data only facilitate identification up to their ordering and signs. Following the seminal articles of Stock and Watson (2012) and Mertens and Ravn (2012), the proxy-SVAR literature is burgeoning, but coming up with exogenous instruments may not be straightforward, and even credibly exogenous proxies may be weak. Apart from strengthening identification by combining the information in the proxies with the statistical properties of the structural shocks, our results may help assess the validity of the external instruments.

3.1 Identification using third moments

In this subsection we consider identification of the B matrix in SVAR model (1) under Assumption 1. Specifically, we show that all structural shocks, or equivalently, all columns of matrix B, are identified, when at least $n-1$ of the n structural shocks follow a skewed distribution. Full identification when at most one of the structural shocks is Gaussian, have been put forth previously (see, e.g., Lanne et al. (2017)), but they require mutual independence of the shocks, while our assumptions allow the shocks to be dependent. Moreover, as pointed out in Section 2, its parts (i)–(iii) cover Lewis's (2021) assumptions (specifically, his Assumption A), so his results guarantee full identification of B, provided the fourth co-moments of ε_t exist. In particular, global identification is achieved if at least all but one of the components of ε_t exhibit time-varying volatility with non-zero autocovariance.

Partial identification can be achieved if at least one of the n components of ε_t is skewed. Specifically, under Assumption 1, the columns of B corresponding to the skewed shocks are globally point identified, while the remaining columns are only set identified. Partial identification by non-Gaussianity and conditional heteroskedasticity has previously been considered by Maxand (2020), Guay (2021), and Bertsche and Braun (2022). However, our approach is more general than these in that it allows for dependent time-varying volatility processes of the structural shocks, and Maxand (2020) also assumes the shocks to be independent.

Our first identification result is stated as Proposition 1, and its proof is found in Appendix A. As always in the case of statistical identification of SVAR models, identification is achieved only up to signs and permutation of the columns of B, which has to be taken into account in conducting statistical inference.

Proposition 1 *Suppose $\varepsilon_t = B^{-1}u_t$ satisfies Assumption 1, and assume that r $(0 < r \leq n)$ components of ε_t have nonzero skewness, and that the remaining $n - r$ components of ε_t have zero skewness. Assume further without loss of generality that these $n - r$ symmetric components are ordered last. Let $B = [B_1, B_2]$ with B_1 $(n \times r)$ and B_2 $(n \times (n - r))$.*

(i) If $r < n - 1$, the $(n \times r)$ matrix B_1, corresponding to the r skewed 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.

(ii) If at least $n - 1$ $(r \in \{n - 1, n\})$ components of ε_t have nonzero skewness, the full matrix B is globally point identified up to sign reversals and ordering of its columns.

3.2 Identification using fourth moments

Leptokurtic shocks are likely to be common in economic applications (see, e.g., Lanne et al. (2022) and the references therein). Lanne et al. (2022) showed how the SVAR model can be identified by leptokurtic shocks in the GMM framwork, while here we consider full and partial identification by the fourth moment structure of the structural errors. To that end, we make the following Assumption 2, which is a slight modification of Assmption 1.

Assumption 2

(i) The error process $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{nt})'$ is a sequence of stationary random vectors with each component ε_{it}, $i = 1, \ldots, n$, having zero mean, a finite positive variance σ_{ii}, and finite third and fourth moments.

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

(iii) *The component processes ε_{it}, $i = 1, \dots, n$, are orthogonal and have no excess co-kurtosis.*

(iv) *Of the components of ε_t s $(0 < s \leq n)$ are all either leptokurtic or platykurtic, and each of the remaining $n - s$ components has zero excess kurtosis.*

Parts (*i*) and (*ii*) are almost identical to those in Assumption 1 (ε_{it} is assumed to have a finite fourth moment in addition to having a finite third moment). While orthogonal, as opposed to independent, shocks are still allowed for, Assumption 2(*iii*) differs from its counterpart in Assumption 1 in that zero co-skewness is replaced by no excess co-kurtosis. Specifically, this means that, normalizing the unconditional variances σ_{ii} $(i = 1, \dots, n)$ to unity, $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}) = 1$ when $i = k, j = l \neq k$ or $i = l, j = k \neq l$ or $i = j \neq k = l$ $(i, j, k, l = 1, \dots, n)$, and zero otherwise (except for $i = j = k = l$). While the shocks need not be independent, dependent time-varying volatility processes are precluded, which makes this assumption somewhat more restrictive than Assumption 1(*iii*). Finally, part (*iv*), adapted from Lanne et al. (2022), states that there is at least one non-Gaussian shock and if there are multiple non-Gaussian shocks, the excess kurtosis of each of them must have the same sign. As discussed in Lanne et al. (2022), this assumption is not restrictive because platykurtic shocks are highly unlikely in economic applications, while leptokurtic shocks abound.

According to part (*i*) of Proposition 2 below, global point identification of the s leptokurtic components of ε_t (or equivalently, of the corresponding columns of B) is achieved (up to their sign reversals and ordering) under Assumption 2. The proof is found in Appendix B). In addition, part (*ii*) states that the full matrix B is globally point identified up to sign reversals and ordering of its columns if at least $n - 1$ elements of ε_t are all leptokurtic (or platykurtic). This result follows from Proposition 2 in Lanne et al. (2022). Notice also that because Assumption 2 does not impose any restrictions on the third (co-)moments of ε_t, also all skewed components of ε_t are globally point identified even if they are mesokurtic (i.e., have zero excess kurtosis) according to Proposition 1 (provided Assumption 1 is satisfied).

Proposition 2 *Suppose $\varepsilon_t = B^{-1}u_t$ satisfies Assumption 2, and assume that s $(0 < s \leq n)$ components of ε_t are leptokurtic (platykurtic), and that the remaining $n - s$ components of ε_t have zero excess kurtosis. Assume further, without loss of generality, that the $n - s$ mesokurtic components are ordered last.*

(i) Suppose $s < n - 1$, and let $B = [B_1, B_2]$ with B_1 $(n \times s)$ and B_2 $(n \times (n - s))$, with B_1 and B_2 corresponding to the s leptokurtic (or platykurtic) and $n - s$ mesokurtic components of ε_t, respectively. Then B_1 is globally point identified up to sign reversals and ordering of its columns, while B_2 is set identified.

(ii) If at least $n - 1$ $(s \in \{n - 1, n\})$ components of ε_t are all leptokurtic (platykurtic), the full matrix B is globally point identified up to sign reversals and ordering of its columns.

3.3 Strengthening Identification via External Instruments

Identification can be strengthened by external instruments, or proxies, each of which is correlated with one of the shocks and uncorrelated with the rest of them. Braun and Brüggemann (2022) combine sign restrictions and external instruments in identification, and following their lead, we consider such instruments to strengthen identification achieved by non-Gaussianity (or conditional heteroskedasticity).

Let us denote by $m_t = (m_{1t}, \ldots, m_{kt})'$ the $k \times 1$ vector of external variables that identify k of the n structural shocks and augment the SVAR model in (1) by equations for the elements of m_t to obtain

$$\tilde{y}_t = \tilde{\nu} + \tilde{A}_1 \tilde{y}_{t-1} + \cdots + \tilde{A}_p \tilde{y}_{t-p} + \tilde{B}\tilde{\varepsilon}_t, \tag{2}$$

where $\tilde{y}_t = (y_t', m_t')'$, $\tilde{\nu} = (\nu', \nu_m')'$, $\tilde{\varepsilon}_t = (\varepsilon_t', \eta_t')$,

$$\tilde{A}_i = \begin{pmatrix} A_i & 0_{n,k} \\ \Gamma_{1i} & \Gamma_{2i} \end{pmatrix} \text{ and } \tilde{B} = \begin{pmatrix} B & 0_{n,k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix}. \tag{3}$$

Here $\tilde{B}$ is assumed invertible, Γ_{1i} and Φ are $(k \times n)$ coefficient matrices, Γ_{2i} is a $(k \times k)$

coefficient matrix, Σ_η is a $(k \times k)$ parameter matrix, and η_t is a $(k \times 1)$ vector of zero mean measurement errors, which are assumed to be orthogonal to the structural errors ε_t.

The $n \times k$ blocks of zeros in the upper right corners of $\tilde{A}_i$ $(i = 1, \ldots, p)$ and $\tilde{B}$ reflect the fact that the instruments are external to the SVAR model in (1). Without further restrictions, the elements of $\tilde{\varepsilon}_t$ are only set identified, and the identification of ε_t and η_t only depends on their separate processes. To see this, consider a SVAR process defined by $\tilde{B}^* = \tilde{B}\tilde{Q}$, where $\tilde{B}^*$ has the same structure as $\tilde{B}$ (that is, $\tilde{B}^*$ is a lower triangular matrix), and $\tilde{\varepsilon}_t^* = \tilde{Q}^{-1}\tilde{\varepsilon}_t$. If $\tilde{Q}$ is an $((n+k) \times (n+k))$ orthogonal matrix, this SVAR process is observationally equivalent to (2). As shown in Appendix C, then $\tilde{Q} = \text{diag}(\tilde{Q}_1, \tilde{Q}_4)$, where $\tilde{Q}_1$ and $\tilde{Q}_4$ are $(n \times n)$ and $(k \times k)$ orthogonal matrices, respectively.

The result that $\tilde{Q} = \text{diag}(\tilde{Q}_1, \tilde{Q}_4)$ implies that $\tilde{\varepsilon}_t^* = \tilde{Q}^{-1}\tilde{\varepsilon}_t$ can be written as $\varepsilon_t = \tilde{Q}_1\varepsilon_t^*$ and $\eta_t = \tilde{Q}_4\eta_t^*$. Thus, if all the elements of ε_t (η_t) are Gaussian, $\tilde{Q}_1$ $(\tilde{Q}_4)$ remains an orthogonal matrix, and hence the elements of ε_t (η_t) are indeed only set identified, as stated in Braun and Brüggemann (2023) and Arias et al. (2021). However, if ε_t and/or η_t satisfy Assumption 1, all their skewed components are point identified and the remaining shocks are set identified, as Proposition 1 can be directly applied to $\varepsilon_t = \tilde{Q}_1\varepsilon_t^*$ and $\eta_t = \tilde{Q}_4\eta_t^*$ separately.[1] Obviously, if at most one of the components of ε_t (η_t) has zero skewness, all the components of ε_t (η_t) are point identified (i.e., $\tilde{Q}_1$ $(\tilde{Q}_4)$ is a signed permutation matrix). In the same vein, if ε_t and/or η_t satisfy Assumption 2, all their leptokurtic components are point identified, and their remaining components are set identified. Also, if at least $n-1$ $(k-1)$ components of ε_t (η_t) are all leptokurtic, all the components of ε_t (η_t) are point identified.

In much of the proxy SVAR literature, external instruments are assumed to be unpredictable by lagged values of $\tilde{y}_t$, which means that $\Gamma_{1i} = \Gamma_{2i} = 0$ for $i = 1, \ldots, p$ in (2). With this restriction, the last k equations of (2) can be expressed as

$$m_t = \nu_m + \Phi\varepsilon_t + \Sigma_\eta^{1/2}\eta_t. \tag{4}$$

This suggests that identifying information from external instruments can be incorporated

[1]In this case, $\tilde{Q}_1$ $(\tilde{Q}_4)$ is of the form $\text{diag}(P, D)$, where P is a signed permutation matrix and D is an orthogonal matrix, with the dimensions depending on the length of ε_t (η_t) and the number of skewed components (see (A.14) in Appendix A.

into (2) by imposing zero restrictions on the elements of Φ. In particular, without loss of generality, let us collect the k shocks of interest into the vector ε_{1t} and the remaining shocks into the vector ε_{2t}, so $\varepsilon_t = (\varepsilon_{1t}', \varepsilon_{2t}')'$ and partition Φ accordingly as $\Phi = [\Phi_1, \Phi_2]$ with Φ_1 and Φ_2 $(k \times k)$ and $(k \times (n-k))$ matrices, respectively. Then, if the instruments in (4) are valid for ε_{1t}, the following two conditions hold:

$$\Phi_2 = 0_{k,n-k}, \tag{5}$$

and

$$\Phi_1 \neq 0, \quad \text{rank}(\Phi_1) = k, \tag{6}$$

where (5) and (6) are the exogeneity and the relevance conditions, respectively. Provided Φ is point identified, these conditions can be assessed by Bayes factors as discussed in Section 4.

Whether Φ is identified, in turn, depends on the properties of the structural errors ε_t: if at most one of the components of ε_t is symmetric and/or at least $n-1$ components of ε_t are all leptokurtic (or exhibit persistent conditional heteroskedasticity), all columns of Φ are point identified. This can be seen by comparing the observationally equivalent SVAR models characterized by $\tilde{B}$ and $\tilde{B}^* = \tilde{B}\tilde{Q}$:

$$\tilde{B}^* = \begin{pmatrix} B^* & 0_{n,k} \\ \Phi^* & \Sigma_\eta^{*1/2} \end{pmatrix} = \begin{pmatrix} B & 0_{n,k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix} \begin{pmatrix} \tilde{Q}_1 & 0_{n,k} \\ 0_{k,n} & \tilde{Q}_4 \end{pmatrix} = \begin{pmatrix} B\tilde{Q}_1 & 0_{n,k} \\ \Phi\tilde{Q}_1 & \Sigma_\eta^{1/2}\tilde{Q}_4 \end{pmatrix}. \tag{7}$$

As shown above, $\tilde{Q}_1$ is a signed permutation matrix when ε_t is point identified, and, hence, Φ is point identified because $\Phi^* = \Phi\tilde{Q}_1$.

4 Bayesian Inference

We estimate the parameters of the SVAR model in (1) by Bayesian methods. As Anttonen et al. (2023) point out, Bayesian analysis of the model is possible even if the parameters are only set identified because it only requires a proper posterior, which can be established by using proper priors. This facilitates valid Bayesian inference without additional restrictions also when point identification of some (or all) shocks due to Gaussianity and homoskedasticity fails. Furthermore, the identified set of B is not unconstrained, but its bounds are actually relatively narrow: the absolute value of the (i, j)-element of B cannot exceed $\omega_{ii}^{1/2}$ (see, e.g., Anttonen et al. (2023) and the references therein).

Because of set identification, the posterior distribution of the parameters can be obtained by simulation also when point identification fails. This only requires an estimation algorithm that is efficient enough to facilitate estimation of a large number of parameters under the complex topology of the identified set. To this end, we employ the very efficient Hamiltonian Monte Carlo (HMC) algorithm of Anttonen et al. (2023), based on the No-U-Turn Sampler (NUTS) of Gelman and Hoffman (2014), which is able to provide accurate estimates of the posterior distribution of the parameters even when all structural shocks are not point identified.

It is important to realize that the existence of third and/or fourth moments of the elements of the error vector is required only for establishing identification, but estimation based on a non-Gaussian parametric distribution (such as the skewed t-distribution used in this paper) need not depend in any way on the higher moments per se. In particular, no empirical estimation of higher moments is required in this case. This should alleviate the concerns of some authors (e.g., Montiel Olea et al. (2021)), who have been worried about potential weak identification of non-Gaussian SVAR models because accurate empirical estimation of the third and higher moments may require more data than is often available.

4.1 Likelihood function

For Bayesian inference, we need the likelihood function and the prior distribution of its parameters. Here we only derive the likelihood function of the SVAR model in (1), but the derivations generalize in a straightforward manner to the augmented SVAR model in (2) by replacing ε_t by $\tilde{\varepsilon}_t = (\varepsilon_t', \eta_t')'$. Let us start out by specifying the distribution of the structural errors, which should be sufficiently flexible to fully exploit various types of deviations from Gaussianity, as well as potential heteroskedasticity. Moreover, to avoid misspecification, it is important to approximate the true error distribution as accurately as possible. To this end, following Anttonen et al. (2023), we reparametrize ε_t as $\Sigma_t^{1/2}\epsilon_t$, where $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{nt})'$, and assume that each individual structural error ϵ_{it} $(i = 1, \ldots, n)$ has zero mean and unit variance and follows a skewed t-distribution. Hence, Σ_t is the conditional covariance matrix of ε_t, $Var[\varepsilon_t|\sigma_t, \mathcal{F}_{t-1}]$, where $\sigma_t = (\sigma_{1t}, \ldots, \sigma_{nt})'$ and $\mathcal{F}_{t-1} = \{\varepsilon_1, \ldots, \varepsilon_{t-1}, \sigma_1, \ldots, \sigma_{t-1}\}$. Furthermore, $\Sigma_t = \text{diag}(\sigma_t^2)$ with $\sigma_t^2 = \sigma_t \odot \sigma_t$, where $\odot$ denotes the Hadamard product. While the conditional variances σ_{it}^2 of the components of ε_t may be mutually dependent, the elements of ϵ_t are assumed mutually and temporally independent, and also σ_t and ϵ_t are assumed independent of each other. The skewed t-distribution depends on two parameters, λ_i and q_i that control the skewness and excess kurtosis of ϵ_{it} (and, hence, of ε_{it}), respectively, and it nests the Gaussian distribution as a limiting case.

While these assumptions cover a wide variety of volatility processes, including stochastic volatility (SV) and autoregressive conditional heteroskedasticity (ARCH) type processes, we provide the likelihood function in the special case of a generalized ARCH (GARCH) law of motion for σ_t. Assuming, say, a stochastic volatility process would lead to only minor changes in the likelihood function, but computational complexity would increase because in that case integration over a high-dimensional σ_t is required. Specifically, we parametrize σ_{it} as

$$\sigma_{it}^2 = c_i + \alpha_i \sigma_{i,t-1}^2 + \beta_i \varepsilon_{i,t-1}^2, \quad i = 1, \ldots, n, \tag{8}$$

where both α_i and β_i are restricted positive. In addition, in order to normalize the unconditional shock variance to unity and to keep the volatility processes stationary, we set $c_i = 1 - \alpha_i - \beta_i$, where $\nu_i > 0$.

Let us next collect the parameters controlling the law of motion of σ_t and the shape of ϵ_{it} $(i = 1, \ldots, n)$ into the vectors $\delta = (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)'$ and $\gamma_i = (\lambda_i, q_i)'$ $(i = 1, \ldots, n)$, respectively. Then, substituting ε_t for $\Sigma_t^{1/2}\epsilon_t$ in (1) and recalling the mutual (and temporal) independence of the elements of ϵ_t, we can write the density function of the distribution of the data y as

$$p(y|\theta) = |\det(B)|^{-T} \prod_{i=1}^{n} \prod_{t=1}^{T} \sigma_{it}^{-1} f_i(\sigma_{it}^{-1} \iota_i' B u_t(\pi); \gamma_i), \tag{9}$$

where $\theta = (\pi', \text{vec}(B)', \delta', \gamma')'$, $\pi = \text{vec}(\nu, A_1, \ldots, A_p)'$, $\gamma = (\gamma_1', \ldots, \gamma_n')'$, ι_i is the ith unit vector, $u_t(\pi) = y_t - \nu - A_1 y_{t-1}, \cdots, A_p y_{t-p}$, and $f_i(\cdot)$ $(i = 1, \ldots, n)$ is the density function of ϵ_{it}. To retain the elements of B unconstrained, the unconditional variances of ε_{it} are normalized to unity: $E(\varepsilon_{it}^2) = E(\sigma_{it}^2) = \sigma_{ii} = 1$ $(i = 1, \ldots, n)$. Alternatively, the diagonal elements of B can be normalized to unity, in which case we assume that $\sigma_{ii} > 0$ $(i = 1, \ldots, n)$.

4.2 Prior Distribution

In this subsection, we briefly describe the priors used in our empirical applications in Section 5 that are likely to be even more widely applicable in economic applications. They closely resemble those in Anttonen et al. (2023), and we refer to their paper for a more detailed discussion.

As mentioned in Subsection 4.1, two parameters, λ_i and q_i, control the shape ϵ_{it} $(i = 1, \ldots, n)$ that follows a skewed t-distribution. Skewness is controlled by $\lambda_i \in (-1, 1)$, with negative (positive) values indicating negative (positive) skewness, and $\lambda_i = 0$ for a symmetric random variable. We assume a symmetric Beta prior with equal shape parameters on $2\lambda_i - 1$. A uniform prior would be obtained by setting the value of the shape parameters of the Beta distribution at unity, but we set them at four, which gives a slightly higher prior probability to symmetric than extremely skewed distributions.

The parameter q_i controls the excess kurtosis of ϵ_{it} in the same manner as the degree-of-freedom parameter of a Student's t-distribution (the degree-of-freedom parameter of a skewed t distributed random variable is $2q_i$). It takes only positive values, and, according to our experience, its distribution is skewed and has a long tail. Therefore, we sample from it in terms of $\log(q_i)$. In particular, we assume a normal prior (with mean unity and standard deviation equal to 2) on $\log(q_i - 1)$, resulting in a shifted log-normal prior on the degree-of-freedom parameter that guarantees a well defined variance $(2q_i > 2)$. As discussed in Anttonen et al. (2023), this prior on q_i results in efficient posterior geometry, gives significant prior probability to an approximately Gaussian shock distribution and reflects our prior notion of reasonable values of the degree-of-freedom parameter.

As for the parameters α_i and β_i controlling the volatility processes, we assume priors that favor persistent volatility. In particular, the restriction $c_i + \alpha_i + \beta_i = 1$ (where $c_i > 0$ captures the constant part of the volatility process) suggests a Dirichlet prior on α_i and β_i. Based on the properties of a Dirichlet distribution, this, in turn, implies marginal beta priors on these parameters, and we tailor the parameters of the prior distribution to favor values of α_i close to unity (giving a higher prior probability to persistent volatility processes). Following Anttonen et al. (2023), we set the prior such that the marginal priors on α_i and β_i coincide with beta distributions with the shape parameters equal to 10 and 2, and 1 and 11, respectively.

We assume the Gaussian independent Minnesota prior for the elements of π characterized by the following equations (see Litterman (1986) and Doan, Litterman, and Sims (1984)):

$$E[A_1] = I_n \tag{10}$$

$$E[A_h] = 0_{n\times n}, \quad \text{for } h \geq 2 \tag{11}$$

$$\mathrm{Var}[A_{h,i,j}] = \begin{cases} \left(\frac{\kappa_1}{h^{\kappa_2}}\right)^2 & \text{if } i = j \\ \left(\frac{\kappa_1\kappa_3}{h^{\kappa_2}}\right)^2 \frac{\omega_{ii}}{\omega_{jj}} & \text{if } i \neq j, \end{cases} \tag{12}$$

where $\kappa_1 \geq 0$ controls the overall tightness of the prior, $\kappa_2 \geq 0$ controls for the tightness of the higher lags (greater values imply a faster decay of the coefficients towards zero) and $\kappa_3 \in [0,1]$ controls the additional cross-equation shrinkage.

The fraction $\frac{\omega_{ii}}{\omega_{jj}}$ accounts for the different scales of the variables of the SVAR model, and, in the literature, this term is usually approximated a priori by the estimated variances of the univariate autoregressive processes. However, as discussed in Anttonen et al. (2023), the efficient implementation of the NUTS algorithm requires scaling the elements in B such that they have a roughly similar scale. This can be accomplished, for instance, by first demeaning the time series and then multiplying each of the resulting series by a factor that results in residual series with a unit variance. As a result, $\omega_{ii} \approx \omega_{jj}$, for all $i, j = 1, ..., n$ and consequently $\frac{\omega_{ii}}{\omega_{jj}} \approx 1$. This data transformation also makes it much easier to scale the prior distribution of the parameter matrix B. In particular, it results in B whose diagonal elements are approximately one, giving us a natural candidate (i.e., an identity matrix) for the prior mean.

To avoid the arbitrary element in setting the prior variance, we treat κ_1 as a hyperparameter and estimate it from the data as explained in Anttonen et al. (2023). 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 the common rule-of-thumb value 0.2. For simplicity and convenience, we fix $\kappa_2 = 1$ and $\kappa_3 = 0.5$.

We restrict the signs of the diagonal elements of B positive. Although this technically restricts the parameter space to some extent, in our experience, it rarely has any effect on the posterior distribution. However, it greatly alleviates the practical difficulties related to occasional switching of the shocks due to the fact that ε_t is only identified up to sign reversals and ordering of its elements (see Brunnermeier et al. (2021) and Anttonen et al. (2022, 2023) for further discussion). After restricting the diagonal elements positive, we set a log-normal prior with a sufficiently large log-variance (4.0) 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 than in the case of the diagonal elements, and it may enhance the identification of shocks to a large extent, if properly chosen. However, as the appropriate amount of shrinkage is not only a function of the model (including the prior), but also of the data, it is also here ideal to estimate the hyperparameters controlling the shrinkage. To this end, we set a standard log-normal hyperprior on the standard deviation of the prior of the off-diagonal elements of B. Such a hyperprior with mode equal to $\mathrm{e}^{-1} \approx 0.37$ gives a significant prior probability to values of standard deviation close to, but above, zero. This seems sensible given that the prior on the diagonal elements of B gives the most probability mass to the diagonal elements around unity. Importantly, imposing this hyperprior does not fix the prior of the elements of B, but it can be interpreted as some kind of a suggestion for a reasonable prior that is updated if it contradicts the data to a sufficiently large extent.

4.3 Checking for Specification and Identification

According to Propositions 1 and 2, the skewed and leptokurtic shocks are point identified. The strength of identification based on skewness and excess kurtosis of the shocks can be assessed in a straightforward manner by inspecting the posterior densities of the respective parameters of the skewed t-distributions specified for the structural errors. Moreover, as discussed in Subsection 3.1, if at least $n-1$ shocks exhibit persistent conditional heteroskedasticity, all shocks are point identified. The Bayes factor of a SVAR specification with the errors following conditional volatility processes against a homoskedastic SVAR model may yield evidence in favor of sufficient conditional heteroskedasticity. Even if identification is achieved by higher moments, checking for the presence of conditional heteroskedasticity by the Bayes factor is, in general, advisable to avoid misspecification.

The Bayes factor is obtained as the ratio of the marginal likelihoods of two competing models. Specifically, the Bayes factor of Model 1 (M_1) against Model 2 (M_2) equals

$$BF_{12} = \frac{\int p(\theta_1 \mid M_1)p(Y \mid \theta_1, M_1)d\theta_1}{\int p(\theta_2 \mid M_2)p(Y \mid \theta_2, M_2)d\theta_2}, \tag{13}$$

where $p(\theta_i \mid M_i)$ is the prior density of the parameters θ_i of the model M_i, $i = 1, 2$. The quantities $p(Y \mid \theta_1, M_1)$ and $p(Y \mid \theta_2, M_2)$ are the corresponding likelihood functions, and $Y = (y_1', \ldots, y_T')'$ is the vector of data. To interpret the values of the Bayes factor, we use the widely acknowledged reference categories of Kass and Raftery (1995), with values from 1 to 3.2, from 3.2 to 10, from 10 to 100 and greater than 100 indicating virtually zero, substantial, strong and decisive evidence in favor of M_1 against M_2, respectively. Typically, M_1 is more general than M_2 and BF_{12} is greater than unity. It is, however, possible that the evidence still supports M_2 more than M_1, in which case the Bayes factor is less than unity, and the reference categories of Kass and Raftery (1995) should be applied to BF_{21}, the inverse of BF_{12} instead.

In addition to assessing whether there is sufficient heteroskedasticity for identification, the Bayes factor can be used to check for the validity of identifying external instruments. Recall from Subsection 3.3 that, with k instruments, the $(k \times (n-k))$ matrix Φ_2 in (2) must equal zero for the instruments to be exogenous, and the $(k \times k)$ matrix Φ_1 must deviate from zero for the instruments to be relevant. Hence, exogeneity can be assessed by imposing $\Phi_2 = 0$ in (2) and using (13) to compare the resulting model M_1 against the model M_2, where no restrictions are imposed on Φ_2. In the same vein, instrument relevance can be assessed by comparing the unrestricted model M_1 against M_2, where $\Phi_1 = 0$. Values of the Bayes factor greater than 3.2 (10), indicate substantial (strong) evidence in favor of exogoneity or relevance of the instruments. If evidence in favor of exogeneity is found, it may be a good idea to restrict Φ_2 equal to zero in M_1 when checking for instrument relevance, as this increases the probability of the Bayes factor being informative.

Above as well as in Subsection 3.3, we assumed that the k instruments are correlated with the first k elements of ε_t (if any). This poses no problem when sign restrictions are used to provide the shocks with economic labels, as in Braun and Brüggeman (2022), whereas the shocks are identified only up to ordering (and signs) by statistical properties of the data and have no economic interpretation a priori. Therefore, it is not known which of the statistically identified shocks are the likeliest to be correlated with each of the instruments. However, following the previous statistical identification literature, inspection of the impulse responses and forecast error variance decompositions can be used for labeling, and subsequently the shocks of interest can be ordered first before introducing the instruments. In addition, Bayes factors can be used to match the shocks with instruments such that the exogeneity and relevance conditions are likely to be satisfied. For instance, in the case of one instrument, the exogeneity and relevance of the instrument for each statistically identified shock can be assessed in turn, and the shock with the strongest evidence in favor of validity is deemed the shock of interest. Alternatively, the posterior distributions of the elements of Φ can be inspected, which is our approach in Section 5.

5 Empirical Application

We illustrate the methods by means of an empirical application related to the crude oil market. In the previous literature, different strategies have been used to identify the structural shocks in SVAR models of the oil market. However, typically the shocks have been assumed Gaussian, so without additional restrictions, they are only set identified, and to achieve point identification, additional information must be incorporated into the model. For instance, Kilian and Murhpy (2014) use sign restrictions on impulse responses, while Baumeister and Hamilton (2019) impose short-run restrictions on the impact matrix (or its inverse) by means of the prior distribution. Montiel Olea et al. (2021) use an external instrument as a proxy for the oil supply shock, and Braun and Brüggemann (2022) combine sign restrictions with a proxy variable in identification.

Instead of assuming the structural errors to be Gaussian, our approach is to use a highly flexible error distribution to capture most deviations from normality. This way, as shown by the identification results in Section 3, we can efficiently learn about the structural shocks of interest from data when these shocks exhibit non-Gaussianity. Specifically, we assume that each of the errors follows a skewed t distribution. Because it nests the Gaussian distribution (as a limiting case), checking for identification is straightforward, as discussed in Subsection 4.3. To avoid misspecification due to unmodeled heteroskedasticity, we assume that the conditional variance of each error term follows the GARCH(1,1) process in (8).

In addition to relying on non-Gaussianity for identification, we consider two ways of strengthening it. First, following Braun and Brüggemann (2022), we use Kilian's (2008) measure of exogenous oil supply shocks as an instrument for the oil supply shock. This proxy variable also helps in labelling the one of the statistically identified shocks as the supply shock. Second, following Baumeister and Hamilton (2019) and Braun and Brüggeman (2022), we consider imposing strongly informative priors on the oil supply elasticities. However, in terms of impulse responses, the impact of the proxy variable and tight priors turns out to be negligible.

Our SVAR(13) model comprises the same four variables as Kilian and Murphy's (2014) model: global oil production (prod_t), global real activity (ip_t), the real price of oil (rpo_t), and the seasonally adjusted OECD crude oil inventories (inv_t), i.e., $y_t = (\text{prod}_t, \text{ip}_t, \text{rpo}_t, \text{inv}_t)'$. We use the global industrial production index of Baumeister and Hamilton (2019) as a measure of ip_t. All these variables are expressed in logs, and the monthly series run from October 1978 to November 2018. Kilian's (2008) oil supply shock series covers the period from January 1973 to September 2004 only, and we use its extended version that spans from January 1973 to November 2018 in our analysis (see Braun and Brüggeman (2022)).[2]

5.1 Strength of Identification

Using Bayesian methods described in Section 4, we estimate the SVAR(13) model (1) augmented by univariate GARCH(1,1) processes (8) for the structural errors. Thirty chains, each consisting of four thousand draws, are generated using the NUTS algorithm. The first thousand draws are used for automatic tuning of the algorithm, so the posterior sample

[2]The same data were used by Braun and Brüggeman (2022), and we downloaded them from `https://www.tandfonline.com/doi/suppl/10.1080/07350015.2022.2104857`. For a detailed discussion on the variables, see Kilian (2008, 2009), Baumeister and Hamilton (2019), and Braun and Brüggeman (2022).

consists of ninety thousand draws that are only slightly autocorrelated. The convergence of these chains is assessed using the potential scale reduction factor (PSRF) introduced by Gelman et al. (2013). For each parameter of the model, the value of PSFR is below the threshold of 1.01 specified by Vehtari et al. (2021), indicating satisfactory convergence.

In view of Propositions 1 and 2, identification depends on the statistical properties of the structural errors, and its strength can be assessed by checking to what extent the assumptions underlying these results are satisfied. To that end, we depict in Figure 1 the 68% and 90% prior and posterior intervals of λ_i and $2q_i$ $(i = 1, \ldots, 4)$, the parameters controlling the skewness and kurtosis of the shocks, respectively. The posterior intervals of λ_i suggest that Shocks 1, 2 and 3 exhibit relatively strong negative skewness, even after accounting for conditional heteroskedasticity. As only one of the shocks seems symmetric, Proposition 1(ii) suggest that the full matrix B is point identified. In contrast, the posterior intervals of the degree-of-freedom parameter indicate only mild excess kurtosis, with very large values being highly likely especially for Shocks 2 and 4. Thus, Proposition 2 is not useful in establishing identification in this application.

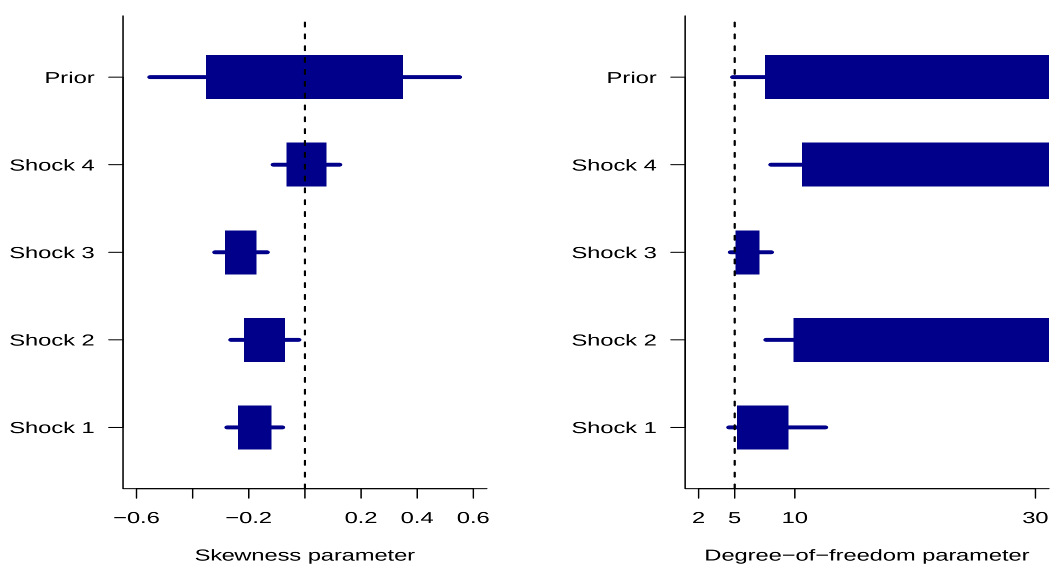

Figure 1: 68% and 90% prior and posterior intervals for the parameters λ_i and $2q_i$ $(i = 1, \ldots, 4)$ controlling for the skewness (left pane) and excess kurtosis (right pane; Degree-of-freedom parameter) of the skewed t-distributed structural shocks, respectively.

While all structural shocks turn out to be strongly identified via skewness, it is still interesting to check whether identification could also be achieved by conditional heteroskedasticity. As discussed in Section 3, if at least three of the four shocks follow autocorrelated volatility processes, the impact matrix B is point identified. Figure 2 depicts the posterior distribution of the conditional volatility of each shock implied by the respective GARCH(1,1) process over time. All four shocks seem to follow a highly persistent conditional volatility process. Hence, the impact matrix B seems globally identified via not only skewness but also conditional heteroskedasticity of the structural shocks.

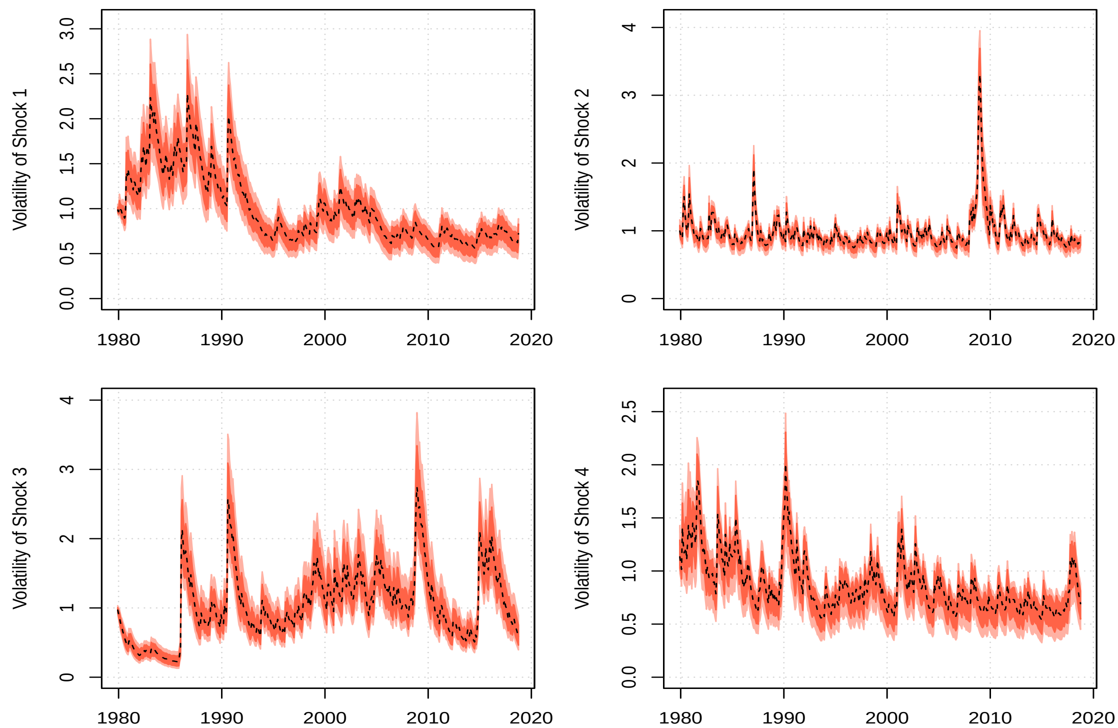

Figure 2: 68% and 90% point-wise credible sets and posterior medians (dashed lines) of the conditional shock volatilities starting from December 1979.

Apart from facilitating identification, capturing conditional heteroskedasticity in the structural errors is important from the viewpoint of avoiding misspecification. To examine conditional heteroskedasticity more closely, we compute the Bayes factor (BF) of the esimated model against its homoskedastic counterpart. The computed Bayes factor equals $e^{123.4}$, lending decisive support to the model with time-varying shock volatilities. Thus, it seems that the Gaussian and homoskedastic SVAR model in, Kilian and Murphy (2014), Baumeister and Hamilton (2019) and Braun and Brüggeman (2022), among others, may be seriously misspecified in addition to unnecessarily overlooking useful identifying information.

As already mentioned, identification of the oil supply shock can be strengthened using Kilian's (2008) measure of exogenous oil supply shocks as an external instrument. Because point identification is achieved by statistical properties of the data, instrument validity conditions in (5) and (6) can be assessed by the Bayes factor. Furthermore, because the shocks are identified only up to their ordering (and signs), but lack an economic interpretation, the instrument can be helpful in labeling the supply shock. Hence, using the same data as above, we estimate the augmented SVAR(13) model (2) with the exogenous oil supply shock as an external instrument. We reparametrize the error vector $\tilde{\varepsilon}_t = (\varepsilon_t', \eta_t')$ as $\Sigma_t^{1/2}\tilde{\epsilon}_t$ and let its each component $\tilde{\epsilon}_{it}$ $(i = 1, \ldots, n)$ follow a skewed t-distribution with mean zero and variance unity. In accordance with the model excluding the instrument, we assume that the diagonal elements of Σ_t follow univariate GARCH(1,1)-processes (8).

The marginal posterior distributions of the parameters λ_{n+1} and $2q_{n+1}$ controlling the skewness and excess kurtosis of the error term of the equation for the instrument, η_t, exhibit negative skewness and long tails, respectively.[3] In addition, strong persistent time-varying heteroskedasticity seems to be present. As expected, the marginal posteriors of λ_i and q_i for $i = 1, \ldots, n$ are very close to those reported in Figure 1, and thus we can conculude that all the parameters of (2) are point identified. Since also Φ, the vector containing the coefficients of the variables in the equation for the instrument (see (3)), is point identified, the validity of the instrument can be assessed. However, because the shocks have no labels, we first have to figure out for which shock the instrument is the likeliest proxy, i.e., which shock is the likeliest to be the oil supply shock. To that end, we report the posterior medians as well as the 16% and 84% quantiles of the elements of $\Phi = (\phi_1, \ldots, \phi_4)$ in Table 1. Among them, only the 68% posterior credible set of ϕ_1 does not contain zero, which suggests that the first shock is the oil supply shock.

In view of the finding that the instrument is likeliest to be a proxy for the first shock, its exogeneity can be assessed by imposing $\Phi_2 = (\phi_2, \phi_3, \phi_4) = 0$ in (2) and comparing the resulting model M_1 against the model M_2, where no restrictions are imposed on Φ_2. The value of the Bayes factor comparing these two models is 42.6, lending very strong support in

[3]The 10% (16%) and 90% (84%) quantiles of λ_{n+1} are –0.259 (–0.237) and 0.013 (–0.103), respectively, whereas those of q_{n+1} are 1.391 (1.422) and 1.755 (1.613), respectively.

Table 1: Posterior medians and 68% posterior credible sets of the elements of $\Phi = (\phi_1, \ldots, \phi_4)$.

Parameter	16%	50%	84%
ϕ_1	0.005	0.026	0.048
ϕ_2	–0.024	–0.010	0.005
ϕ_3	–0.005	0.007	0.019
ϕ_4	–0.038	–0.016	0.005

favor of instrument exogeneity. In the same vein, to assess instrument relevance, we compute the Bayes factor of the less constrained model M_1, where no restrictions are imposed on $\Phi_1 = \phi_1$, against the model M_2, where $\Phi_1 = 0$. We set $\Phi_2 = 0$ in both models to make the Bayes factor more informative. The reciprocal of the Bayes factor comparing the former to the latter model is 1.5699, which lends no support to instrument relevance. However, the results in Table 1 suggest that Kilian's (2008) measure of exogenous oil supply shocks may not be irrelevant but a relatively weak instrument. Also the fact that the posterior probability $P(\Phi_1 > 0 \mid y)$ is only 0.896 backs up this conclusion. Finally, as discussed in the following subsection, the impulse responses to the shocks remain virtually intact whether the instrument is used or not, which attests to this insight. Nevertheless, the instrument is useful in labeling the first shock as the oil supply shock.

5.2 Impulse Responses Analysis

The posterior medians and 68% and 90% credible sets of the impulse responses to the structural shocks up to 40 months are depicted in Figure 3. The shocks are scaled such that the ith shock causes a 1% change in the ith variable on impact. They are based on the model excluding Kilian's (2008) measure of exogenous oil supply shock; the results based on the augmented model are virtually identical. Compared to corresponding impulse responses in the previous literature, they have the advantage that due to versatile error distributions and explicit modeling of conditional heteroskedasticity, severe misspecification is probably avoided and identifying information in the data is efficiently utilized. However, additional information is needed to label the shocks and, hence, to facilitate the interpretation of the impulse responses. Kilian and Murhpy (2014) and Braun and Brüggeman (2022) considered three shocks, namely the flow supply shock, the flow demand shock, and the inventory (speculative) demand shock, and used the sign restrictions summarized in Table 2 for identification.

Table 2: The signs of the impact effects of oil supply, flow demand and inventory demand shocks on prod_t,ip_t, rpo_t and inv_t. A positive (negative) effect is denoted by + (−), while ∗ denotes an unrestricted effect.

	Variable			
	prod_t	ip_t	rpo_t	inv_t
Flow supply shock	−	−	+	∗
Flow demand shock	+	+	+	∗
Inventory demand shock	+	−	+	+

According to the results in the previous subsection, the additional instrument is only a weak proxy for the oil supply shock, but it still facilitates labeling the first shock as the flow supply shock. The impulse responses of this shock come close to satisfying the sign restrictions of the supply shock, and they also resemble those in Kilian and Murphy (2014) and are in line with standard economic intuition, so we label it the flow supply shock. The shock is associated with a sharp and persistent decrease in oil production, but it has no effect on the real activity. In addition, after the initial zero impact, a negative oil supply shock has a positive effect on the real price of oil and a negative effect on crude oil inventories. Both of these effects are highly persistent.

The impulse responses of Shock 2 come closest to satisfying the sign restrictions of the demand shock, and they also resemble the corresponding impulse responses in Kilian and Murphy (2014). Hence, we label Shock 2 the flow demand shock. It has a positive impact on the real activity and the real price of oil (after the initial effect indiscernible from zero). The former jumps to a higher level on impact, and the decay towards zero is very slow, while the real price of oil increases relatively quickly, reaching its new permanent level within ten months. Kilian (2009), Kilian and Murphy (2014), and Inoue and Kilian (2013) also report positive oil price and real activity responses to a flow demand shock. In contrast, Kilian and Murphy (2014) find almost no impact of the aggregate demand shock on inventories, while according to Figure 3 there is a relatively small negative effect.

As for the remaining two shocks (Shocks 3 and 4) in Figure 3, neither of them can be labeled as the inventory demand shock using the sign restrictions. Specifically, the posterior probability of each of these two shocks satisfying the sign restrictions is zero, as all

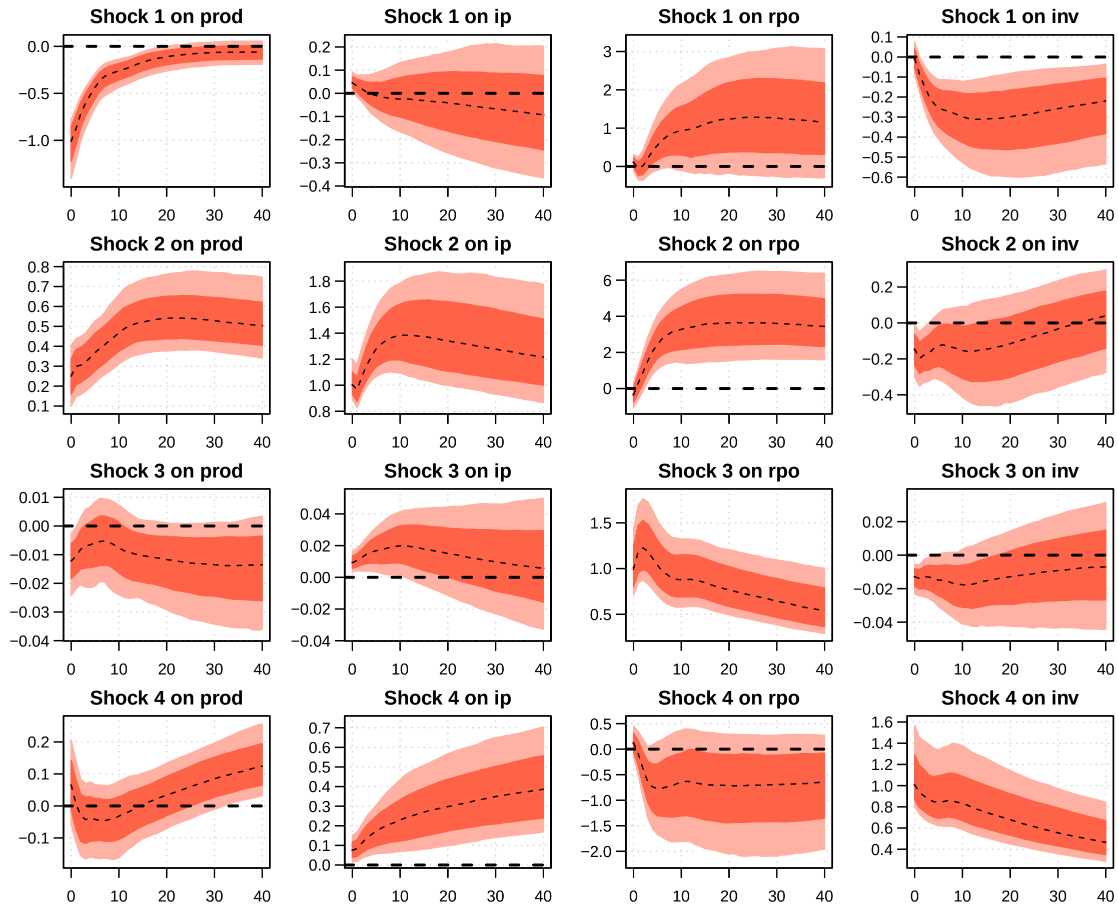

Figure 3: Impulse responses to the statistically identified shocks. The solid lines are the posterior medians, and the shaded red areas are the 68% and 90% point-wise posterior intervals. The effect of the ith shock is normalized such that it causes a 1% increase (or decrease) in the ith variable on impact.

the probability mass of the impact responses of real activity lies above zero in both cases. Recently, Cross et al. (2022) argued that instead of a composite inventory demand shock, precautionary demand and speculative demand shocks should be considered separately. They constructed an oil price uncertainty (OPU) index as a measure of uncertainty driving the former shock (see their Figure 1), and it very closely resembles the conditional volatility of Shock 3 implied by our model in Figure 2. Especially the peaks related to the Persian Gulf war (1990–1991), the Great Recession (2007–2008) and the oil price collapse in 2014 are clearly distinguishable in both series. Hence, assuming that conditional volatilities reflect uncertainty related to the shocks, we can label Shock 3 the precautionary demand shock.

Finally, we additionally consider the bounds on the price elasticities of oil supply used by Kilian and Murphy (2014) for identification. To that end, we impose very informative priors on the supply elasticities. In particular, as explained in Braun and Brüggemann (2022), $\eta_{13k} = B_{1k}/B_{3k}$ can be thought as an oil supply elasticity, measuring the percentage increase of oil production (variable 1) in response to a one percentage increase in the real price of oil (variable 3) caused by a positive demand shock. Following their lead, we assume the Student t prior of Baumeister and Hamilton (2019) with mode 0.1, scale parameter 0.2, and 3 degrees of freedom (and truncated to be positive) for both η_{132} and η_{134} (a detailed description of the implementation of these priors is provided in Appendix D). These priors result in negative impulse responses (with narrower bands than those depicted in Figure 3) of real activity to the supply shock, but otherwise their effects on the impulse responses are minor.[4]

6 Conclusion

In this paper, we consider identification in SVAR models using non-Gaussian features of the structural shocks, including skewness and excess kurtosis. In contrast to most of the previous related literature, we relax the assumption in independent shocks and only require them to be mutually orthogonal. The latter assumption is common in the SVAR literature, and the independence assumption has been criticized on the grounds that independent structural shocks may not always be obtained as linear transformation of the reduced-form residuals. Moreover, the structural shocks are often likely to share a common source of volatility, which is precluded when the shocks are independent.

We show full identification (up to ordering and signs of the structural shocks) when at most one of the shocks is symmetric or at least $n-1$ of the shocks in an n-dimensional SVAR model are leptokurtic. Moreover, if there are fewer than $n-1$ skewed shocks, they are identified, while the remaining shocks are only set identified. A similar partial identification result applies to the leptokurtic shocks. The partial identification results pertaining to the skewed and leptokurtic shocks can also be combined such that some of them can be identified because they are skewed, and some because they are leptokurtic, while the symmetric and mesokurtic shocks are not identified. Finally, we show how identification of the SVAR model by statistical properties of the shocks facilitates checking the validity of exogneous

[4]These impulse responses are not shown to save space but are available upon request.

instruments used to strengthen identification.

Capturing the non-Gaussian features of the data calls for versatile error distributions. In addition, to strengthen identification and to avoid misspecification, it is important to model potential conditional volatility of the errors. To that end, we explicitly consider skewed t-distributed errors following GARCH(1,1) processes. Although not necessary, Bayesian methods are recommended because of the complexity of the model, and we discuss their implementation in detail in this model that is likely to be widely applicable in econometrics. Among other things, they have the advantage that they facilitate checking for identification in a straightforward manner by inspecting the posterior distributions of the shocks.

In an empirical application to the world oil market, we demonstrate how the model is estimated and how identification is checked. The statistically identified shocks do not carry any economic interpretation as such, and we demonstrate how they can be labeled by making use of sign restrictions introduced in the previous literature, conditional volatilities of the shocks and an exogenous instrument. We also show how the validity of the instrument can be checked.

References

Anttonen, J., Lanne, M., Luoto, J. (2023), "Bayesian Inference on Fully and Partially Identified Structural Vector Autoregressions," Available at SSRN: https://ssrn.com/abstract=4358059.

Arias, J., Rubio-Ramírez, J. F., Waggoner, D. F. (2021), "Inference in Bayesian Proxy-SVARs," *Journal of Econometrics*, 225, 88–106.

Baumeister, C., and Hamilton, J. D. (2019). "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks, "*American Economic Review*, 109, 1873–1910.

Bertsche, D., and Braun, R. (2022). "Identification of structural vector autoregressions by stochastic volatility, "*Journal of Business and Economic Statistics*, 40, 328–341.

Bonhomme, S. and, Robin, J.-M. (2009), "Consistent noisy independent component analysis "*Journal of Econometrics*, 149, 12–25.

Braun, R., and Brüggemann, R. (2023), "Identification of SVAR Models by Combining Sign Restrictions With External Instruments," *Journal of Business and Economic Statistics*, forthcoming.

Carriero, A., Clark, T. E., and Marcellino, M. (2016), "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business and Economic Statistics*, 34(3), 375–390.

Chan, J., and Grant, A. (2017), "On the Observed-Data Deviance Information Criterion for Volatility Modeling," *Journal of Financial Econometrics*, 14, 772–802.

Cross, J. L., Nguyen, B. H., and Tran, T. C. (2022), "The Role of Precautionary and Speculative Demand in the Global Market for Crude Oil," *Journal of Applied Econometrics*, 37, 841–1090.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017), "Statistical Inference for Independent Component Analysis: Application to Sructural VAR Models," *Journal of Econometrics*, 196, 111–126.

Guay, A. (2021), "Identification of Structural Vector Autoregressions through Higher Unconditional Moments," *Journal of Econometrics*, 225, 27–46.

Herwartz, H. (2018), "Hodges Lehmann Detection of Structural Shocks - An Analysis of Macroeconomic Dynamics in the Euro Area," *Oxford Bulletin of Economics and Statistics*, 80, 736–754.

Inoue, A. and Kilian, L. (2013), "Inference on impulse response functions in structural VAR models," *Journal of Econometrics*, 177, 1–13.

Kass, R. E., Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Keweloh, S. A. (2021), "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments," *Journal of Business and Economic Statistics*, 39, 772–782. forthcoming.

Kilian, L. (2008), "Exogenous Oil Supply Shocks: How Big are they and how much do they Matter for the U.S. Economy?" *The Review of Economics and Statistics*, 90, 216–240.

Kilian, L. (2009), "Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–1069.

Kilian, L., and Murphy, D. P. (2014), "The Role Of Inventories And Speculative Trading In The Global Market For Crude Oil," *Journal of Applied Econometrics*, 29, 454–478.

Lanne, M., Liu, K., and Luoto, J. (in press), "Identifying Structural Vector Autoregression via Leptokurtic Economic Shocks," *Journal of Business and Economic Statistics*.

Lanne, M., and Luoto, J. (2021), "GMM Estimation of Non-Gaussian Structural Vector Autoregression," *Journal of Business and Economic Statistics*, 39, 69–81.

Lanne, M., Meitz, M., and Saikkonen, P. (2017), "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions," *Journal of Econometrics*, 196, 288–304.

Lanne, M., and Saikkonen, P. (2007), "A multivariate generalized orthogonal factor GARCH model," *Journal of Business and Economic Statistics*, 25(1), 61–75.

Lewis, D. J. (2021), "Identifying Shocks via Time-Varying Volatility," *The Review of Economic Studies*, 88(6), 3086–3124.

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

Mencia, J., and Sentana, E. (2006), "Estimation and testing of dynamic models with generalised hyperbolic innovations," *CEPR* Discussion Paper Series No. 5177.

Mertens, K., Ravn, M. O. (2012), "Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks," *American Economic Journal-Economic Policy*, 4, 145–181.

Mesters, G., and Zwiernik, P. (2022), "Non-independent Components Analysis," arXiv:2206.13668v2 [math.ST]

Montiel Olea, J. L., Plagborg-Møller, M., and Qian, E. (2022), "SVAR Identification From Higher Moments: Has the Simultaneous Causality Problem Been Solved?," *AEA Papers and Proceedings*, 112, 481–485.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140(2), 425–449.

Schlaak, T., Rieth, M., and Podstawski, M. (2023), "Monetary policy, external instruments, and heteroskedasticity," *Quantitative Economics*, 14, 161–200.

Stock, J. H., Watson, M. W. (2012), "Disentangling the channels of the 2007–09 recession," *Brookings Papers on Economic Activity*, 43, 81–156.

Velasco, C. (2023), "Identification and estimation of structucal VARMA models using higher order dynamics," *Journal Business and Economic Statistics*, 41, 819–832.

Yang, M. (2008), "Normal log-normal mixture, leptokurtosis and skewness," *Applied Economics Letters*, 15(9), 737–742.

Yu, J. (2005), "On leverage in a stochastic volatility model," *Journal of Econometrics*, 127(2), 165–178.

Zhou, X. (2020), "Refining the Workhorse Oil Market Model," *Journal of Applied Econometrics*, 35, 130–140.

Appendix A Proof of Proposition 1

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 1 and r 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 $\mathrm{E}[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}] \equiv \Gamma_{ijk}$. By $\varepsilon_t^* = Q^{-1}\varepsilon_t$, it can be expressed as

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

where $\Gamma_{ijk}^* \equiv \mathrm{E}[\varepsilon_{i,t}^*\varepsilon_{j,t}^*\varepsilon_{k,t}^*]$, $\varepsilon_{i,t}^*$ is the ith, $i = 1, \dots, n$, element of ε_t^*, and Q_{ij} is the (i, j)-element, $i, j = 1, \dots, n$, of Q. Assumption 1(*iii*) implies that $\mathrm{E}[\varepsilon_{it}^*\varepsilon_{jt}^*\varepsilon_{kt}^*] = \mathrm{E}[\varepsilon_{it}^{*3}]$ when $i = j = k$ and zero otherwise. Therefore, (A.1) above, can be written as

$$\begin{aligned}
\Gamma_{ijk} &= \sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n} Q_{ip}Q_{jq}Q_{kr}\Gamma_{pqr}^* \\
&= \sum_{p=1}^{n} Q_{ip}Q_{jp}Q_{kp}\Gamma_{ppp}^*. \qquad \text{(A.2)}
\end{aligned}$$

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

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \Gamma_{ijk}^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n} Q_{ip}Q_{jq}Q_{kr}\Gamma_{pqr}^*\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n}\sum_{x=1}^{n}\sum_{y=1}^{n}\sum_{z=1}^{n} Q_{ip}Q_{jq}Q_{kr}Q_{ix}Q_{jy}Q_{kz}\Gamma_{pqr}^*\Gamma_{xyz}^* \\
&= \sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n}\sum_{x=1}^{n}\sum_{y=1}^{n}\sum_{z=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} Q_{ip}Q_{jq}Q_{kr}Q_{ix}Q_{jy}Q_{kz}\Gamma_{pqr}^*\Gamma_{xyz}^*, \qquad \text{(A.3)}
\end{aligned}$$

where the first equality follows from (A.1). By the orthogonality of Q, we have

$$\sum_{i=1}^{n} Q_{ip} Q_{iq} = \delta_{pq} = \begin{cases} 1, & p = q \\ 0, & p \neq q \end{cases} \tag{A.4}$$

Using (A.4) above, (A.3) simplifies to

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \Gamma_{ijk}^2 &= \sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n}\sum_{x=1}^{n}\sum_{y=1}^{n}\sum_{z=1}^{n} \delta_{px}\delta_{qy}\delta_{rz}\Gamma_{pqr}^{*}\Gamma_{xyz}^{*} \\ &= \sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n} \Gamma_{pqr}^{*2} \end{aligned} \tag{A.5}$$

Recall that based on $1(iii)$, $\Gamma_{ijk} \equiv \mathrm{E}[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}] = \mathrm{E}[\varepsilon_{it}^3]$ when $i = j = k$ and zero otherwise. Therefore, (A.5) reduces to

$$\sum_{i=1}^{n} \Gamma_{iii}^2 = \sum_{i=1}^{n} \Gamma_{iii}^{*2} \tag{A.6}$$

Using (A.2), (A.6) can be written as

$$\sum_{i=1}^{n} \left(\sum_{p=1}^{n} Q_{ip}^3 \Gamma_{ppp}^{*} \right)^2 = \sum_{i=1}^{n} \Gamma_{iii}^{*2}, \tag{A.7}$$

or equivalently,

$$\sum_{i=1}^{n} \left(\sum_{p=1}^{n} Q_{ip}^3 \, \mathrm{E}[\varepsilon_{p,t}^{*3}] \right)^2 = \sum_{i=1}^{n} \mathrm{E}[\varepsilon_{i,t}^{*3}]^2, \tag{A.8}$$

From Lemma 15 of Comon (1994), we obtain

$$\sum_{i=1}^{n} \left(\sum_{p=1}^{n} Q_{ip}^2 \left| \mathrm{E}[\varepsilon_{p,t}^{*3}] \right| \right)^2 \leq \sum_{i=1}^{n} \mathrm{E}[\varepsilon_{i,t}^{*3}]^2. \tag{A.9}$$

Combining (A.8) and (A.9), we have

$$\sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{2}\left|\mathrm{E}[\varepsilon_{p,t}^{*3}]\right|\right)^{2} \leq \sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{3}\,\mathrm{E}[\varepsilon_{p,t}^{*3}]\right)^{2}, \tag{A.10}$$

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

$$\sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{2}\left|\mathrm{E}[\varepsilon_{p,t}^{*3}]\right|\right)^{2} \geq \sum_{i=1}^{n}\left(\sum_{p=1}^{n} \left|Q_{ip}^{3}\right|\left|\mathrm{E}[\varepsilon_{p,t}^{*3}]\right|\right)^{2} \geq \sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{3}\,\mathrm{E}[\varepsilon_{p,t}^{*3}]\right)^{2}. \tag{A.11}$$

By (A.10) and (A.11), it must thus be that

$$\sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{2}\left|\mathrm{E}[\varepsilon_{p,t}^{*3}]\right|\right)^{2} = \sum_{i=1}^{n}\left(\sum_{p=1}^{n} \left|Q_{ip}^{3}\right|\left|\mathrm{E}[\varepsilon_{p,t}^{*3}]\right|\right)^{2} = \sum_{i=1}^{n}\left(\sum_{p=1}^{n} Q_{ip}^{3}\,\mathrm{E}[\varepsilon_{p,t}^{*3}]\right)^{2}, \tag{A.12}$$

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

$$Q_{ij}^{2}(|Q_{ij}|-1)|E(\varepsilon_{j,t}^{*3})| = 0. \quad i, j = 1, \ldots, n \tag{A.13}$$

Now, according to Assumption 1(iv) at least one structural error has nonzero skewness. Suppose $E\left(\varepsilon_{j,t}^{*3}\right) \neq 0$. Then, by (A.13), Q_{ij} must be either zero or ± 1 for all $i = 1, \ldots, n$. By the orthogonality of Q, it hence follows that the jth 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 kth, multiplied by ± 1. By $B^* = BQ$, (A.13) also means that the jth column of B^* is equal to the kth column of B corresponding to the kth structural error ε_{kt}. Obviously, if the number of the skewed structural errors $r > 1$, then (A.13) and the orthogonality of Q ensure that each of the r 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, r elements of ε_t^* are equal to the r skewed structural errors in ε_t, and also r columns of B^* are equal to the r columns of B

corresponding to these r skewed structural errors in ε_t.

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

$$Q = \begin{pmatrix} P & 0 \\ 0 & D \end{pmatrix} \tag{A.14}$$

with some $(r \times r)$ signed permutation matrix P, and an $((n-r) \times (n-r))$ orthogonal matrix D (notice that $QQ' = I_n$ together with (A.14), implies that $DD' = I_{n-r}$). This means that 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.14) into $B^* = BQ$, we immediately see that

$$B_1^* = B_1 P, \tag{A.15}$$

$$B_2^* = B_2 D, \tag{A.16}$$

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. 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.13), Q_{ij} must be either zero or ± 1 for all $i, j = 1, \ldots, n$, $j \neq l$, and therefore, by the orthogonality of Q, we know that each column of Q except the lth 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 $j = 1, \ldots, n$, $j \neq l$, 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.13), Q must be a signed permutation matrix, so B is identified by sign reversals and ordering of its columns.

Appendix B Proof of Proposition 2

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 the first s $(0 < s < n-1)$ components of ε_t^* are either leptokurtic or platykurtic, and the last $n-s$ $(0 < s < n-1)$ components have zero excess kurtosis. We partition ε_t^* accordingly: $\varepsilon_t^* = (\varepsilon_t^{*1\prime}, \varepsilon_t^{*2\prime})'$ with ε_t^{*1} an $(s \times 1)$ and ε_t^{*2} an $((n-s) \times 1)$ vector. We also partition the orthogonal matrix Q as

$$Q = \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix}, \tag{B.1}$$

where Q_1, Q_2, Q_3 and Q_4 are $(s \times s)$, $(s \times (n-s))$, $((n-s) \times s))$ and $((n-s) \times (n-s))$ matrices.

The results in Appendix F of Lanne, Liu, and Luoto (2021), imply that under Assumption 2, we have

$$E(\varepsilon_{it}^2 \varepsilon_{jt}^2) - 1 = \sum_{k=1}^{n} Q_{ik}^2 Q_{jk}^2 \Gamma_k^* = 0, \quad i \neq j \tag{B.2}$$

where we denote by $\Gamma_i^* = E(\varepsilon_{it}^{*4}) - 3$ the excess kurtosis of the structural errors ε_{it}^*, $i = 1, \ldots, n$, and Q_{ij} are the (i,j)-elements, $i, j = 1, \ldots, n$, of Q in $\varepsilon_t^* = Q^{-1}\varepsilon_t$. Now, according to Assumption 2(iv), the components of ε_t^{*1} are all either leptokurtic or platykurtic, whereas the components of ε_t^{*2} have zero excess kurtosis. Thus, by (B.2), it must be that

$$Q_{ik}^2 Q_{jk}^2 = 0. \quad i, j = 1, \ldots, n, i \neq j, k = 1, \ldots, s \tag{B.3}$$

This means that each column of $(Q_1', Q_3')'$ has at most one nonzero element. And, by the orthogonality of Q, each column of $(Q_1', Q_3')'$ has exactly one nonzero element equal to ± 1.

Thus, the first s columns $(Q_1', Q_3')'$ of Q can be expressed as

$$\begin{pmatrix} Q_1 \\ Q_3 \end{pmatrix} = P \begin{pmatrix} I_s \\ 0_{n-s,s} \end{pmatrix}, \tag{B.4}$$

where P is an $(n \times n)$ signed permutation matrix. By left multiplying (B.1) by P^{-1}, and using (B.4) above, we obtain

$$P^{-1}Q = O = \begin{pmatrix} I_s & O_2 \\ 0_{n-s,s} & O_4 \end{pmatrix}, \tag{B.5}$$

where we let

$$P^{-1}\begin{pmatrix} Q_2 \\ Q_4 \end{pmatrix} = \begin{pmatrix} O_2 \\ O_4 \end{pmatrix}. \tag{B.6}$$

By the fact that a signed permutation transformation conserves orthogonality, O must be orthogonal. By the orthogonality of O, we obtain

$$\begin{aligned} I_n &= OO' \\ &= \begin{pmatrix} I_s & O_2 \\ 0_{n-s,s} & O_4 \end{pmatrix} \begin{pmatrix} I_s & 0'_{n-s,s} \\ O'_2 & O'_4 \end{pmatrix} \\ &= \begin{pmatrix} I_s + O_2 O'_2 & O_2 O'_4 \\ O_4 O'_2 & O_4 O'_4 \end{pmatrix}, \end{aligned} \tag{B.7}$$

Thus, O_4 is an orthogonal matrix: $O_4 O'_4 = I_{n-s}$. By the orthogonality of O_4, it follows that O_4 is of full rank, and, hence, the conditions $O_4 O'_2 = 0_{n-s,s}$ and $O_2 O'_4 = 0_{s,n-s}$, provided by (B.7) above, hold if and only if $O_2 = 0_{s,n-s}$.

Based on these results, (B.5) can be rewritten as

$$Q = P \begin{pmatrix} I_s & 0_{s,n-s} \\ 0_{n-s,s} & O_4 \end{pmatrix}. \tag{B.8}$$

From $\varepsilon_t^* = Q^{-1}\varepsilon_t$, we hence obtain

$$\varepsilon_t = P \begin{pmatrix} \varepsilon_t^{*1} \\ O_4 \varepsilon_t^{*2} \end{pmatrix}. \tag{B.9}$$

Because ε_t^{*1} contains errors with nonzero excess kurtosis, the leptokurtic (or platykurtic) structural errors in ε_t are point identified, but their order (and signs) are unknown. The remaining $n-s$ structural errors in ε_t have zero excess kurtosis, since ε_t^{*2} contains mesokurtic errors, and they are only set identified, as O_4 is an orthogonal matrix.

Similarly, let us partition B^* as $B^* = [B_1^*, B_2^*]$ with B_1^* and B_2^* $(n \times s)$ and $(n \times (n-s))$ matrices, respectively. Then, by substituting (B.8) into $B^* = BQ$, we immediately see that

$$B = \begin{pmatrix} B_1^* & B_2^* O_4^{-1} \end{pmatrix} P^{-1}. \tag{B.10}$$

The fact that $P^{-1} = P'$, a signed permutation matrix, implies that the s columns of B corresponding to the leptokurtic (platykurtic) structural errors, are point identified, but their order and signs are unknown. And the remaining $n-s$ columns of B are set identified, as O_4^{-1} is an orthogonal matrix.

Finally, it can be readily checked that these results hold for any ordering of the components of ε_t^*. To see this, notice that by (B.3), there are s columns each of which has at most one nonzero element, despite their order. By the orthogonality of Q, in turn, these nonzero elements on each column are equal to ± 1, and, for the same reason, this ± 1 is also the only nonzero element in the corresponding row of Q. Thus, (B.8) can also be obtained using (B.3) and the orthogonality of Q.

To prove part (ii), suppose first that only one component of ε_t, say, ε_{lt} has zero excess kurtosis (i.e., $\Gamma_l^* = 0$, $1 \leq l \leq n$). Then, because the remaining components of ε_t are all either leptokurtic or platykurtic, by (B.2), $Q_{ik}^2 Q_{jk}^2 = 0$ for all $i, j, k = 1, \ldots, n$, $i \neq j$, $k \neq l$, and therefore we know that each column of Q except the lth has at most one nonzero element. By the orthogonality of Q, it thus follows that Q_k, the kth column of Q, $(k = 1, \ldots, n$, $k \neq l)$ 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 $Q_k' Q_l = 0$ for $k = 1, \ldots, n$, $k \neq l$, 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, $Q = P$, a signed permutation matrix. Obviously, if all components of ε_t have nonzero skewness, by the orthogonality of Q and (B.2), Q must be a signed permutation matrix, so B is identified by sign reversals and ordering of its columns.

Appendix C Proof that $\tilde{Q} = \textbf{diag}(\tilde{Q}_1, \tilde{Q}_4)$

Consider the SVAR process in (2) and an observationally equivalent SVAR process defined by $\tilde{B}^* = \tilde{B}\tilde{Q}$ and $\tilde{\varepsilon}_t^* = \tilde{Q}^{-1}\tilde{\varepsilon}_t$ with $\tilde{Q}$ an $((n+k) \times (n+k))$ orthogonal matrix, where $\tilde{B}^* = \tilde{B}\tilde{Q}$ has the same structure as $\tilde{B}$ (that is, $\tilde{B}^*$ is a lower triangular matrix).

From (3), we obtain

$$\begin{aligned}
\tilde{B}^* &= \tilde{B}\tilde{Q} \\
&= \begin{pmatrix} B & 0_{n,k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix} \begin{pmatrix} \tilde{Q}_1 & \tilde{Q}_2 \\ \tilde{Q}_3 & \tilde{Q}_4 \end{pmatrix} \\
&= \begin{pmatrix} B\tilde{Q}_1 & B\tilde{Q}_2 \\ \Phi\tilde{Q}_1 + \tilde{Q}_3\Sigma_\eta^{1/2} & \Phi\tilde{Q}_2 + \tilde{Q}_4\Sigma_\eta^{1/2} \end{pmatrix}.
\end{aligned} \tag{C.1}$$

Because $\tilde{B}^*$ is a lower triangular matrix $B\tilde{Q}_2 = 0_{n,k}$. Based on the fact that B is of full rank, $B\tilde{Q}_2 = 0_{n,k}$ holds if and only if $\tilde{Q}_2 = 0_{n,k}$.

By the orthogonality of $\tilde{Q}$, we have

$$\begin{aligned}
I_{n+k} &= \tilde{Q}\tilde{Q}' \\
&= \begin{pmatrix} \tilde{Q}_1 & 0'_{n,k} \\ \tilde{Q}_3 & \tilde{Q}_4 \end{pmatrix} \begin{pmatrix} \tilde{Q}'_1 & \tilde{Q}'_3 \\ 0_{k,n} & \tilde{Q}'_4 \end{pmatrix} \\
&= \begin{pmatrix} \tilde{Q}_1\tilde{Q}'_1 & \tilde{Q}_1\tilde{Q}'_3 \\ \tilde{Q}_3\tilde{Q}'_1 & \tilde{Q}_3\tilde{Q}'_3 + \tilde{Q}_2\tilde{Q}'_2 \end{pmatrix}.
\end{aligned} \tag{C.2}$$

Therefore, $\tilde{Q}_1$ is an orthogonal matrix: $\tilde{Q}_1\tilde{Q}'_1 = I_n$. By the orthogonality of $\tilde{Q}_1$, it follows that $\tilde{Q}_1$ is of full rank, and, hence, the conditions $\tilde{Q}_1\tilde{Q}'_3 = 0_{k,n}$ and $\tilde{Q}_3\tilde{Q}'_1 = 0_{n,k}$ hold if and only if $\tilde{Q}_3 = 0_{k,n}$. Thus $\tilde{Q} = \text{diag}(\tilde{Q}_1, \tilde{Q}_4)$ as claimed in Section 3.3.

Appendix D Priors on Normalized Structural Shocks

Suppose the researcher is interested in the effect on variable i of a shock that increases the value of jth variable by one unit, and they want to impose a prior on the instantaneous response of the ith variable to this unit shock. Let $p_\eta(B_{jk}, \eta_{ijk})$ be the joint prior density of B_{jk} and the quantity of interest $\eta_{ijk} = B_{ik}/B_{jk}$ $(i, j, k = 1, \ldots, n)$. The joint prior density of (B_{ik}, B_{jk}), $p_B(B_{ik}, B_{jk})$, can be deduced from $p_\eta(B_{jk}, \eta_{ijk})$ using the change-of-variables formula:

$$p_B(B_{ik}, B_{jk}) = |J|\, p_\eta(B_{jk}, B_{ik}/B_{jk})$$

where the Jacobian matrix J is given by

$$J = \begin{bmatrix} \frac{\partial(B_{ik}/B_{jk})}{\partial B_{ik}} & \frac{\partial(B_{ik}/B_{jk})}{\partial B_{jk}} \\ \frac{\partial B_{jk}}{\partial B_{ik}} & \frac{\partial B_{jk}}{\partial B_{jk}} \end{bmatrix} = \begin{bmatrix} B_{jk}^{-1} & -B_{ik}B_{jk}^{-2} \\ 0 & 1 \end{bmatrix}.$$

Assuming that η_{ijk} and B_{jk} are a priori independent, we obtain

$$p_B(B_{ik}, B_{jk}) = |J|\, p_\eta(B_{ik}/B_{jk}) p_B(B_{jk}) \tag{D.1}$$

where $p_\eta(\eta_{ijk})$ and $p_B(B_{jk})$ are the marginal prior densities of η_{ijk} and B_{jk}, respectively.

For instance, suppose that η_{ijk} is the oil supply elasticity measuring the percentage increase of oil production (variable i) in response to a one percentage increase in the real price of oil (variable j) caused by a positive demand shock (shock k) (see Braun and Brüggeman (2022)). Then, we can follow Baumeister and Hamilton (2015) in using a truncated Student t density for η_{ijk}. In particular,

$$p_\eta(\eta_{ijk}) = \frac{\Gamma\left(\frac{\nu_{ijk}+1}{2}\right)}{(1 - F(0))(\nu_{ijk}\pi)^{1/2}\Gamma\left(\frac{\nu_{ijk}}{2}\right)} \left(1 + \frac{(\eta_{ijk} - c_{\eta,ijk})^2}{\sigma^2_{\eta,ijk}\nu_{ijk}}\right)^{-(\nu_{ijk}+1)/2}$$

if $\eta_{ijk} > 0$ and zero otherwise, where $c_{\eta,ijk}$, ν_{ijk}, and $\sigma_{\eta,ijk}$ are prior hyper-parameters, and $F(0)$ is the shorthand notation for the cumulative distribution $F(0; c_{ijk}, \sigma_{\eta,ijk}, \nu_{ijk})$. Thus, plugging in the expression above into (D.1), we have

$$p_B(B_{ik}, B_{jk}) = \left|B_{jk}^{-1}\right| \frac{\Gamma\left(\frac{\nu_{ijk}+1}{2}\right)}{(1 - F(0))(\nu_{ijk}\pi)^{1/2}\Gamma\left(\frac{\nu_{ijk}}{2}\right)} \left(1 + \frac{(B_{ik}/B_{jk} - c_{\eta,ijk})^2}{\sigma^2_{\eta,ijk}\nu_{ijk}}\right)^{-(\nu_{ijk}+1)/2} p_B(B_{jk})$$

if $B_{ik}/B_{jk} > 0$ and zero otherwise.