

A Non-Gaussian GDP Anatomy

Andrzej Kociecki* Christian Matthes† Michele Piffer‡

January 27, 2025

Abstract

Which shocks are the main contributors to the volatility of U.S. real GDP? We study this question using a vector autoregressive (VAR) model identified via non-Gaussianity. We first show that a simple reparameterization of a VAR with t-distributed shocks allows for the development of the first Gibbs sampler for this model, allowing us to increase the size of such models. Applying our framework to U.S. data reveals that no single shock dominates GDP. Instead, two shocks drive GDP fluctuations: a weakly inflationary demand shock at short horizons and a supply shock at longer horizons. No single shock explains substantially more than 20 percent of fluctuations at any horizon.

JEL classification: C32, E52.

Keywords: Non-Gaussianity, Shocks, identification, Bayesian inference, Gibbs sampler.

*University of Warsaw, Warsaw, Poland. e-mail: akociecki@wne.uw.edu.pl

†Indiana University, Bloomington, USA. e-mail: matthesc@iu.edu

‡Bank of England and King's Business School, King's College London, UK. e-mail: m.b.piffer@gmail.com. We would like to thank seminar and conference participants at the Barcelona Summer Forum, the Cleveland Fed, the Bank of England, the DNB, University of Lancaster, and the Milan Time Series Seminars. The views expressed in this paper do not represent the views of the Bank of England.

1 Introduction

Since at least [Tinbergen \(1939\)](#), economists have recognized that economic outcomes depend on both endogenous responses to economic conditions and unexpected changes – or shocks – that we want to interpret. A natural question then arises: Which shocks account for the bulk of business cycle fluctuations? In a key contribution, [Angeletos et al. \(2020\)](#) provide a contemporary answer. They use structural vector autoregressions (SVARs) to identify a shock that maximizes fluctuations in key measures of U.S. real activity. Specifically, they use a variant of the “max-share” approach of [Uhlig \(2004\)](#), which identifies a shock that maximizes the share of the forecast error variance for a given variable. However, this identification scheme delivers an estimated shock series that can be a linear combination of multiple underlying structural shocks. [Wolf \(2020\)](#) refers to this phenomenon as the “masquerading problem” in sign-restricted VAR settings, and [Francis and Kindberg-Hanlon \(2022\)](#) show that this issue can also arise with the max-share identification scheme.¹

To offer a different perspective on what drives business cycles and GDP more broadly, we use the same observables as [Angeletos et al. \(2020\)](#) but adopt alternative identification assumptions. In particular, we assume that US economic data are non-Gaussian – specifically, that structural shocks follow independent t-distributions – so that all SVAR parameters, including the shocks, are automatically identified. This result follows from [Comon \(1994\)](#), who shows that if at most one shock is Gaussian, the SVAR is identified up to sign and column permutations of the impact matrix, a normalization we solve in this paper.² After identifying these shocks, we compute variance decompositions at various horizons to determine which shocks explain most of the variance of real GDP at business cycle frequencies. Importantly, while our assumption of non-Gaussianity serves as an identification tool, it is also testable: we can check ex-post whether a Gaussian model would fit the data better. The estimated degrees of freedom of the t-distributed shocks measure the strength of identification – higher degrees of freedom mean the data are closer to Gaussian, leaving less information to exploit. We use a Bayesian approach, so weaker identification (i.e. larger degrees of freedom) naturally leads to an increase in the width of our posterior intervals.³

¹[Angeletos et al. \(2020\)](#) acknowledge this and refer to their identified shock as a “reduced-form” shock. Unless one adopts strong assumptions about which shocks truly drive the data, this leaves the door open to the masquerading problem.

²If only a subset of structural shocks is non-Gaussian, those shocks remain identified ([Maxand, 2020](#)).

³For a frequentist approach that provides correct coverage when the deviations from Gaussianity

Applying this method to the same data as [Angeletos et al. \(2020\)](#) reveals a shock whose impulse responses resemble their “main business cycle shock,” yet it does not account for the majority of GDP fluctuations. Nevertheless, among the shocks we identify, it remains one of the two most influential shocks for GDP, explaining 15-20 percent of GDP variance across horizons. The other key shock, especially relevant at longer horizons, moves total factor productivity and resembles a supply shock, consistent with traditional real-business-cycle (RBC) models ([Kydland and Prescott, 1982](#)).

The existing literature that uses non-Gaussianity for identification in VARs (e.g. [Brunnermeier et al., 2021](#)) has faced technical challenges that have limited the scope of applications. In particular, Bayesian inference has relied on the Metropolis-Hastings algorithm to estimate the matrix capturing how variables respond to the structural shocks, which is inefficient at exploring large parameter spaces and cannot handle the larger VARs commonly used in Gaussian settings.⁴ Furthermore, while non-Gaussianity removes any meaningful economic identification issues, the aforementioned impact matrix is only identified up to scale and sign. If not handled properly, lack of normalization can produce misleading posterior uncertainty estimates ([Hamilton et al., 2007](#)). Existing solutions to this problem are generally only applicable in models with relatively few observables. Relative to this existing literature, the technical contributions of our paper are (i) to develop a convenient sampling procedure for a non-Gaussian SVAR model in which the structural shocks are assumed to be t -distributed, and (ii) to provide a new normalization algorithm to tackle the sign and column switches of the impact matrix of the structural shocks, extending the likelihood-preserving normalization of [Waggoner and Zha \(2003b\)](#) and deriving a new, efficient algorithm to implement this normalization. Notably, our Gibbs sampler and normalization algorithm work for large-scale VARs, which have grown in popularity ([Baíbura et al., 2010](#)).

Student t distributions have attracted substantial interest in the literature, not only because they provide a natural framework to identify structural shocks, but also thanks to their ability to model fat tails, compared to Gaussian models ([Cúrdia et al., 2014](#), [Chiu et al., 2017](#)). Previous important contributions that exploit t -distributed shocks for identification in SVARs include [Lanne and Luoto, 2020](#), and [Anttonen](#)

are small, see [Hoesch et al. \(2024\)](#)

⁴Bayesian inference is especially useful in VAR settings because of the regularization that a judicious prior choice allows. In larger VAR systems direct maximization of the likelihood is generally not feasible, so full-information frequentist inference is not commonly used in our setting.

[et al., 2024](#). Our paper also relates to an important recent contribution by [Jarociński \(2024\)](#), who identified monetary policy shocks in the US using t -distributed shocks. In the forecasting literature, researchers often rely on a recursive ordering to estimate VAR models with independent t -distributed shocks equation by equation ([Clark and Ravazzolo, 2015](#); [Chiu et al., 2017](#); [Chan, 2020](#)), which scales well to larger VARs.⁵ However, this assumption imposes strong timing restrictions on the shocks' within-period effects, making it less suitable for structural analysis.⁶

We follow a Bayesian approach and propose a new parametrization of the model that allows us to derive in closed form the conditional posterior distributions of the parameters that currently require a Metropolis-Hastings step. By doing so, we are in a position to propose a Gibbs sampler to explore the joint posterior distribution of the model. To the best of our knowledge, this is the first paper that develops a Gibbs sampler for SVAR models with t -distributed structural shocks. Gibbs samplers exploit conjugate prior distributions so that the conditional posterior in each block is of known form. Because we want all our identification to come from the non-Gaussianity of the data, we derive a Gibbs sampler that puts an uninformative prior on the effects of the structural shocks on impact.

Our paper relates to the broader literature on non-Gaussianity in SVARs, even when shocks are not assumed to be t -distributed. For instance, [Braun \(2023\)](#) and [Herwartz \(2018\)](#) use non-parametric procedures, while [Drautzburg and Wright \(2023\)](#) combine moment conditions on higher order moments with sign restrictions. Using these higher order moments, they can detect whether the identified shocks are independent and discard parameter values that would imply shocks that are not independent. We also maintain the assumption of independent shocks to aid identification.⁷ [Andrade et al. \(2023\)](#) present a methodology to directly use flexible restrictions on higher moments and combine those with standard sign restrictions. By assuming t -distributed shocks, our approach provides direct estimates of the degrees of freedom, which measure deviations from Gaussianity. It is also more parsimonious than non-parametric methods, though any parametric assumption risks misspecification. We therefore provide Monte Carlo evidence that our method still performs well even if the

⁵Another approach that allows researchers to estimate the model equation by equation is to introduce as many Gaussian residuals as observables, in addition to any non-Gaussian shocks, thus effectively introducing a factor structure in forecasts errors ([Prüser, 2024](#)).

⁶This is the same argument that has traditionally been levied against early SVAR studies that assumed exactly such a relationship between shocks and variables ([Sims, 1980](#)).

⁷For theoretical results on what conditions can replace independence in non-Gaussian models to achieve identification, see [Mesters and Zwiernik \(2024\)](#).

true process is not t -distributed. Our paper extends the Gibbs-sampling framework for SVARs with Gaussian errors (Waggoner and Zha, 2003a) to a non-Gaussian setting. Like recent work on Gibbs samplers for structural VARs with stochastic volatility (Chan et al., 2024; Wu and Koop, 2023), our approach is invariant to variable ordering.

The paper is structured as follows: Section 2 provides an illustration that highlights how higher moments help with identification in non-Gaussian settings. Section 3 then presents details of our VAR model, while Section 4 presents our results on US GDP. Section 5 reports the results on Monte Carlo simulations using our methodology. Section 6 concludes.

2 An Illustration

In this section, we use a bivariate model to illustrate how higher moments identify structural parameters when shocks follow independent t -distributions rather than Gaussian distributions. We consider the following data generating process:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}, \quad (1)$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})'$ are independently t distributed structural shocks with degrees of freedom $\mathbf{v} = (v_1, v_2)'$. We normalize the variance of these shocks to 1 (which means that we assume the variance is finite and thus $v_1, v_2 > 2$). Model (1) can be rewritten as

$$\mathbf{y}_t = B_c Q \boldsymbol{\epsilon}_t, \quad (2)$$

where we define $\Sigma = BB'$ as the covariance matrix of \mathbf{y}_t , B_c the Cholesky decomposition of Σ , and $Q = B_c^{-1}B$ an orthogonal matrix so that $QQ' = I$. Finally, we define the individual elements of B as $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. Our goal is to identify (B, \mathbf{v}) , or equivalently, (B_c, Q, \mathbf{v}) .

Gaussianity appears here as a special case with $v_i = \infty$, $i = 1, 2$. Under Gaussianity, only second moments distinguish different parameter values of the model, so identification relies on $\Sigma = BB'$ because higher moments do not vary across Gaussian distributions. The following restrictions on elements of B hold under both Gaussianity

and more generally with a t -distribution:

$$E(y_{1t}^2) = b_{11}^2 + b_{12}^2, \quad (3a)$$

$$E(y_{1t}y_{2t}) = b_{11}b_{21} + b_{12}b_{22}, \quad (3b)$$

$$E(y_{2t}^2) = b_{21}^2 + b_{22}^2. \quad (3c)$$

Because these three equations involve four unknowns, B_c is identified, but B is not ([Kilian and Lütkepohl, 2018](#)). Thus, in the Gaussian setting, we would need to impose a restriction (e.g., $b_{12} = 0$ to obtain $B = B_c$).

Higher moments help when distributions are non-Gaussian. Since we assume symmetric distributions (Gaussian or t), third moments vanish, so we must rely on fourth moments. In this bivariate example, the relevant expressions (derived in the Online Appendix) are:

$$E(y_{1t}^4) = b_{11}^4\kappa_1 + 6b_{11}^2b_{12}^2 + b_{12}^4\kappa_2, \quad (4a)$$

$$E(y_{1t}^3y_{2t}) = b_{11}^3b_{21}\kappa_1 + 3b_{11}b_{12}^2b_{21} + 3b_{11}^2b_{12}b_{22} + b_{12}^3b_{22}\kappa_2, \quad (4b)$$

$$E(y_{1t}^2y_{2t}^2) = b_{11}^2b_{21}^2\kappa_1 + b_{12}^2b_{21}^2 + 4b_{11}b_{12}b_{21}b_{22} + b_{11}^2b_{22}^2 + b_{12}^2b_{22}^2\kappa_2, \quad (4c)$$

$$E(y_{1t}y_{2t}^3) = b_{11}b_{21}^3\kappa_1 + 3b_{11}b_{21}b_{22}^2 + 3b_{12}b_{21}^2b_{22} + b_{12}b_{22}^3\kappa_2, \quad (4d)$$

$$E(y_{2t}^4) = b_{21}^4\kappa_1 + 6b_{21}^2b_{22}^2 + b_{22}^4\kappa_2, \quad (4e)$$

with κ_i both the fourth moment and kurtosis of ϵ_{it} , since it holds that $\kappa_i = E(\epsilon_{it}^4)$ due to $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = 1$. Gaussianity implies $\kappa_1 = \kappa_2 = 3$, and all equations (4) become constant in the orthogonal matrix Q . Hence, the fourth moments provide no additional information to identify B . To see this, note that under Gaussianity, the right-hand sides of (4) can be written as functions of the right-hand sides of the second-moment conditions in (3). Since those second moments are invariant to orthogonal rotations, so must be the fourth moments under Gaussianity. Instead, under a t distribution, where $\kappa_i \neq 3$, the expressions in (4) vary with Q (i.e., with alternative choices of B such that $BB' = \Sigma$). As a result, the fourth moments provide additional information beyond what is contained in the second moments.

To illustrate the above point, we provide a numerical example by setting

$$B_{\text{true}} = \begin{pmatrix} 1 & -1.25 \\ 2 & 0.5 \end{pmatrix}.$$

Although these parameter values are arbitrary, one motivation might be the following: If \mathbf{y}_t includes output and prices, and $\boldsymbol{\epsilon}_t$ includes a demand shock and a supply shock (in that order), then a positive demand shock increases output and prices, while a positive supply shock decreases output and increases prices. To highlight identification issues that persist even as sample size grows, we work in population, assuming knowledge of all moments of the data \mathbf{y}_t .⁸

Equations (3) and (4) form a system of eight polynomial equations (up to fifth order) in six parameters, $(b_{11}, b_{12}, b_{21}, b_{22}, \kappa_1, \kappa_2)$.

To reduce the parameter space, we set

$$Q = Q(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad (5)$$

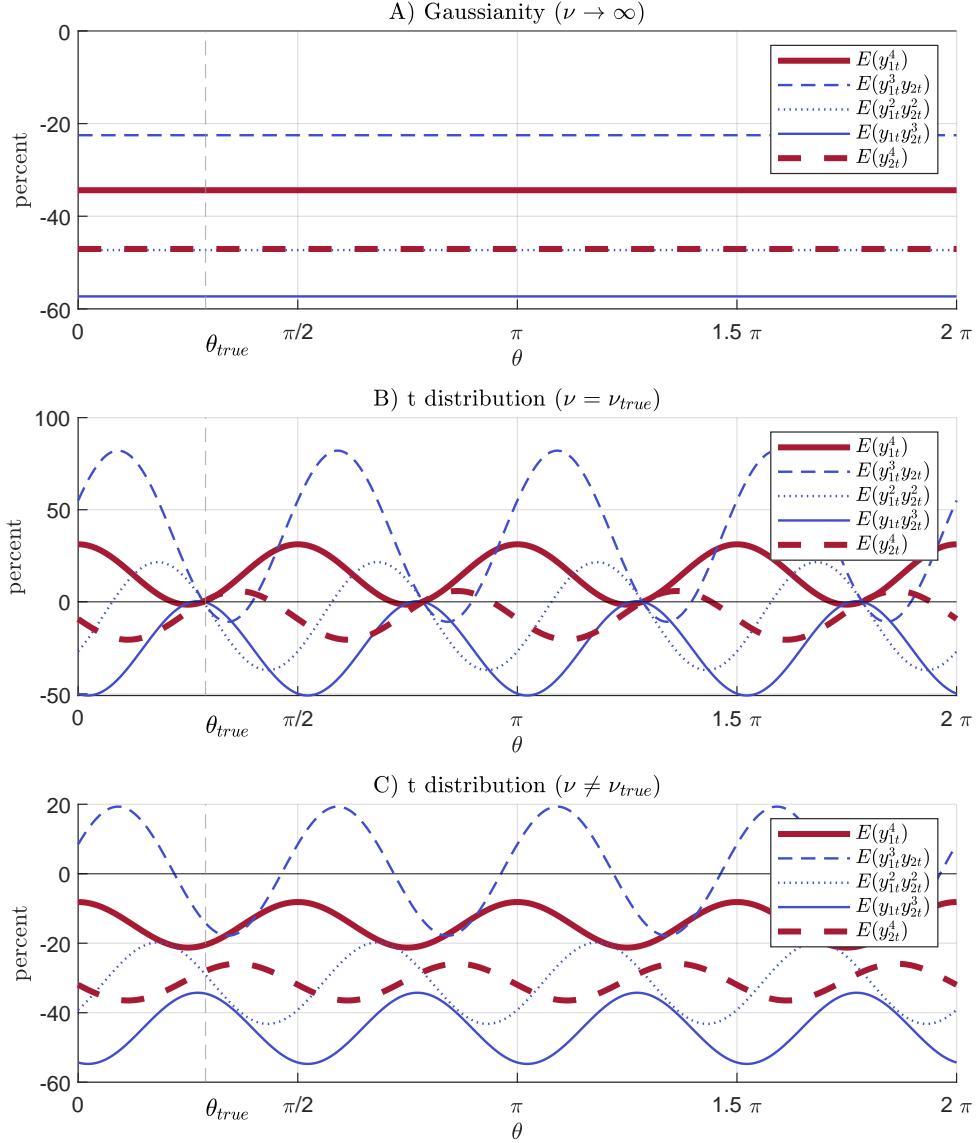
a Givens rotation that produces an orthogonal matrix with $\det(Q(\theta)) = 1$. This assumption is standard in the literature (Canova and De Nicoló, 2002) and loses no generality aside from fixing the determinant (a restriction we come back to below). Next, we impose $\mathbf{v} = (v, v)'$, reducing the unknown parameter space from $(b_{11}, b_{12}, b_{21}, b_{22}, \kappa_1, \kappa_2)$ to (θ, v) , and we treat B_c as known (since B_c is identified by second moments alone). The value of θ consistent with B_{true} is $\theta_{\text{true}} = 0.29\pi$, and we set $v_{\text{true}} = 6$. We then compute the fourth moments implied by $(\theta_{\text{true}}, v_{\text{true}})$. Checking whether these fourth moments identify the model parameters amounts to asking if the system (4) has a solution other than $(\theta, v) = (\theta_{\text{true}}, v_{\text{true}})$. By construction, any θ is consistent with the second moments alone.

Each of the three panels in Figure 1 show the fourth moments from equation (4) evaluated over a grid of θ , conditioning on a value of v shown in each panel. The moments are shown in percentage deviation from the true moments associated with $(\theta_{\text{true}}, v_{\text{true}})$. Panel A) reports the case when evaluating moments (4) for $v = \infty$, hence $\kappa_i = 3$, $i = 1, 2$ (Gaussianity). The moments are constant in θ , showing that fourth moments do not provide any information when the model is Gaussian. As an example, the true moments $E(y_{1t}^4) = 30.01$ and $E(y_{1t}^2) = 2.56$ imply the population kurtosis of y_{1t} equal to $\frac{E(y_{1t}^4)}{(E(y_{1t}^2))^2} = 4.57$. Evaluating (4) using $\kappa_i = 3$, $i = 1, 2$ gives $E(y_{1t}^2) = 19.70$ and kurtosis $\frac{E(y_{1t}^4)}{(E(y_{1t}^2))^2} = 3$ for *any* value of θ , with $E(y_{1t}^4)$ being 34% below the true value (red thick line), showing that the Gaussian model is misspecified. Since the

⁸We examine our approach in finite samples, under a more realistic data-generating process, in Section 5.

likelihood function of model under Gaussianity is uniquely pinned down by the first and second moments, higher moments provide no additional information.

Figure 1: Illustrative example



Note: Fourth moments are computed using equations (4) evaluated for $B = B_{chol} \cdot Q$, with B_{chol} set equal to the Cholesky decomposition of $B_{true} \cdot B'_{true}$ and Q set as from equation (5) over a grid on θ . The figure shows the moments in percentage deviation from the true values associated with B_{true} . The top panel evaluates equations (4) for $\kappa_i = 3$, $i = 1, 2$. The middle and lower panel set κ_i implied by $v = v_{true} = 6$ or $v = 9$.

Things change dramatically when we set $\kappa_i = 6$ for $i = 1, 2$, the true value in our illustration. The middle panel of Figure 1 shows that the fourth moments in

(4) are no longer constant in θ . Instead of infinitely many possibilities (as in the Gaussian case), there are now exactly four values of θ that match the five moments from (4): $(0.29\pi, 0.78\pi, 1.26\pi, 1.78\pi)'$. We return to this residual identification issue below when we discuss normalization. In the middle panel, we specifically evaluate the moments at the true degrees of freedom. This step matters only if the data's fourth moments let us infer the correct degrees of freedom in the first place. Panel C of [Figure 1](#) demonstrates that using $v \neq v_{\text{true}}$ (specifically $v = 9$ here) prevents an exact match of all moments. Hence, the model's fourth moments - and thus the likelihood function - are informative about both the degrees of freedom in the t -distribution and the elements of B .

[Table 1](#) clarifies how the four values of θ identified in Panel B of [Figure 1](#) relate to B . Following [Lanne et al. \(2017\)](#), the likelihood function of the t -distributed model has $2! \cdot 2^2 = 8$ modes, which differ from B_{true} by the sign or the ordering of the shocks. Matrix a in [Table 1](#) shows B_{true} . The remaining matrices on the left differ from B_{true} by flipping the sign of one or both columns, while those on the right swap the columns. From [Lanne et al. \(2017\)](#), we know that the model's likelihood peaks at these eight solutions. Economically, all solutions produce the same impulse responses, once we account for shock ordering and sign. The four θ -values from Panel B of [Figure 1](#) correspond to matrices a, f, g , and d in [Table 1](#). To generate all solutions listed in the table, we can generalize the example to

$$\mathbf{y}_t = B_c Q(\theta) P \boldsymbol{\epsilon}_t,$$

where P is a permutation matrix. Under this specification, $\theta = 0.29\pi$ yields matrix a (no column permutation) or matrix b (with column permutation), but the implied Q for matrix b has determinant -1 and thus cannot be generated by a Givens rotation. The same reasoning applies to the remaining θ -solutions.

[Table 1](#) also shows why it is important to address the inference challenges posed by identification up to the sign and ordering of shocks. Although the eight modes in [Table 1](#) are economically identical, if the posterior sampler jumps among them, the resulting posterior approximation can be highly misleading about the genuine uncertainty ([Hamilton et al., 2007](#)). To avoid this problem, prior work has proposed normalization schemes ([Jarociński, 2024; Waggoner and Zha, 2003b](#)). In this paper, we adopt their insight of selecting parameter draws closest to a particular mode during sampling, and we demonstrate how to implement this strategy efficiently, even in

Table 1: Modes of the likelihood of the model in Panel B) from [Figure 1](#)

	B	θ	permute		B	θ	permute
a)	$\begin{pmatrix} 1 & -1.25 \\ 2 & 0.5 \end{pmatrix}$	0.29π	no	b)	$\begin{pmatrix} -1.25 & 1 \\ 0.5 & 2 \end{pmatrix}$	0.29π	yes
c)	$\begin{pmatrix} -1 & -1.25 \\ -2 & 0.5 \end{pmatrix}$	0.78π	yes	d)	$\begin{pmatrix} 1.25 & 1 \\ -0.5 & 2 \end{pmatrix}$	1.78π	no
e)	$\begin{pmatrix} 1 & 1.25 \\ 2 & -0.5 \end{pmatrix}$	1.78π	yes	f)	$\begin{pmatrix} -1.25 & -1 \\ 0.5 & -2 \end{pmatrix}$	0.78π	no
g)	$\begin{pmatrix} -1 & 1.25 \\ -2 & -0.5 \end{pmatrix}$	1.26π	no	h)	$\begin{pmatrix} 1.25 & -1 \\ -0.5 & -2 \end{pmatrix}$	1.26π	yes

Note: $B_{true} = \begin{pmatrix} 1 & -1.25 \\ 2 & 0.5 \end{pmatrix}$ corresponds to matrix a). The remaining matrices c, e, g on the left differ from B_{true} up to the sign of the first or second column, or both. Matrices b, d, f, h differ from B_{true} regarding the ordering of the columns, and sign of the columns.

large-scale systems.

3 Our Approach

This section describes our linear VAR model with independent t -distributed structural shocks. We present our proposed reparametrization, explain how it yields a Gibbs sampler, and then detail the computational steps we use to address the model's identification issues due to sign and permutation switches of the impact responses. We refer to Appendix E of the Online Appendix for a full description of the posterior sampling procedure.

3.1 The Model

The model is given by

$$\mathbf{y}_t = \mathbf{c} + \sum_{l=1}^p \Pi_l \mathbf{y}_{t-l} + B \boldsymbol{\epsilon}_t, \quad (6a)$$

$$\epsilon_{it} \sim t(v_i). \quad (6b)$$

The $k \times 1$ vector \mathbf{y}_t collects the k endogenous variables of the model, whose evolution depends on the $k \times 1$ vector of constants \mathbf{c} as well as p lags of the endogenous variables. The $k \times m$ matrix $\Pi = [\mathbf{c}, \Pi_1, \dots, \Pi_p]$ contains the constant terms and the autoregressive parameters, with $m = 1 + kp$. We will refer to $\boldsymbol{\phi} = \text{vec}(\Pi)$ as the vector that stacks the

columns of Π vertically. The $k \times 1$ vector $\boldsymbol{\epsilon}_t = (\epsilon_1, \dots, \epsilon_i, \dots, \epsilon_k)'$ contains the structural shocks whose effects we want to study. Following previous work such as [Jarociński \(2024\)](#) and [Brunnermeier et al. \(2021\)](#), we assume that the structural shocks are t -distributed, statistically independent across time and cross-sectionally. v_i represents the degrees of freedom of shock i , where we define $\mathbf{v} = (v_1, \dots, v_i, \dots, v_k)$. We set the scale parameters of the t -distribution to normalize the variance of the structural shocks to unity, and assume $v_i > 2, \forall i$.

Following [Geweke \(1993\)](#), we use a first reparametrization that delivers an equivalent model that explicitly interprets the t distributions as scale-mixtures of Normal distributions. We introduce the variable $D_t = \text{diag}(d_{1t}, \dots, d_{it}, \dots, d_{kt})$ and specify this model as

$$\mathbf{y}_t = \mathbf{c} + \sum_{l=1}^p \Pi_l \mathbf{y}_{t-l} + BD_t^{\frac{1}{2}} \boldsymbol{\epsilon}_t, \quad (7a)$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, I). \quad (7b)$$

The entries $\{D_t\}_{t=1}^T$, which are combined into the $kT \times kT$ block diagonal matrix $D = \text{diag}(D_1, \dots, D_t, \dots, D_T)$, are treated as unknown parameters. Conditioning on D , model (7) is Gaussian. [Geweke \(1993\)](#) derives a distribution of D_t so that the two models imply the same distribution for \mathbf{y}_t , as we discuss next.

3.2 Prior Distributions

We define $p(\phi, B, \mathbf{v})$ as the joint prior distribution for the parameters of model (6), and use the same prior for the same parameters in (7). We then introduce an inverse Gamma prior on the latent variables d_{it} parametrized as

$$p(d_{it}|h_i, r_i) = \frac{r_i^{h_i}}{\Gamma(h_i)} \cdot d_{it}^{-h_i-1} e^{-r_i \cdot \frac{1}{d_{it}}}, \quad (8a)$$

$$h_i = \frac{v_i}{2}, \quad (8b)$$

$$r_i = \frac{v_i - 2}{2}. \quad (8c)$$

d_{it} is prior independent across (i, t) and prior independent on (ϕ, B) . As shown by [Geweke \(1993\)](#) in the context of univariate models, when prior (8) is used, the posterior distribution from the alternative model (7) after marginalizing out D coincides

with the joint posterior distribution from the original model (6). Yet, working with the alternative model (7) is convenient because, conditioning on D , the model is Gaussian and hence existing results from the Gaussian literature apply. In employing this approach, we follow a large literature on univariate and multivariate models, see, for instance, [Clark and Ravazzolo \(2015\)](#), [Chiu et al. \(2017\)](#), [Karlsson and Mazur \(2020\)](#) and [Karlsson et al. \(2023\)](#).

More precisely, we work with the following prior distribution:

$$p(\boldsymbol{\phi}, B, \mathbf{v}, D) = p(\boldsymbol{\phi}, B, \mathbf{v}) \cdot p(D|\mathbf{v}), \quad (9)$$

$$= p(\boldsymbol{\phi}) \cdot p(B) \cdot p(\mathbf{v}) \cdot p(D|\mathbf{v}), \quad (10)$$

$$p(D|\mathbf{v}) = \prod_{t=1}^T \prod_{i=1}^k p(d_{it}|v_i). \quad (11)$$

The prior $p(d_{it}|v_i)$ is the inverse Gamma prior from equation (8). We work with a Normal prior for $\boldsymbol{\phi}$, which nests priors already used for large VARs ([Ba  bura et al., 2010](#)), but our method can be extended to other priors specifically derived for large VARs with many observables ([Chan, 2022](#)). We use a flat, improper prior on B , and discuss alternative possible priors in the next section. Last, our method works with a wide range of candidate priors for \mathbf{v} . In our simulations and application we use a truncated Normal prior.

3.3 Posterior Sampling

We want to explore the joint posterior distribution of model (7) via a Gibbs sampler. This, in turn, requires deriving the conditional posterior distributions of each of the parameters of the model.

The literature already knows how to conveniently draw from some of these distributions. It is straightforward to show that $p(\boldsymbol{\phi}|Y, B, D, \mathbf{v})$ is a Normal distribution, and [Geweke \(1993\)](#) shows that $p(D|Y, \boldsymbol{\phi}, B, \mathbf{v})$ is a Gamma distribution. These results have been extensively used in the literature, see for instance [Chiu et al. \(2017\)](#), [Lanne and Luoto \(2020\)](#) and [Anttonen et al. \(2024\)](#). For a Gibbs sampler to be feasible for this model, it remains to derive a convenient way of sampling from the remaining

conditional distributions,

$$p(B|Y, \phi, D, v), \quad (12)$$

$$p(v|Y, \phi, B, D). \quad (13)$$

We concentrate our analysis on these two, starting from the former.

Developing a convenient way of sampling from $p(B|Y, \phi, D, v)$ has so far proved challenging. Following Cogley and Sargent (2005), many papers in the literature parametrize the model using

$$A = B^{-1}, \quad (14)$$

and achieve a computationally convenient sampling procedure for $p(B|Y, \phi, D, v)$ that is feasible under the assumption that A (and hence B) is triangular (Clark and Ravazzolo, 2015, Chiu et al., 2017, Karlsson and Mazur, 2020 and Karlsson et al., 2023). However, since the shocks of the model are independent and non-Gaussian, all elements in A and B are identified, and the triangular restrictions on (A, B) become over-identifying. So far, the literature that introduces no zero restrictions on (A, B) has resorted to simulation-based methods, either employing a Metropolis-Hastings step on $p(B|Y, \phi, v, D)$ (as in Lanne and Luoto, 2020, Brunnermeier et al., 2021), or employing a more involving MCMC procedure on the full joint posterior distribution $p(\phi, B, v, D|Y)$ (Anttonen et al., 2024). Our main contribution is to develop a way of drawing from the conditional posterior distributions (12) and (13) using only distributions of common form. This means that no Metropolis-Hastings step is required, which makes the analysis computationally much less demanding.

The core insight for how to sample from (12) hinges on a new reparametrization, which, to our knowledge, has not been used in Bayesian econometrics. Although our benchmark approach sets a flat prior on B , here we parametrize A instead, introducing the decomposition

$$A = \Lambda L U, \quad (15)$$

where Λ is a diagonal $k \times k$ matrix with entries λ_i , and L and U are lower- and upper-triangular $k \times k$ matrices, respectively, both constrained to have ones on their main diagonals. We assume $\lambda_i \neq 0$ for all i , which is a necessary and sufficient condition for (15) to exist and be unique.⁹

The choice between using the A or B parametrization is inconsequential for our

⁹Under this reparametrization, $B = U^{-1} L^{-1} \Lambda^{-1}$.

method: the Jacobian of the inverse mapping is straightforward to derive, so one could also impose a flat prior directly on A , which we do in the Online Appendix. Our decomposition is related to, but distinct from, the standard LDU decomposition. One can derive it in two steps: first, apply a LU decomposition to A , then decompose the resulting L into the product of a diagonal matrix and a lower triangular matrix with unit diagonal entries. Thus, the existence and uniqueness conditions match those of the LU decomposition. However, since the likelihood is invariant under premultiplication of A by any permutation matrix, for every non-singular A , there always exists a permutation matrix P such that PA admits the (unique) LU decomposition. Hence, essentially (15) entails no restrictions other than the nonsingularity of B (or A). As a consequence, we can think of (15) much like other reparametrizations employed in the SVAR literature.¹⁰

As is well known (see, for example, Brunnermeier et al., 2021), the determinant of A enters the likelihood function, complicating posterior derivation unless additional assumptions are imposed (see, e.g., Kocięcki et al., 2012; Arias et al., 2018). However, our (Λ, L, U) reparametrization in (15) conveniently yields

$$|\det(A)| = \prod_{i=1}^k |\lambda_i|, \quad (16)$$

meaning the determinant depends only on the diagonal entries of Λ (because L and U both have ones on their main diagonals). Moreover, $|\det(A)|$ depends on Λ in a way that resembles the kernel of a Gamma distribution. These two properties make a Gibbs sampler feasible, because they imply a Gamma conditional posterior for a function of Λ and a Normal conditional posterior for a function of both L and U . Details of those posterior distributions can be found in the Online Appendix.

Our Gibbs sampler requires a flat prior on either B or A . In an independent, non-Gaussian framework, B and A are identified (up to sign and column permutations, addressed in the next section), so the exact form of the prior is less critical than in the Gaussian case, where it encodes necessary identification restrictions (Baumeister and Hamilton, 2015; Arias et al., 2019; Inoue and Kilian, 2020).

This design streamlines posterior sampling for B . To complete our Gibbs proce-

¹⁰For example, Arias et al. (2018) apply sign restrictions by setting $A = \mathbf{Q} h(\Sigma)$, where Σ is the reduced-form covariance matrix of a Gaussian VAR, $h(\cdot)$ is a unique decomposition, and \mathbf{Q} is orthogonal. In contrast, Wu and Koop (2023) use the eigenvalue-eigenvector decomposition of Σ to simplify sampling in large-scale VAR models.

dure, we must also show that the conditional posterior for the degrees of freedom in (13) has a tractable form. We accomplish this via a Griddy-Gibbs sampler (Ritter and Tanner, 1992), which discretizes ν . In our simulations and empirical application, we let ν range over a fine grid between 3 and 60, specifying a truncated Normal prior over this interval. Further details appear in the Online Appendix.

3.4 Identification up to Sign and Permutations of the Shocks

Because the model features independent non-Gaussian shocks, theoretical restrictions are not needed for statistical identification. However, as Lanne et al. (2017) show, identification holds only up to the sign and permutation of the columns of B (or equivalently, the rows of A). Accounting for this indeterminacy can pose computational challenges: any $k \times k$ matrix B the algorithm samples corresponds to as many as $k! \cdot 2^k$ alternative matrices that differ solely by column sign or permutation. Although these differences are economically meaningless, they must be managed in the sampler to avoid mixing shocks of different types.

A second methodological contribution of our paper is thus a new, computationally efficient way to handle sign and permutation indeterminacies in VARs with non-Gaussian shocks.¹¹ We first compute a matrix \hat{B} that approximates one of the model's modes, which serves as a target in the Gibbs sampler. For example, \hat{B} could be the maximum-likelihood estimator of B . Let P_s be a signed-permutation matrix of dimension $k \times k$, with non-zero elements equal to either 1 or -1. In each sampler iteration, we draw (Λ, L, U) , thus obtaining A and B . We then store BP_s , where P_s minimizes

$$\Delta = \text{trace} \left[(BP_s - \hat{B})' (\hat{B} \hat{B}')^{-1} (BP_s - \hat{B}) \right]. \quad (17)$$

In other words, we choose the signed-permutation matrix P_s that minimizes the weighted distance between BP_s and the target \hat{B} . As shown in Appendix D of the Online Appendix, minimizing Δ generalizes the Likelihood Preserving (LP) normalization of Waggoner and Zha (2003b) - originally designed for Gaussian SVARs - to non-Gaussian SVARs (including t -distributions). We therefore refer to this approach as a *generalized LP normalization*.

The most important feature of our approach is computational. Naively, one might

¹¹A researcher could use restrictions to remove this multiplicity, as we do, or sample the full unnormalized posterior. In the latter case, strategies from mixture-model sampling that address permutation issues (Geweke, 2007; Frühwirth-Schnatter, 2001) might be applied.

compute the distance (17) for all $k! \cdot 2^k$ possible signed-permutation matrices in each Gibbs iteration, which is infeasible. Instead, as we discuss in Appendix D of the Online Appendix, we show how to apply the Hungarian algorithm from combinatorial optimization to this problem, which finds the optimal P_s at negligible computational cost—even for large models.

Our method also differs from other approaches to identification up to sign and permutation in non-Gaussian SVARs. Compared to [Lanne et al. \(2017\)](#), our method reduces the risk of mode-switching by explicitly targeting a single mode \hat{B} . Relative to [Brunnermeier et al. \(2021\)](#), we avoid imposing an informative prior on A , letting the likelihood’s shape provide identification. Finally, unlike [Jarociński \(2024\)](#), we do not evaluate a target function for each of the $k! \cdot 2^k$ permutations every Gibbs iteration. This makes our approach tractable for large systems. For example, on a standard laptop, we can solve the minimization problem for a 20-variable VAR in 0.0001 seconds (during each Gibbs iteration). Evaluating the target function for each permutation for such a large model would require $20! \cdot 2^{20} \approx 2.5 \cdot 10^{24}$ functions evaluations. See the Online Appendix for details and additional speed comparisons.

4 What Shocks Drive Real GDP?

We now apply our framework to determine which shocks matter most for real GDP in the United States. [Angeletos et al. \(2020\)](#) use a structural VAR with Gaussian shocks to uncover a “Main Business Cycle” shock that explains the bulk of volatility in real activity over business-cycle frequencies. They show that (i) a single shock can account for almost 80% of GDP volatility at business cycle frequencies when GDP itself is used to measure real activity,¹² (ii) the shock behaves like a weakly inflationary demand shock based on impulse responses, (iii) it is largely unrelated to TFP dynamics, and (iv) it does not explain long-run fluctuations in real activity.

We revisit and extend these findings under non-Gaussian shocks. Specifically, we ask whether the business cycle can be attributed to a single shock or if multiple shocks share responsibility for GDP volatility. Our specification closely follows [Angeletos et al. \(2020\)](#), using 10 variables (GDP, investment, consumption, hours worked, unemployment rate, labor share, Fed funds rate, inflation, labor productivity, and TFP), all in log levels. All variables are real and in per capita terms where applicable. The

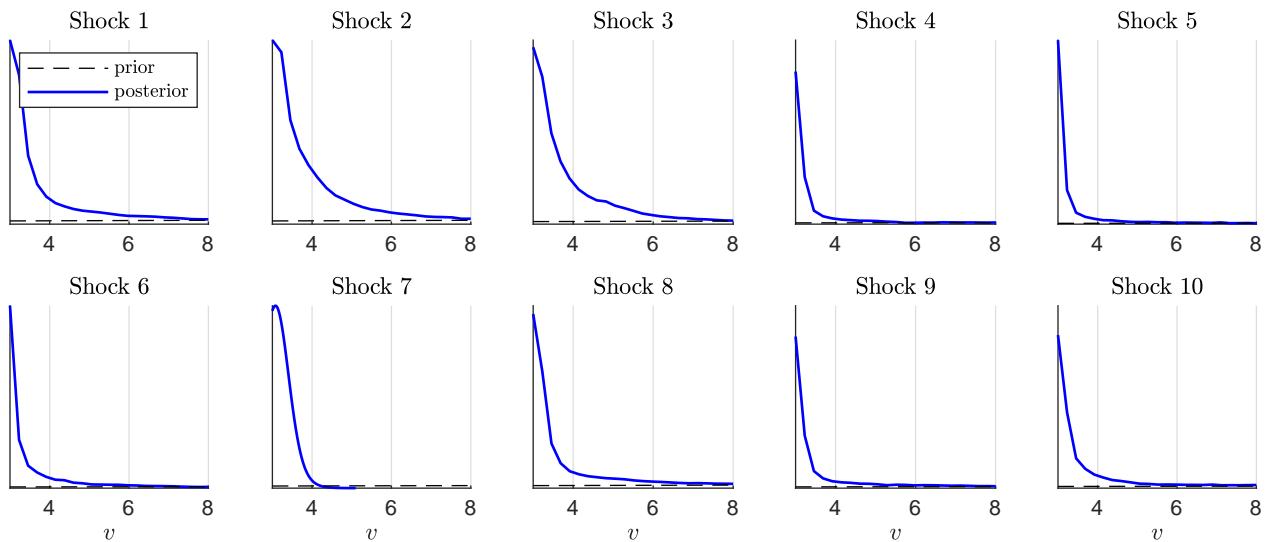
¹²When they instead target unemployment, they still find that their shock explains more than half of the business cycle fluctuations of GDP.

sample is quarterly from 1955Q1 to 2017Q4, and, as in the original study, we include two lags.

We specify the baseline prior for the analysis as follows. For the autoregressive parameters we use the traditional Minnesota prior. For the contemporaneous impulse responses, we use a flat prior. For the degrees of freedom we use a Normal distribution centered at 20 and with variance 4, truncated to be positive to be between (3, 60). Posterior sampling is carried out using the method discussed in Section 3, generating 50,000 draws.

Before interpreting our results, we first check the credibility of our identification assumption. Under our framework, shocks are t -distributed, and we estimate their degrees of freedom as a natural measure of identification strength. Lower degrees of freedom imply stronger deviations from Gaussianity. This contrasts with previous studies, such as Brunnermeier et al. (2021), that fix the degrees of freedom. Figure 2 reports the posterior distributions for the degrees of freedom for the 10 shocks in our model. We see that the prior is updated sharply downward for all shocks, suggesting non-Gaussian shocks. This finding supports our approach, since our identification strategy relies on non-Gaussianity.

Figure 2: Evidence of non-Gaussianity: prior and posterior on the degrees of freedom



Note: The dashed black line shows the prior distribution, which is a $N(20, 20)$ truncated to be positive in the support [3,60]. The solid blue line shows the posterior distribution obtained from 50,000 posterior draws.

Next, we turn to our main question: Is there one key shock that drives GDP? To

answer this question, we compute forecast error variance decompositions at different horizons. To be more specific, we denote by $y_{i,t+h}$ the i -th element of \mathbf{y}_{t+h} and by \mathbf{s}_i the selection vector of the same dimension as \mathbf{y}_{t+h} that has zeros everywhere except at position i . The formula for the forecast error variance decomposition (FEVD) of variable i of \mathbf{y}_t at horizon h due to the k -th structural shock is:

$$\text{FEVD}_{i,k}(h) = \frac{\sum_{m=0}^{h-1} (\mathbf{s}'_i \boldsymbol{\Theta}_m \mathbf{s}_k)^2}{\sum_{m=0}^{h-1} \mathbf{s}'_i \boldsymbol{\Theta}_m \boldsymbol{\Theta}'_m \mathbf{s}_i},$$

where $\boldsymbol{\Theta}_m$ are the impulse response coefficients that can be obtained by repeated substitution into the VAR model.¹³ We use this time domain variance decomposition as it is arguably more common in the literature than the frequency domain version used in [Angeletos et al. \(2020\)](#). It is, however, worth noting, that [Angeletos et al. \(2020\)](#) find that in their model, their frequency domain-based variance decomposition based identification behaves similarly to a time domain variance decomposition-based identification at horizon $h = 4$. We consider variance decompositions up to 24 quarters and find that *no shock* explains more than 25 % of the relative forecast error variance of real GDP at *any horizon*.¹⁴ The two shocks that matter most are shock 3, peaking around four quarters, and shock 7, peaking at about three years. What do these shocks represent?

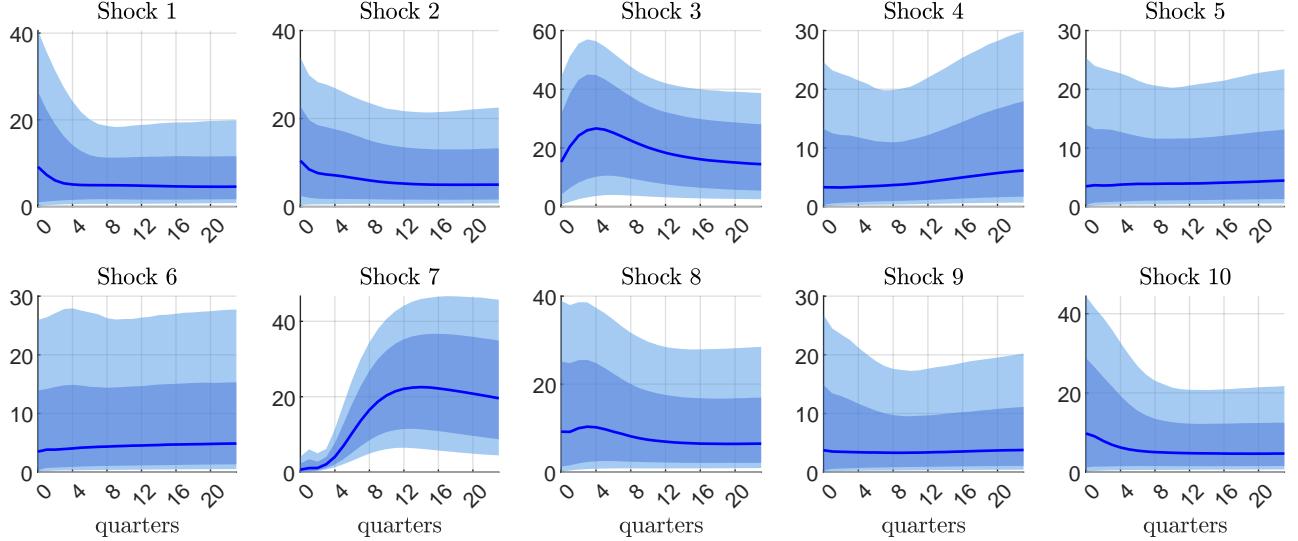
Figure 4 shows the impulse responses to these shocks, normalized to a unit standard deviation impulse. The top two rows show the responses to shock 3, which generally raises real activity, lowers unemployment, but has no effect on TFP and one a small (if any) effect on inflation. Thus, this shock closely resembles the main business cycle shock of [Angeletos et al. \(2020\)](#), with the caveat that we find it is less important than suggested in that work (albeit still one of the two key drivers of GDP in the US). The second shock we identify, which is important at longer horizons, also increases real activity, but at the same time lowers inflation and moves TFP. Thus, it is reminiscent of standard TFP shocks ([Kydland and Prescott, 1982](#)). In fact, in the canonical [Smets and Wouters \(2007\)](#) model, a TFP shock leads to lower inflation, just as we find.

Before describing our robustness checks, one might wonder how we arrive at different conclusions from [Angeletos et al. \(2020\)](#) regarding the demand shock (shock 3).

¹³For details on this derivation, see [Lütkepohl \(2005\)](#).

¹⁴In contrast, Figure 12 in their online appendix shows that even the version of their shock that targets unemployment and not GDP directly in the identification gives [Angeletos et al. \(2020\)](#) a time domain forecast error variance decomposition where their shock explains more than 50 percent of GDP fluctuations at horizons up to 5 quarters.

Figure 3: Forecast Error Variance Decomposition of real GDP



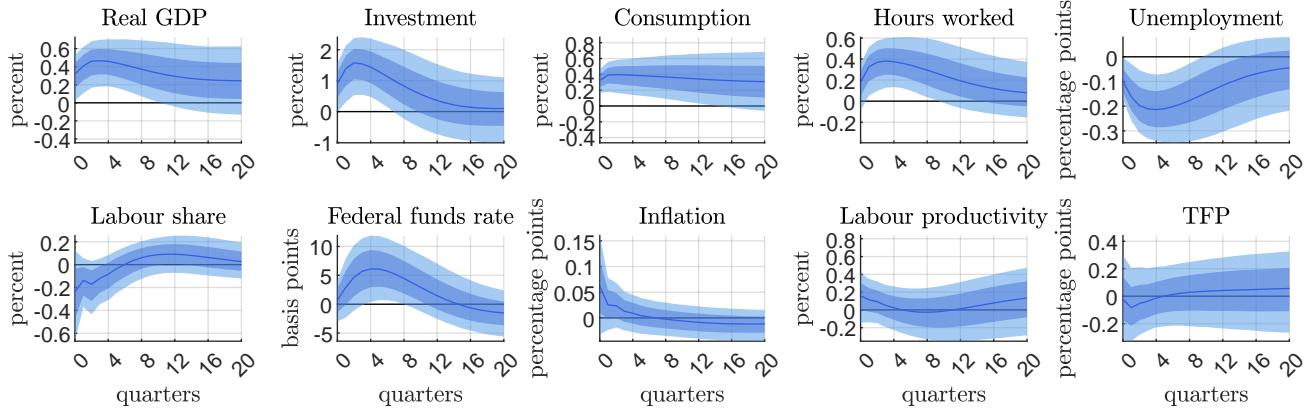
Note: The solid lines show the pointwise posterior median, the shaded areas show the 68% and 90% pointwise posterior credible sets.

After all, we use the same data and lag length, and we find similar impulse responses. The difference stems from our assumption of t -distributed shocks, which changes the shape of the likelihood function and thus affects all parameter estimates. We see clear evidence of non-Gaussianity in this dataset, which influences the posterior distribution and differentiates our results from those of [Angeletos et al. \(2020\)](#). While we still recover a shock that resembles their main business cycle shock, our model's better fit reveals a second shock that significantly affects GDP.¹⁵

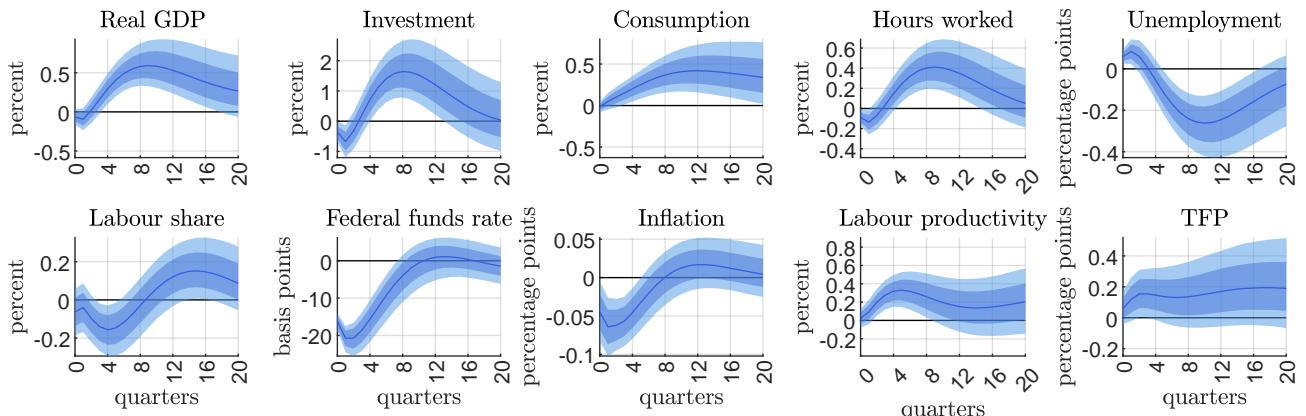
In the Appendix, we show that our results are robust to many alternative specifications. In particular, we consider a model with 4 lags, a model with a flat prior on A rather than $B = A^{-1}$, a model with a longer sample (where the Fed Funds rate is replaced with a T-bill rate), and a model with credit spreads.

¹⁵Our model estimates low degrees of freedom for the t-distribution. If the degrees of freedom were high, the model would be nearly Gaussian, matching the fit of the model in [Angeletos et al. \(2020\)](#) more closely (remember that we have a nearly uninformative prior on the degrees of freedom that, if anything, favors Gaussianity).

Figure 4: Impulse responses
Shock 3: demand shock



Shock 7: supply shock



Note: The solid lines show the pointwise posterior median, the shaded areas show the 68% and 90% pointwise posterior credible sets.

5 Validation Using Simulated Data

We use a simulation exercise to validate our methodology, and to further illustrate how it operates. The data-generating process (DGP) is a bivariate VAR of order 6

with no constant term, where the structural shocks are t -distributed. Specifically,

$$\mathbf{y}_t = \sum_{l=1}^6 \tilde{\Pi}_l \mathbf{y}_{t-l} + \tilde{B} \boldsymbol{\epsilon}_t, \quad (18)$$

$$p(\boldsymbol{\epsilon}_t) = \prod_{i=1}^2 p(\epsilon_{it}), \quad (19)$$

$$\epsilon_{it} \sim t(\tilde{v}_i). \quad (20)$$

The model generates the dynamics of two variables, which we label as “output” (ordered first) and the “price level.” The driving shocks-labeled “demand” (ordered first) and “supply”-are t -distributed.

Following [Canova et al. \(2024\)](#), we set the true parameter values of the data-generating process by first specifying a functional form for the impulse responses, then choosing $(\tilde{\Pi}, \tilde{B})$ to match those responses. We parametrize the true impulse responses via the Gaussian basis functions of [Barnichon and Matthes \(2018\)](#), adapted by [Canova et al. \(2024\)](#). Let $\tilde{\psi}_{ij,h}$ be the true impulse response of variable i to shock j at horizon h . We specify

$$\tilde{\psi}_{ij,h} = a_{ij} \cdot \exp \left[- \left(\frac{(h - b_{ij})^2}{c_{ij}^2} \right) + \frac{b_{ij}^2}{c_{ij}^2} \right], \quad (21)$$

where a_{ij} governs the impact effect of shock j on variable i , b_{ij} is an integer marking the horizon at which the peak response occurs (and equals 0 if there is no hump shape), and c_{ij} captures the persistence of the response. Hence, each response over $H + 1$ horizons is fully determined by three parameters, (a_{ij}, b_{ij}, c_{ij}) .

We set $\{a_{ij}, b_{ij}, c_{ij}\}$ to generate the following impulse responses. The first shock is a demand shock that raises output and the price level on impact by 0.6 and 0.7, respectively. Both responses then decline to zero, reaching half their impact effect 5 and 3 horizons later, respectively. The second shock is a supply shock that raises output by 0.4 on impact but lowers the price level by 0.7. These responses both exhibit a hump shape: output peaks 4 horizons after the shock at 50% above its impact level, and the price level peaks 5 horizons later at 100% above (in absolute value) its impact effect.

We specify impulse responses up to horizon $H = 6$, matching the number of lags in the true model. As discussed by [Canova et al. \(2024\)](#), there is a unique solution in $(\tilde{\Pi}, \tilde{B})$ such that the VAR-implied responses coincide with the true responses from equation (21) for $h \leq 6$. In our baseline analysis, we set $\tilde{v}_i = 6$ for $i = 1, 2$. See the

Online Appendix for additional details on the data-generating process.

We first generate a single dataset from our specified data-generating process. We initialize the simulation at the unconditional mean (zero) and create 350 observations. We discard the first 100, retain the next 50 as a training sample, and use the remaining 200 for estimation. We then estimate a VAR with 6 lags (no constant) on the simulated data, applying the following priors: For the autoregressive parameters, we use a Normal prior centered at zero, calibrated according to [Kadiyala and Karlsson \(1997\)](#) and [Canova \(2007\)](#) (using the aforementioned training sample). For the impact matrix B , we specify a flat prior. Finally, for the degrees of freedom, we adopt a truncated Normal prior (support between 3 and 60), centered at 20 with variance 20 for each degrees-of-freedom parameter.

We compute an initial estimate \hat{B} by maximizing the likelihood function via the three-step estimator of [Lanne et al. \(2017\)](#). We initialize the sampler as described in Appendix E of the Online Appendix. The sampler runs for 21,000 draws, discarding the first 1,000 as burn-in. On a standard computer with an Intel i7-7700K 4.2 GHz Quad Core processor and 64 GB RAM, this takes about four minutes.

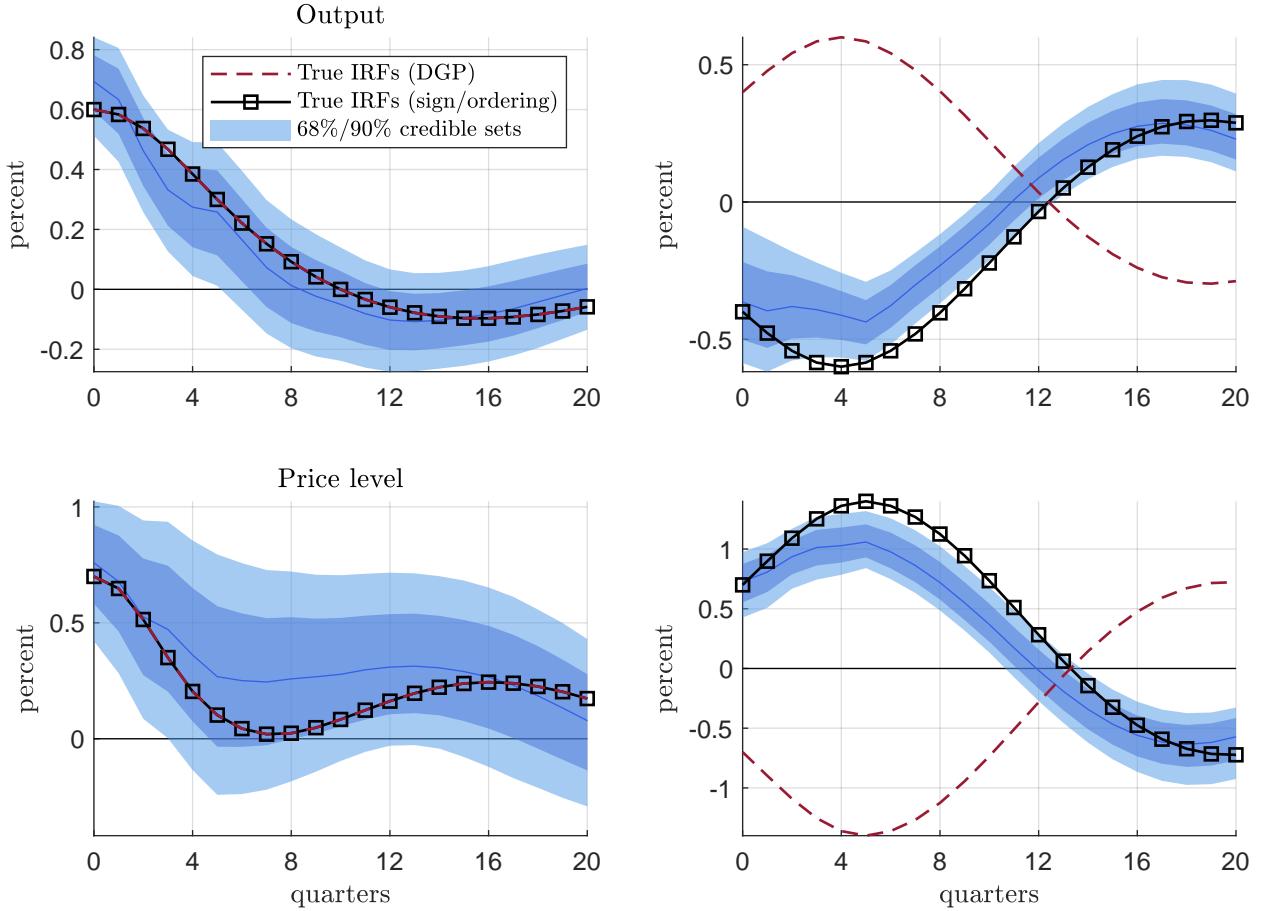
Figure 5 presents our methodology's performance in estimating impulse responses. The blue solid line and shaded regions depict the pointwise median and 68/90% credible sets, respectively. The red dashed line shows the true impulse response from the data-generating process. For comparison, the black squared line shows an alternative set of responses identical to the true ones except that the supply shock's response is flipped in sign.

The figure shows that the sampler correctly recovers the true sign of the impact effects. It also captures that a demand shock exhibits no hump-shaped response, while a supply shock does generate a delayed response. In addition, both the persistence and timing of each response are estimated precisely.

Figure 6 illustrates why our sampler succeeds in recovering the true impulse responses. The solid line represents the posterior distribution of B under our normalization procedure, which addresses sign and permutation indeterminacies. The dotted line shows the posterior without normalization. The blue diamonds mark the true values of B , while the blue dots depict the corresponding entries of the target matrix \hat{B} . By comparing these two, we see that \hat{B} ranks the demand shock first, treating a positively signed shock as expansionary - both consistent with the data-generating process.

Indeed, once our normalization is applied, the marginal posterior for the demand

Figure 5: Impulse response functions

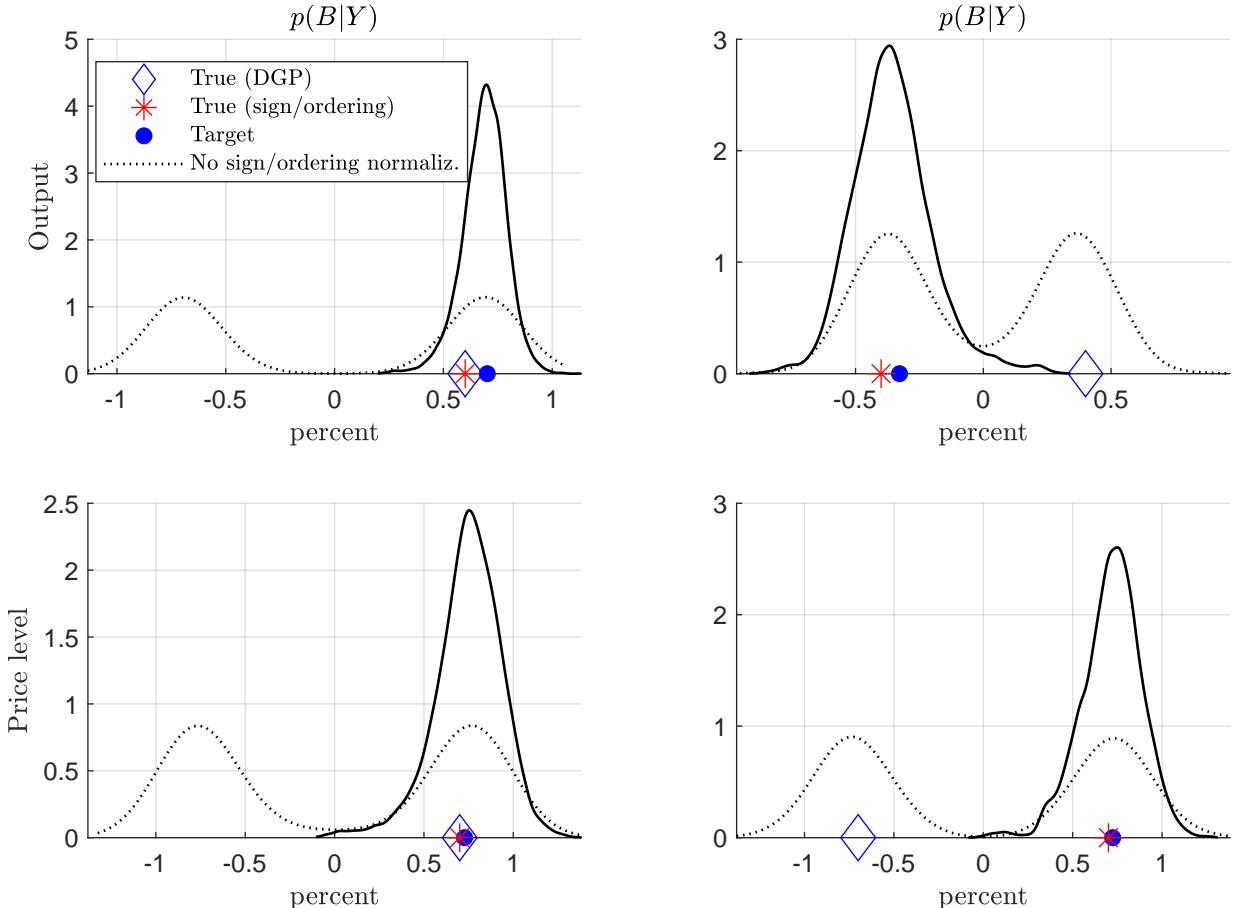


Note: The red dashed lines show the true impulse responses associated with the data generating process. The blue lines and shaded areas show the pointwise posterior median and 68%/90% credible sets. The black squared lines show the sign/permuation of the true impulse responses that is closest to the impulse responses estimated in the posterior.

shock's impact effect concentrates tightly around the true value. In contrast, \hat{B} interprets a positively signed supply shock as contracting output, while the true supply shock actually raises output. The red star shows the true impact effect of the supply shock, but with its sign flipped to match \hat{B} . After correcting for this sign, the model recovers the shock's true impact.

Figure 7 displays the marginal prior (dashed line) and posterior (solid lines) distributions for the degrees of freedom. Our prior on ν is relatively uninformative, yet the figure shows that our method uncovers clear evidence of non-Gaussianity in this sample. The demand shock's degrees of freedom are estimated more precisely than

Figure 6: Impact effect of the shocks (B)

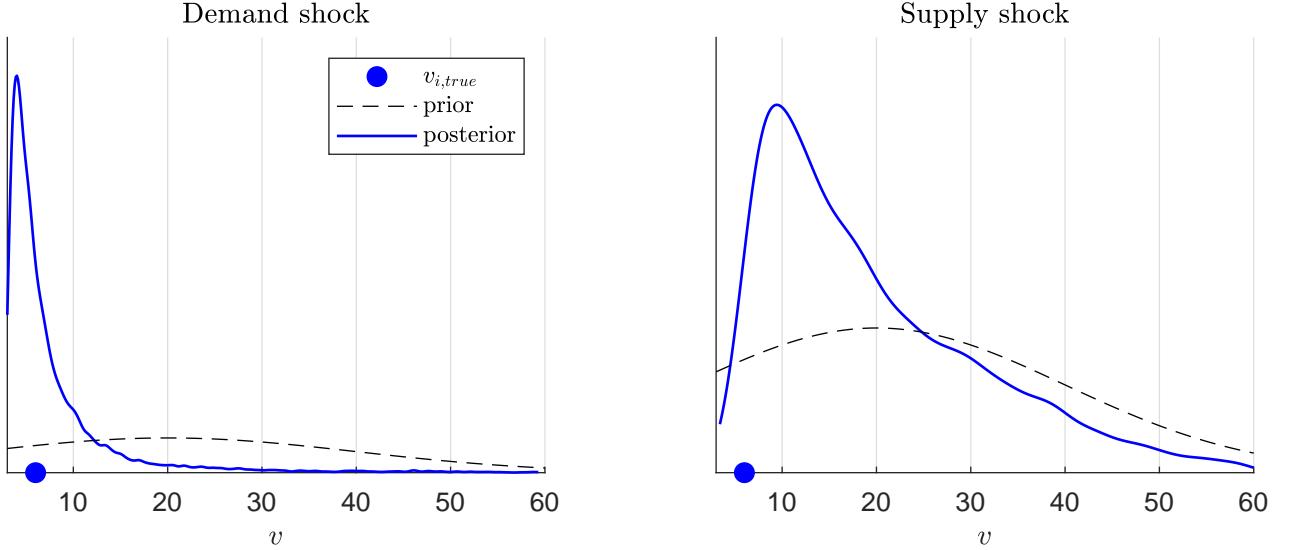


Note: The blue diamond indicates B_{true} . The blue dots indicate the target matrix \hat{B} used for the normalization. The red star indicates the sign/permutation of B_{true} that is the closest to \hat{B} . The continuous line shows the marginal posterior of the entries of B from 20,000 posterior draws when applying the generalized LP normalization. The dotted line shocks the marginal posterior when no normalization is used.

that of the supply shock. However, our estimated impulse responses show that impulse responses are estimated precisely for both shocks. Figure A-3 in the Online Appendix confirms that our ability to estimate \mathbf{v}_{true} improves with larger samples, as expected.

So far, we have illustrated our Bayesian procedure using a single pseudo-dataset. Figure 8 summarizes results from 100 replications. We hold the data-generating process fixed and draw new pseudo-datasets for each replication. The black squared line shows the true impulse response, matching the sign and ordering of the shocks in Figure 5. The blue lines depict the pointwise medians of the estimated responses for each

Figure 7: Degrees of freedom



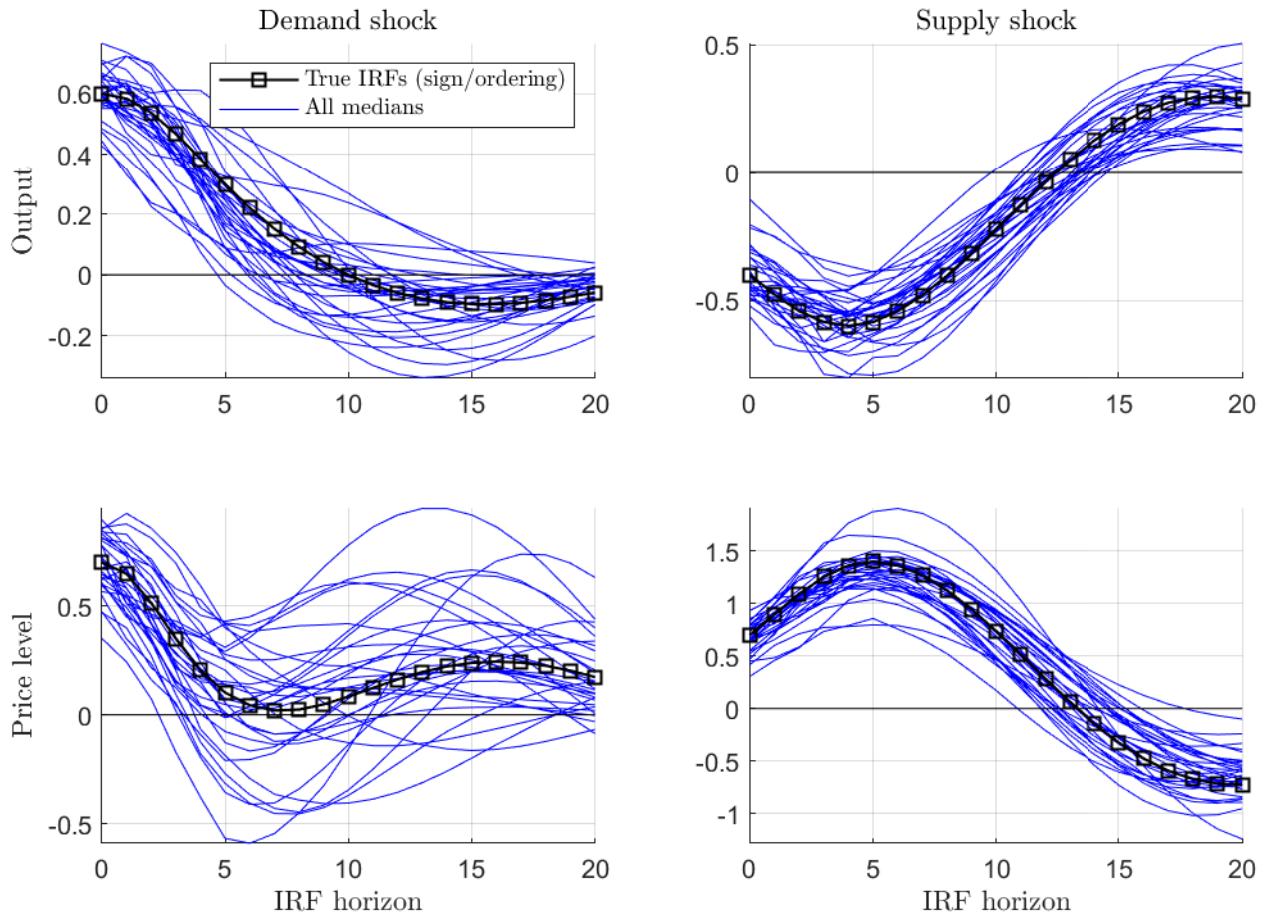
Note: The blue dot indicates the true value of the degrees of freedom. The dashed black line shows the prior distribution, which is a $N(20, 20)$ truncated to be positive in the support $[3, 60]$. The solid blue line shows the posterior distribution obtained from 20,000 posterior draws.

of the 100 replications. When necessary, we reorder and flip the sign of the estimated shocks so they match the sign and ordering of the black squared line. Figure 8 indicates that our method reliably recovers the true dynamics behind each pseudo-dataset.

We find that our approach succeeds at uncovering the true impulse responses even when the model is misspecified. Figure 9 replicates Figure 8 under a scenario where the data-generating process features Laplace (rather than t -distributed) shocks, normalized to have unit standard deviation. As the figure shows, the estimated impulse responses remain accurate. This aligns with the remarks by Sims (2021) and Brunnermeier et al. (2021) (see their footnote 8) that if the true shocks are mutually independent, symmetric around zero, and fat-tailed (though not necessarily t -distributed), a t -based SVAR can still recover the model's true dynamics.

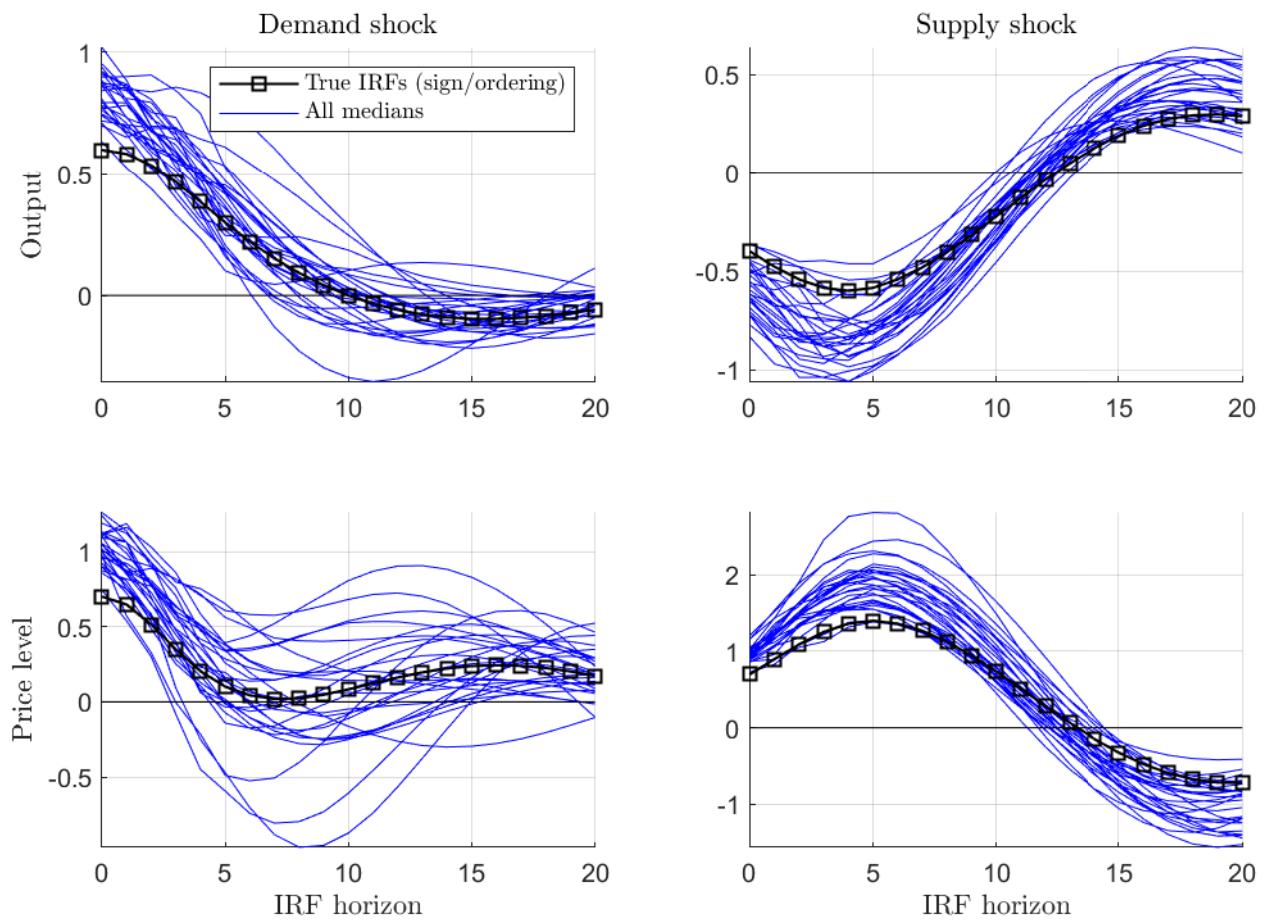
We refer to Appendix F in the Online Appendix for additional Monte Carlo results. There, we study what happens when we target the identity matrix in our normalization, highlighting that an informative normalization target is necessary. We also vary the prior on the degrees of freedom and use a flat prior on B^{-1} instead of B , showing that our results are robust. Finally, the appendix also shows the estimated posterior credible sets for each Monte Carlo replication.

Figure 8: Impulse response functions
Robustness across 100 samples



Note: See also Figure A-6 in the Online Appendix

Figure 9: Impulse response functions
 Robustness across 100 samples and DGP featuring Laplace shocks



Note: See also Figure A-7 in the Online Appendix

6 Conclusions

This paper introduces a new Gibbs sampler for structural VARs with t -distributed shocks. A simple reparametrization allows each conditional posterior to take a standard form, eliminating the computational bottlenecks that often arise in non-Gaussian models. Because our identification relies on non-Gaussianity alone, we avoid the strong theoretical restrictions that often accompany traditional approaches.

Applying this method to U.S. data shows that no single shock dominates GDP volatility. Instead, both demand and supply shocks play meaningful roles in explaining GDP. These findings suggest that understanding macroeconomic dynamics requires considering multiple sources of variation, rather than attributing most outcomes to a single “main” business-cycle shock.

References

- Andrade, P., Ferroni, F. and Melosi, L. (2023), Identification Using Higher-Order Moments Restrictions, Working Paper Series WP 2023-28, Federal Reserve Bank of Chicago.
- URL:** <https://ideas.repec.org/p/fip/fedhwp/96666.html>
- Angeletos, G.-M., Collard, F. and Dellas, H. (2020), ‘Business-Cycle Anatomy’, *American Economic Review* **110**(10), 3030–3070.
- URL:** <https://ideas.repec.org/a/aea/aecrev/v110y2020i10p3030-70.html>
- Anttonen, J. J., Lanne, M. and Luoto, J. (2024), ‘Statistically identified structural VAR model with potentially skewed and fat-tailed errors’, *Journal of Applied Econometrics* **39**(3), 422–437.
- Arias, J. E., Caldara, D. and Rubio-Ramirez, J. F. (2019), ‘The systematic component of monetary policy in SVARs: an agnostic identification procedure’, *Journal of Monetary Economics* **101**, 1–13.
- Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018), ‘Inference based on Structural Vector Autoregressions identified with sign and zero restrictions: Theory and applications’, *Econometrica* **86**(2), 685–720.
- Bańbura, M., Giannone, D. and Reichlin, L. (2010), ‘Large Bayesian VARs’, *Journal of Applied Econometrics* **25**, 71–92.
- Barnichon, R. and Matthes, C. (2018), ‘Functional approximation of impulse responses’, *Journal of Monetary Economics* **99**, 41–55.
- Baumeister, C. and Hamilton, J. D. (2015), ‘Sign restrictions, structural vector autoregressions, and useful prior information’, *Econometrica* **83**(5), 1963–1999.
- Braun, R. (2023), ‘The importance of supply and demand for oil prices: Evidence from non-gaussianity’, *Quantitative Economics* **14**(4), 1163–1198.
- Brunnermeier, M., Palia, D., Sastry, K. A. and Sims, C. A. (2021), ‘Feedbacks: financial markets and economic activity’, *American Economic Review* **111**(6), 1845–1879.
- Canova, F. (2007), *Methods for applied macroeconomic research*, Vol. 13, Princeton University Press.

- Canova, F. and De Nicoló, G. (2002), ‘Monetary disturbances matter for business fluctuations in the G-7’, *Journal of Monetary Economics* **49**(6), 1131–1159.
- Canova, F., Kocięcki, A. and Piffer, M. (2024), ‘Flexible prior beliefs on impulse responses in Bayesian vector autoregressive models’, *CEPR Discussion Paper No. 18992*.
- Chan, J. C. (2020), ‘Large Bayesian VARs: A flexible kronecker error covariance structure’, *Journal of Business & Economic Statistics* **38**(1), 68–79.
- Chan, J. C. C. (2022), ‘Asymmetric conjugate priors for large Bayesian VARs’, *Quantitative Economics* **13**(3), 1145–1169.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE1381>
- Chan, J. C. C., Koop, G. and Yu, X. (2024), ‘Large Order-Invariant Bayesian VARs with Stochastic Volatility’, *Journal of Business & Economic Statistics* **42**(2), 825–837.
URL: <https://ideas.repec.org/a/taf/jnlbes/v42y2024i2p825-837.html>
- Chiu, C.-W. J., Mumtaz, H. and Pinter, G. (2017), ‘Forecasting with VAR models: Fat tails and stochastic volatility’, *International Journal of Forecasting* **33**(4), 1124–1143.
- Clark, T. E. and Ravazzolo, F. (2015), ‘Macroeconomic forecasting performance under alternative specifications of time-varying volatility’, *Journal of Applied Econometrics* **30**(4), 551–575.
- Cogley, T. and Sargent, T. J. (2005), ‘Drifts and volatilities: monetary policies and outcomes in the post WWII US’, *Review of Economic Dynamics* **8**(2), 262–302.
- Comon, P. (1994), ‘Independent component analysis, a new concept?’, *Signal processing* **36**(3), 287–314.
- Cúrdia, V., Del Negro, M. and Greenwald, D. L. (2014), ‘Rare shocks, great recessions’, *Journal of Applied Econometrics* **29**(7), 1031–1052.
- Drautzburg, T. and Wright, J. H. (2023), ‘Refining set-identification in VARs through independence’, *Journal of Econometrics* **235**(2), 1827–1847.

Francis, N. and Kindberg-Hanlon, G. (2022), ‘Signing out confounding shocks in variance-maximizing identification methods’, *AEA Papers and Proceedings* **112**, 476–80.

URL: <https://www.aeaweb.org/articles?id=10.1257/pandp.20221046>

Frühwirth-Schnatter, S. (2001), ‘Markov chain monte carlo estimation of classical and dynamic switching and mixture models’, *Journal of the American Statistical Association* **96**(453), 194–209.

URL: <http://www.jstor.org/stable/2670359>

Geweke, J. (1993), ‘Bayesian treatment of the independent student-t linear model’, *Journal of Applied Econometrics* **8**(S1), S19–S40.

Geweke, J. (2007), ‘Interpretation and inference in mixture models: Simple mcmc works’, *Computational Statistics & Data Analysis* **51**(7), 3529–3550.

URL: <https://www.sciencedirect.com/science/article/pii/S0167947306004506>

Hamilton, J. D., Waggoner, D. F. and Zha, T. (2007), ‘Normalization in Econometrics’, *Econometric Reviews* **26**(2-4), 221–252.

URL: <https://ideas.repec.org/a/taf/emetrv/v26y2007i2-4p221-252.html>

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**(4), 736–754.

Hoesch, L., Lee, A. and Mesters, G. (2024), ‘Locally robust inference for non-gaussian SVAR models’, *Quantitative Economics* **15**(2), 523–570.

URL: <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE2274>

Inoue, A. and Kilian, L. (2020), ‘The role of the prior in estimating VAR models with sign restrictions’.

Jarociński, M. (2024), ‘Estimating the fed’s unconventional policy shocks’, *Journal of Monetary Economics* **144**, 103548.

Kadiyala, K. R. and Karlsson, S. (1997), ‘Numerical methods for estimation and inference in Bayesian VAR-models’, *Journal of Applied Econometrics* **12**(2), 99–132.

Karlsson, S. and Mazur, S. (2020), ‘Flexible fat-tailed vector autoregression’.

- Karlsson, S., Mazur, S. and Nguyen, H. (2023), ‘Vector autoregression models with skewness and heavy tails’, *Journal of Economic Dynamics and Control* **146**, 104580.
- Kilian, L. and Lütkepohl, H. (2018), *Structural Vector Autoregressive Analysis*, number 9781107196575 in ‘Cambridge Books’, Cambridge University Press.
- Kocięcki, A. (2010), ‘A prior for impulse responses in Bayesian structural VAR models’, *Journal of Business & Economic Statistics* **28**(1), 115–127.
- Kocięcki, A., Rubaszek, M. and Ca’Zorzi, M. (2012), ‘Bayesian analysis of recursive SVAR models with overidentifying restrictions’, *ECB Working Paper No. 1492*.
- Kydland, F. E. and Prescott, E. C. (1982), ‘Time to build and aggregate fluctuations’, *Econometrica* **50**(6), 1345–1370.
URL: <http://www.jstor.org/stable/1913386>
- Lanne, M. and Luoto, J. (2020), ‘Identification of economic shocks by inequality constraints in Bayesian structural vector autoregression’, *Oxford Bulletin of Economics and Statistics* **82**(2), 425–452.
- Lanne, M., Meitz, M. and Saikkonen, P. (2017), ‘Identification and estimation of non-gaussian structural vector autoregressions’, *Journal of Econometrics* **196**(2), 288–304.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, number 978-3-540-27752-1 in ‘Springer Books’, Springer.
URL: <https://ideas.repec.org/b/spr/sprbok/978-3-540-27752-1.html>
- Maxand, S. (2020), ‘Identification of independent structural shocks in the presence of multiple Gaussian components’, *Econometrics and Statistics* **16**, 55–68.
URL: <https://www.sciencedirect.com/science/article/pii/S2452306218300923>
- Mesters, G. and Zwiernik, P. (2024), ‘Non-independent component analysis’, *The Annals of Statistics* **52**(6), 2506 – 2528.
URL: <https://doi.org/10.1214/24-AOS2373>
- Prüser, J. (2024), ‘A large non-Gaussian structural VAR with application to monetary policy’.
URL: <https://arxiv.org/abs/2412.17598>

- Ritter, C. and Tanner, M. A. (1992), ‘Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler’, *Journal of the American Statistical Association* **87**(419), 861–868.
- Sims, C. A. (1980), ‘Macroeconomics and reality’, *Econometrica* **48**(1), 1–48.
URL: <http://www.jstor.org/stable/1912017>
- Sims, C. A. (2021), ‘SVAR identification through heteroskedasticity with misspecified regimes’, *Technical report* .
- Smets, F. and Wouters, R. (2007), ‘Shocks and frictions in US business cycles: A Bayesian DSGE approach’, *American Economic Review* **97**(3), 586–606.
- Tinbergen, J. (1939), *Statistical Testing of Business Cycle Theories: Part I: A Method and Its Application to Investment Activity*.
URL: <http://hdl.handle.net/1765/14936>
- Uhlig, H. (2004), What moves GNP?, Econometric Society 2004 North American Winter Meetings 636, Econometric Society.
URL: <https://ideas.repec.org/p/ecm/nawm04/636.html>
- Waggoner, D. F. and Zha, T. (2003a), ‘A Gibbs sampler for structural vector autoregressions’, *Journal of Economic Dynamics and Control* **28**(2), 349–366.
- Waggoner, D. F. and Zha, T. (2003b), ‘Likelihood preserving normalization in multiple equation models’, *Journal of Econometrics* **114**(2), 329–347.
- Wolf, C. K. (2020), ‘SVAR (mis)identification and the real effects of monetary policy shocks’, *American Economic Journal: Macroeconomics* **12**(4), 1–32.
- Wu, P. and Koop, G. (2023), ‘Fast, order-invariant Bayesian inference in VARs using the eigendecomposition of the error covariance matrix’.

Online Appendix for “A Non-Gaussian GDP Anatomy”

A	Derivations of the fourth moments in section 2 of the paper	A-2
B	The model	A-4
C	Derivations of the conditional posterior distributions	A-6
C.1	Conditional posterior $p(\phi Y, B, D, \mathbf{v})$	A-6
C.2	Conditional posterior $p(D Y, \phi, B, \mathbf{v})$	A-7
C.3	Conditional posterior $p(\mathbf{v} Y, \phi, B, D)$	A-8
C.4	Conditional posterior $p(B Y, \phi, D, \mathbf{v})$	A-8
D	Identification up to sign and ordering	A-12
E	Summary of the algorithm to sample from the posterior of a SVAR with t-distributed strutural shocks	A-17
F	Additional material on the simulation exercise	A-18
G	Additional material on the application to the US GDP	A-26

A Derivations of the fourth moments in section 2 of the paper

The bivariate illustrative model is given by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}, \quad (\text{A-1})$$

where

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{11} \end{pmatrix}, \quad (\text{A-2})$$

and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})'$ are independently t distributed structural shocks with degrees of freedom $\mathbf{v} = (v_1, v_2)'$ and variances normalized to 1. $E(\epsilon_{it}) = 0$ implies $E(y_{it}) = 0$, $i = 1, 2$. Then

$$E(y_{1t}^4) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})^4, \quad (\text{A-3})$$

$$= E(b_{11}^4\epsilon_{1t}^4 + 4b_{11}^3\epsilon_{1t}^3b_{12}\epsilon_{2t} + 6b_{11}^2\epsilon_{1t}^2b_{12}^2\epsilon_{2t}^2 + 4b_{11}\epsilon_{1t}b_{12}^3\epsilon_{2t}^3 + b_{12}^4\epsilon_{2t}^4), \quad (\text{A-4})$$

$$= b_{11}^4 E(\epsilon_{1t}^4) + 6b_{11}^2 b_{12}^2 + b_{12}^4 E(\epsilon_{2t}^4), \quad (\text{A-5})$$

with the last step deriving from the independence assumption between ϵ_{1t} and ϵ_{2t} , and $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = 1$. section 2 of the paper defines k_i the kurtosis of ϵ_{it} . Since ϵ_{it} has mean 0 and variance normalized to 1, it holds that

$$E(\epsilon_{it}^4) = \kappa_i = \frac{3(\nu_i - 2)}{\nu_i - 4}. \quad (\text{A-6})$$

Hence

$$E(y_{1t}^4) = b_{11}^4 \kappa_1 + 6b_{11}^2 b_{12}^2 + b_{12}^4 \kappa_2, \quad (\text{A-7})$$

while similar derivations give

$$E(y_{2t}^4) = b_{21}^4 \kappa_1 + 6b_{21}^2 b_{22}^2 + b_{22}^4 \kappa_2. \quad (\text{A-8})$$

Next:

$$E(y_{1t}^3 y_{2t}) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})^3(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t}), \quad (\text{A-9})$$

$$= E(b_{11}^3\epsilon_{1t}^3 + 3b_{11}^2\epsilon_{1t}^2b_{12}\epsilon_{2t} + 3b_{11}\epsilon_{1t}b_{12}^2\epsilon_{2t}^2 + b_{12}^3\epsilon_{2t}^3)(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t}), \quad (\text{A-10})$$

$$\begin{aligned} &= E(b_{11}^3\epsilon_{1t}^3b_{21}\epsilon_{1t} + 3b_{11}^2\epsilon_{1t}^2b_{12}\epsilon_{2t}b_{21}\epsilon_{1t} + 3b_{11}\epsilon_{1t}b_{12}^2\epsilon_{2t}^2b_{21}\epsilon_{1t} + b_{12}^3\epsilon_{2t}^3b_{21}\epsilon_{1t} + \\ &\quad + b_{11}^3\epsilon_{1t}^3b_{22}\epsilon_{2t} + 3b_{11}^2\epsilon_{1t}^2b_{12}\epsilon_{2t}b_{22}\epsilon_{2t} + 3b_{11}\epsilon_{1t}b_{12}^2\epsilon_{2t}^2b_{22}\epsilon_{2t} + b_{12}^3\epsilon_{2t}^3b_{22}\epsilon_{2t}), \end{aligned} \quad (\text{A-11})$$

$$= E(b_{11}^3\epsilon_{1t}^4b_{21} + 3b_{11}\epsilon_{1t}^2b_{12}^2\epsilon_{2t}^2b_{21} + 3b_{11}^2\epsilon_{1t}^2b_{12}\epsilon_{2t}^2b_{22} + b_{12}^3\epsilon_{2t}^4b_{22}), \quad (\text{A-12})$$

$$= b_{11}^3b_{21}\kappa_1 + 3b_{11}b_{12}^2b_{21} + 3b_{11}^2b_{12}b_{22} + b_{12}^3b_{22}\kappa_2. \quad (\text{A-13})$$

$$E(y_{1t}y_{2t}^3) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t})^3, \quad (\text{A-14})$$

$$= E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})(b_{21}^3\epsilon_{1t}^3 + 3b_{21}^2\epsilon_{1t}^2b_{22}\epsilon_{2t} + 3b_{21}\epsilon_{1t}b_{22}^2\epsilon_{2t}^2 + b_{22}^3\epsilon_{2t}^3), \quad (\text{A-15})$$

$$\begin{aligned} &= E(b_{11}\epsilon_{1t}b_{21}^3\epsilon_{1t}^3 + 3b_{11}\epsilon_{1t}b_{21}^2\epsilon_{1t}^2b_{22}\epsilon_{2t} + 3b_{11}\epsilon_{1t}b_{21}\epsilon_{1t}b_{22}^2\epsilon_{2t}^2 + b_{11}\epsilon_{1t}b_{22}^3\epsilon_{2t}^3 + \\ &\quad + b_{12}\epsilon_{2t}b_{21}^3\epsilon_{1t}^3 + 3b_{12}\epsilon_{2t}b_{21}^2\epsilon_{1t}^2b_{22}\epsilon_{2t} + 3b_{12}\epsilon_{2t}b_{21}\epsilon_{1t}b_{22}^2\epsilon_{2t}^2 + b_{12}\epsilon_{2t}b_{22}^3\epsilon_{2t}^3), \end{aligned} \quad (\text{A-16})$$

$$= E(b_{11}b_{21}^3\epsilon_{1t}^4 + 3b_{11}\epsilon_{1t}^2b_{21}b_{22}^2\epsilon_{2t}^2 + 3b_{12}\epsilon_{2t}^2b_{21}^2\epsilon_{1t}^2b_{22} + b_{12}b_{22}^3\epsilon_{2t}^4), \quad (\text{A-17})$$

$$= b_{11}b_{21}^3\kappa_1 + 3b_{11}b_{21}b_{22}^2 + 3b_{12}b_{21}^2b_{22} + b_{12}b_{22}^3\kappa_2, \quad (\text{A-18})$$

$$E(y_{1t}^2 y_{2t}^2) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})^2(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t})^2, \quad (\text{A-19})$$

$$= E(b_{11}^2\epsilon_{1t}^2 + 2b_{11}\epsilon_{1t}b_{12}\epsilon_{2t} + b_{12}^2\epsilon_{2t}^2)(b_{21}^2\epsilon_{1t}^2 + 2b_{21}\epsilon_{1t}b_{22}\epsilon_{2t} + b_{22}^2\epsilon_{2t}^2), \quad (\text{A-20})$$

$$\begin{aligned} &= E(b_{11}^2\epsilon_{1t}^2b_{21}^2\epsilon_{1t}^2 + b_{12}^2\epsilon_{2t}^2b_{21}^2\epsilon_{1t}^2 + \\ &\quad + 4b_{11}\epsilon_{1t}b_{12}\epsilon_{2t}b_{21}\epsilon_{1t}b_{22}\epsilon_{2t} + \\ &\quad + b_{11}^2\epsilon_{1t}^2b_{22}^2\epsilon_{2t}^2 + b_{12}^2\epsilon_{2t}^2b_{22}^2\epsilon_{2t}^2), \end{aligned} \quad (\text{A-21})$$

$$= b_{11}^2b_{21}^2\kappa_1 + b_{12}^2b_{21}^2 + 4b_{11}b_{12}b_{21}b_{22} + b_{11}^2b_{22}^2 + b_{12}^2b_{22}^2\kappa_2. \quad (\text{A-22})$$

B The model

The model is given by

$$\mathbf{y}_t = \sum_{l=1}^p \Pi_l \mathbf{y}_{t-l} + \mathbf{c} + U_t, \quad (\text{A-23})$$

$$= \Pi \mathbf{x}_t + U_t, \quad (\text{A-24})$$

$$U_t = B \boldsymbol{\epsilon}_t, \quad (\text{A-25})$$

$$p(\boldsymbol{\epsilon}_t | \boldsymbol{\sigma}, \mathbf{v}) = \prod_{i=1}^k p(\epsilon_{it} | \sigma_i, v_i), \quad (\text{A-26})$$

$$\epsilon_{it} \sim t(\sigma_i, v_i), \quad (\text{A-27})$$

$$p(\epsilon_{it} | \sigma_i, v_i) = \sigma_i^{-1} \cdot v_i^{-\frac{1}{2}} \cdot \frac{\Gamma\left(\frac{v_i+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{v_i}{2}\right)} \cdot \left(1 + \frac{\epsilon_{it}^2}{v_i \cdot \sigma_i^2}\right)^{-\frac{v_i+1}{2}}. \quad (\text{A-28})$$

The $k \times 1$ vector \mathbf{y}_t contains the endogenous variables of the model. The $m \times 1$ vector $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, 1)'$ contains the lagged variables and the constant term, with p the number of lags in the model and $m = k \cdot p + 1$. The structural shocks $\boldsymbol{\epsilon}_t$ are *i.i.d.* with zero mean. Individual components of $\boldsymbol{\epsilon}_t$, i.e. ϵ_{it} , are mutually independent and possess a univariate t distribution, possibly with different degrees of freedom. Following, for example, [Geweke \(1993\)](#), the probability density function of each shock is parametrized according to equation (A-28), with (σ_i, v_i) the shock-specific scale and degrees of freedom, and $\mathbf{v} = (v_1, \dots, v_i, \dots, v_k)'$, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_i, \dots, \sigma_k)'$. As in [Brunnermeier et al. \(2021\)](#) (see their footnote 11), we set

$$\sigma_i = \sqrt{\frac{v_i - 2}{v_i}}, \quad (\text{A-29})$$

which implies that the variance of each structural t -distributed shock is normalized to unity,

$$V(\epsilon_{it}) = 1. \quad (\text{A-30})$$

Under this normalization, the $k \times k$ matrix B captures the impact effect of a one standard deviation shocks.¹

¹This normalization implies that the set of observationally equivalent models differ only up to sign and permutation of the shocks, but not up to the scale of the shocks ([Lanne et al., 2017](#)). We exploit this feature in order to build the generalized LP normalization, see [subsection 3.4](#) of the paper and

As in [Geweke \(1993\)](#), we use the following, alternative specification of the model:

$$\mathbf{y}_t = \sum_{l=1}^p \Pi_l \mathbf{y}_{t-l} + \mathbf{c} + U_t, \quad (\text{A-31})$$

$$= \Pi \mathbf{x}_t + U_t, \quad (\text{A-32})$$

$$U_t = B \sqrt{D_t} \mathbf{e}_t, \quad (\text{A-33})$$

$$\mathbf{e}_t \sim N(\mathbf{0}, I), \quad (\text{A-34})$$

$$D_t = \text{diag}(d_{1t}, \dots, d_{it}, \dots, d_{kt}), \quad (\text{A-35})$$

$$p(d_{it}|h_{d,i}, r_{d,i}) = \frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}}, \quad (\text{A-36})$$

$$h_{d,i} = \frac{v_i}{2}, \quad r_{d,i} = \frac{v_i - 2}{2}. \quad (\text{A-37})$$

The k stochastic terms in the new specification, \mathbf{e}_t , are Gaussian with the identity covariance matrix. The latent variables d_{it} are treated as unknown parameter with inverse Gamma prior that is independent across i and t , parametrized according to [\(A-36\)](#), and $\sqrt{D_t} = \text{diag}(d_{1t}^{\frac{1}{2}}, \dots, d_{it}^{\frac{1}{2}}, \dots, d_{kt}^{\frac{1}{2}})$. It is assumed that D_t is independent of \mathbf{e}_s , for every t and s . The shape and rate parameters $(h_{d,i}, r_{d,i})$ of $p(d_{it}|h_{d,i}, r_{d,i})$ are set as in [\(A-37\)](#), which comply with the normalization $V(\epsilon_{it}) = 1$ made in the original model specification.

Rewrite the parametrization of the model as

$$\boldsymbol{\phi} = \text{vec}(\Pi), \quad (\text{A-38})$$

$$D = \text{diag}(D_1, \dots, D_t, \dots, D_T). \quad (\text{A-39})$$

The vector $\boldsymbol{\phi}$ is of dimensions $km \times 1$. The array D is sparse and of dimensions $kT \times kT$. The joint prior distribution for the alternative specification of the model is

$$p(\boldsymbol{\phi}, B, D, \mathbf{v}) = p(\boldsymbol{\phi}) \cdot p(B) \cdot p(D|\mathbf{v}) \cdot p(\mathbf{v}). \quad (\text{A-40})$$

Our approach requires the prior on $\boldsymbol{\phi}$ to be Normal. While we set $\boldsymbol{\phi}$ as a priori independent of (B, D, \mathbf{v}) , this modelling assumption can be removed. As discussed above, the prior $p(D|\mathbf{v})$ is given by [\(A-36\)](#), with hyperparameters calibrated according to [\(A-37\)](#). By contrast, the prior on (B, \mathbf{v}) can be more flexibly selected by the

[Appendix D](#).

researcher, as discussed below.

C Derivations of the conditional posterior distributions

Posterior sampling is achieved by means of a Gibbs sampler. The conditional posterior for the parameters (ϕ, D, v) are standard in the literature and are reported here for completeness. The conditional posterior of B is discussed in greater length and exploits the reparametrization discussed in the paper.

C.1 Conditional posterior $p(\phi|Y, B, D, v)$

The likelihood function of model (A-31)-(A-34) can be written as

$$p(Y|\phi, B, D) = (2\pi)^{-\frac{Tk}{2}} \cdot |B|^{-T} \cdot |D|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}} - W\phi)'((I_T \otimes B)D(I_T \otimes B'))^{-1}(\tilde{\mathbf{y}} - W\phi)}, \quad (\text{A-41})$$

with

$$Y = [\mathbf{y}_1, \dots, \mathbf{y}_t, \dots, \mathbf{y}_T], \quad (\text{A-42})$$

$$\tilde{\mathbf{y}} = \text{vec}(Y), \quad (\text{A-43})$$

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T], \quad (\text{A-44})$$

$$W = (X' \otimes I_k). \quad (\text{A-45})$$

Using (A-41), we can derive

$$p(\phi|Y, B, D, v) \propto p(\phi) \cdot p(Y|\phi, B, D), \quad (\text{A-46})$$

$$\propto e^{-\frac{1}{2}(\phi - \mu)'V^{-1}(\phi - \mu)} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}} - W\phi)'((I_T \otimes B)D(I_T \otimes B'))^{-1}(\tilde{\mathbf{y}} - W\phi)}, \quad (\text{A-47})$$

$$= e^{-\frac{1}{2}(\phi - \mu)'V^{-1}(\phi - \mu)} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}} - W\phi)' \Omega^{-1}(\tilde{\mathbf{y}} - W\phi)}, \quad (\text{A-48})$$

$$\propto e^{-\frac{1}{2}(\phi'V^{-1}\phi - 2\phi'V^{-1}\mu + \phi'W'\Omega^{-1}W\phi - 2\phi'W'\Omega^{-1}\tilde{\mathbf{y}})}, \quad (\text{A-49})$$

$$\boldsymbol{\phi}|Y, B, D, \mathbf{v} \sim N(\boldsymbol{\mu}^*, V^*), \quad (\text{A-50})$$

$$V^* = (V^{-1} + W'\Omega^{-1}W)^{-1}, \quad (\text{A-51})$$

$$\boldsymbol{\mu} = V^*[V^{-1}\boldsymbol{\mu} + W'\Omega^{-1}\tilde{\mathbf{y}}], \quad (\text{A-52})$$

$$\Omega = (I_T \otimes B)D(I_T \otimes B'). \quad (\text{A-53})$$

C.2 Conditional posterior $p(D|Y, \boldsymbol{\phi}, B, \mathbf{v})$

The conditional posterior for D is also standard. After defining

$$\mathbf{g}_t = B^{-1}[\mathbf{y}_t - \Pi \mathbf{x}_t], \quad (\text{A-54})$$

model (A-31)-(A-34) coincides with $\mathbf{g}_t \sim N(\mathbf{0}, D_t)$, hence

$$p(D|Y, \boldsymbol{\phi}, B, \mathbf{v}) = \prod_{t=1}^T \prod_{i=1}^k p(d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, v_i), \quad (\text{A-55})$$

$$p(d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, v_i) \propto p(d_{it}|h_{d,i}, r_{d,i}) \cdot p(\mathbf{g}_t|\Pi, B, \mathbf{v}, D_t), \quad (\text{A-56})$$

$$\propto \left[\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right] \cdot d_{it}^{-\frac{1}{2}} \cdot e^{-\frac{1}{2} \frac{g_{it}^2}{d_{it}}}, \quad (\text{A-57})$$

$$\propto d_{it}^{-(h_{d,i}+\frac{1}{2})-1} \cdot e^{-(r_{d,i} + \frac{g_{it}^2}{2}) \frac{1}{d_{it}}}, \quad (\text{A-58})$$

$$= d_{it}^{-h_{d,i}^*-1} \cdot e^{-r_{d,it}^* \frac{1}{d_{it}}}, \quad (\text{A-59})$$

$$d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, \mathbf{v} \sim i\Gamma(h_{d,i}^*, r_{d,it}^*), \quad (\text{A-60})$$

$$h_{d,i}^* = h_{d,i} + \frac{1}{2}, \quad (\text{A-61})$$

$$r_{d,it}^* = r_{d,i} + \frac{g_{it}^2}{2}. \quad (\text{A-62})$$

C.3 Conditional posterior $p(\mathbf{v}|Y, \phi, B, D)$

The conditional posterior for \mathbf{v} is

$$p(\mathbf{v}|Y, \Pi, B, D) \propto p(D|\mathbf{v}) \cdot p(\mathbf{v}) \quad (\text{A-63})$$

$$\propto \left[\prod_{t=1}^T \prod_{i=1}^k \frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right] \cdot p(\mathbf{v}), \quad (\text{A-64})$$

$$= \left[\prod_{i=1}^k \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \prod_{t=1}^T \left(d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right) \right] \cdot p(\mathbf{v}), \quad (\text{A-65})$$

$$= \left[\prod_{i=1}^k \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \left(\prod_{t=1}^T d_{it} \right)^{-h_{d,i}-1} e^{-r_{d,i} \cdot \left(\sum_{t=1}^T \frac{1}{d_{it}} \right)} \right] \cdot p(\mathbf{v}). \quad (\text{A-66})$$

We assume prior independence across \mathbf{v} , i.e. $p(\mathbf{v}) = \prod_{i=1}^k p(v_i)$. Since $(h_{d,i}, r_{d,i})$ only depend on entry i of \mathbf{v} , prior independence implies independence in conditional posterior:

$$p(\mathbf{v}|Y, \Pi, B, D) = \prod_{i=1}^k p(v_i|Y, \Pi, B, D), \quad (\text{A-67})$$

$$p(v_i|Y, \Pi, B, D) \propto \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \left(\prod_{t=1}^T d_{it} \right)^{-h_{d,i}-1} e^{-r_{d,i} \cdot \left(\sum_{t=1}^T \frac{1}{d_{it}} \right)} \cdot p(v_i). \quad (\text{A-68})$$

Last, we use a Griddy-Gibbs sampler and discretize the support for v_i .

C.4 Conditional posterior $p(B|Y, \phi, D, \mathbf{v})$

Posterior sampling on B is achieved indirectly via sampling on the parametrization $A = B^{-1}$. After defining

$$\mathbf{z}_t = \mathbf{y}_t - \Pi \mathbf{x}_t, \quad (\text{A-69})$$

model (A-31)-(A-34) coincides with $\mathbf{z}_t \sim N(\mathbf{0}, A^{-1} D_t A^{-1}')$, hence the likelihood function is

$$p(Z|\Pi, A, D) = (2\pi)^{-\frac{T_k}{2}} \cdot |A|^T \cdot \left(\prod_{t=1}^T |D_t|^{-\frac{1}{2}} \right) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' A' D_t^{-1} A \mathbf{z}_t}, \quad (\text{A-70})$$

with $Z = [\mathbf{z}_1, \dots, \mathbf{z}_t, \dots, \mathbf{z}_T]$. Note that in fact $|A| := |\det(A)|$

We rewrite A as

$$A = \Lambda LU. \quad (\text{A-71})$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_i, \dots, \lambda_k)$ is a $k \times k$ diagonal matrix, L and U are lower and upper triangular matrices, respectively, of dimension $k \times k$, both with unit diagonal entries. Assuming that all λ_i are nonzero, the underlying decomposition of A exists and is unique. The free entries of (L, U) are (L, U) , respectively. We use the notation

$$\text{vec}(L) = \mathbf{s} + S_L L, \quad (\text{A-72})$$

$$\text{vec}(U) = \mathbf{s} + S_U U, \quad (\text{A-73})$$

with (L, U) of dimension $k(k - 1)/2 \times 1$, (S_L, S_U) of dimension $k^2 \times k(k - 1)/2$, \mathbf{s} of dimension $k^2 \times 1$, and (S_L, S_U, \mathbf{s}) having zero or one entries as appropriate. Note that because (L, U) are triangular matrices with unit diagonal entries, the determinant of A evaluated in the parametrization (Λ, L, U) is only a function of Λ , namely

$$|A| = \prod_{i=1}^k |\lambda_i|. \quad (\text{A-74})$$

Define $p_A(A)$ as the prior distribution on A . In order to write it down explicitly, one should take a stand on the chosen parametrization of the SVAR model i.e. whether A or B is our basic parameter. In either case our baseline prior setup assumes flat prior. Hence two cases are of interest for us:

$$p_A(A) \propto 1, \quad (\text{A-75})$$

$$p_A(A) \propto |A|^{-2k}. \quad (\text{A-76})$$

(A-75) is the case in which the researcher expresses an uninformative prior directly on A (hence implicitly the SVAR uses A as its basic parameter). (A-76) is the case in which the researcher expresses an uninformative prior on B , with $|A|^{-2k}$ the Jacobian term in the transformation from B to A , assuming the entries of B are functionally unconstrained (Kocięcki, 2010). When evaluated in the (Λ, L, U) , (A-76) is only a function of λ_i 's via equation (A-74).

Combining (A-70)-(A-75)-(A-76) gives the conditional posterior for A :

$$p(A|Y, \phi, D, \mathbf{v}) \propto p_A(A) \cdot |A|^T \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' A' D_t^{-1} A \mathbf{z}_t}. \quad (\text{A-77})$$

In order to proceed further we need the Jacobian of the transformation from A to (Λ, L, U) . It may be shown that:

$$J(A \rightarrow \Lambda, L, U) = \prod_{i=1}^k |\lambda_i|^{k-1}. \quad (\text{A-78})$$

Hence, (A-77) implies the following conditional posterior jointly for (Λ, L, U) :

$$\begin{aligned} p(\Lambda, L, U | Y, \phi, D, v) &\propto \left(\prod_{i=1}^k |\lambda_i|^{k-1} \right) \cdot p_A(\Lambda L U) \cdot \left(\prod_{i=1}^k |\lambda_i|^T \right) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \\ & \quad (\text{A-79}) \end{aligned}$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}. \quad (\text{A-80})$$

Our approach to sample from $p(\Lambda, L, U | Y, \phi, D, v)$ consists in showing that the conditional posteriors of Λ , L and U all have a common form. Starting with L , under either (A-75) or (A-76) we can derive

$$p(L | Y, \phi, D, v, \Lambda, U) \propto e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{A-81})$$

$$= e^{-\frac{1}{2} \text{vec}(L)' \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}_t' U' \otimes \Lambda D_t^{-1} \Lambda) \text{vec}(L)}, \quad (\text{A-82})$$

$$= e^{-\frac{1}{2} \text{vec}(L)' \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}_t' U' \otimes \Lambda^2 D_t^{-1}) \text{vec}(L)}, \quad (\text{A-83})$$

$$p(L | Y, \phi, D, v, \Lambda, U) \propto e^{-\frac{1}{2} (\mathbf{s} + S_L L)' W_L (\mathbf{s} + S_L L)}, \quad (\text{A-84})$$

$$\propto e^{-\frac{1}{2} (L' S_L' W_L S_L L + 2L' S_L' W_L \mathbf{s})}, \quad (\text{A-85})$$

hence,

$$L | Y, \phi, D, v, \Lambda, U \sim N(\boldsymbol{\mu}_L^*, V_L^*), \quad (\text{A-86})$$

$$V_L^* = (S_L' W_L S_L)^{-1} \quad (\text{A-87})$$

$$\boldsymbol{\mu}_L^* = -V_L^* S_L' W_L \mathbf{s}, \quad (\text{A-88})$$

$$W_L = \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}_t' U' \otimes \Lambda^2 D_t^{-1}), \quad (\text{A-89})$$

The derivations for U are similar. Under either (A-75) or (A-76) we can derive

$$p(U|Y, \phi, D, v, \Lambda, L) \propto e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{A-90})$$

$$= e^{-\frac{1}{2} \text{vec}(U)' \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}'_t \otimes L' \Lambda D_t^{-1} \Lambda L) \text{vec}(U)}, \quad (\text{A-91})$$

$$= e^{-\frac{1}{2} \text{vec}(U)' \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}'_t \otimes L' \Lambda^2 D_t^{-1} L) \text{vec}(U)}, \quad (\text{A-92})$$

$$p(U|Y, \phi, D, v, \Lambda, U) \propto e^{-\frac{1}{2} (\mathbf{s} + S_U U)' W_U (\mathbf{s} + S_U U)}, \quad (\text{A-93})$$

$$\propto e^{-\frac{1}{2} (U' S'_U W_U S_U U + 2U' S'_U W_U \mathbf{s})}, \quad (\text{A-94})$$

hence

$$U|Y, \phi, D, v, \Lambda, L \sim N(\boldsymbol{\mu}_U^*, V_U^*) \quad (\text{A-95})$$

$$V_U^* = (S'_U W_U S_U)^{-1}, \quad (\text{A-96})$$

$$\boldsymbol{\mu}_U^* = -V_U^* S'_U W_U \mathbf{s}, \quad (\text{A-97})$$

$$W_U = \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}'_t \otimes L' \Lambda^2 D_t^{-1} L). \quad (\text{A-98})$$

It remains to derive the conditional posterior for Λ , which is

$$p(\Lambda|Y, \phi, D, v, L, U) \propto \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{A-99})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' D_t^{-0.5} \Lambda^2 D_t^{-0.5} L U \mathbf{z}_t}, \quad (\text{A-100})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{i=1}^k \sum_{t=1}^T c'_t \Lambda^2 c_t}, \quad (\text{A-101})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{i=1}^k \sum_{t=1}^T c_{it}^2 \lambda_i^2}, \quad (\text{A-102})$$

$$= \prod_{i=1}^k |\lambda_i|^{T+k+\alpha-1} \cdot e^{-\frac{1}{2} \lambda_i^2 \sum_{t=1}^T c_{it}^2}, \quad (\text{A-103})$$

where $\alpha = 0$ if we adopt (A-75), or $\alpha = -2k$ if we use (A-76).

Let us define $x_i = \lambda_i^2$. Although this transformation is not 1-1 since λ_i may be

both positive and negative, from standard probability we know that if λ_i has pdf $p(\lambda_i)$ then the pdf of x_i is $g(x_i) = \frac{1}{2}x_i^{-\frac{1}{2}}p(\sqrt{x_i}) + \frac{1}{2}x_i^{-\frac{1}{2}}p(-\sqrt{x_i})$, for $x_i > 0$. Since in our case $p(\sqrt{x_i}) = p(-\sqrt{x_i})$, hence $g(x_i) = x_i^{-\frac{1}{2}}p(\sqrt{x_i})$, it follows that

$$p(x_1, x_2, \dots, x_k | Y, \phi, D, v, L, U) \propto \prod_{i=1}^k x_i^{-\frac{1}{2}} |x_i^{\frac{1}{2}}|^{T+k+\alpha-1} \cdot e^{-\frac{1}{2}x_i \sum_{t=1}^T c_{it}^2}, \quad (\text{A-104})$$

$$= \prod_{i=1}^k x_i^{\frac{T+k+\alpha}{2}-1} \cdot e^{-\frac{1}{2}x_i \sum_{t=1}^T c_{it}^2}, \quad (\text{A-105})$$

hence

$$x_i | Y, \phi, D, v, L, U \sim \Gamma(h_{\lambda,i}^*, r_{\lambda,i}^*), \quad (\text{A-106})$$

$$r_{\lambda,i}^* = \frac{\sum_{t=1}^T c_{it}^2}{2}, \quad (\text{A-107})$$

$$\mathbf{c}_t = D_t^{-0.5} L U \mathbf{z}_t, \quad (\text{A-108})$$

where it holds

$$h_{\lambda,i}^* = \frac{T+k}{2}, \quad (\text{A-109})$$

if prior (A-75) is used, and

$$h_{\lambda,i}^* = \frac{T-k}{2}, \quad (\text{A-110})$$

if prior (A-76) is used. Note that we are using the following shape-rate parametrization of the Gamma distribution

$$p(x|h, r) = \frac{r^h}{\Gamma(h)} \cdot x^{h-1} e^{-r \cdot x}. \quad (\text{A-111})$$

Having drawn x_i we set $\lambda_i = \sqrt{x_i}$ with probability $\frac{1}{2}$, or $\lambda_i = -\sqrt{x_i}$ with probability $\frac{1}{2}$, see [Waggoner and Zha \(2003a\)](#), p. 357, for analogous treatment.

D Identification up to sign and ordering

The non-Gaussian, statistically independent nature of the structural shocks combined with the normalization of the variance to unity implies that the model is identified up to sign and permutation of the shocks ([Lanne et al., 2017](#)). In order to avoid the

associated multimodality of the posterior of A and B , we have to be sure that we uncover the posterior uncertainty surrounding only one (possibly arbitrarily chosen) mode. We achieve this via a normalization rule that builds on [Waggoner and Zha \(2003b\)](#) (WZ hereafter). WZ work with the Gaussian SVAR model with variance of the structural shocks normalized to 1, and develop the Likelihood Preserving (LP) normalization that addresses the indeterminacy of the model up to the sign of the shocks. We extend their method to address indeterminacy up to sign and ordering (or permutation) of the shocks in a SVAR with independent t -distributed shocks. We refer to this as the generalized LP normalization. In this section we first define the generalized LP normalization and then relate it to the original specification by WZ. Last, we show that existing combinatorial optimization techniques allow for a very fast computation of the matrix needed to operationalize the normalization rule. This makes the normalization rule feasible also for large models.

Let P denote the permutation matrix and P_s the signed permutation matrix. Our criterion to choose P_s is

$$\min_{P_s} \text{tr}\{(BP_s - \hat{B})' \hat{A}' \hat{A} (BP_s - \hat{B})\}, \quad (\text{A-112})$$

where $\hat{\theta}$ denotes the Maximum Likelihood (ML) estimator of θ .² By multiplying matrices in (A-112) and using the fact that $\hat{B} = \hat{A}^{-1}$, one can show that (A-112) is equivalent to one of the following maximization problems:

$$\max_{P_s} \text{tr}\{P_s \hat{A} B\} = \max_{P_s} \text{tr}\{\hat{A} B P_s\} = \max_{P_s} \text{tr}\{\hat{A} A^{-1} P_s\} = \max_{P_s} \text{tr}\{P_s' A'^{-1} \hat{A}'\}. \quad (\text{A-113})$$

In order to appreciate the similarity of our criterion to the original LP normalization, we show that if P_s were the diagonal matrix with ± 1 on the diagonal, then the solution to (A-112) would be exactly the LP normalization. Since WZ use the SVAR with its transposed form, their (A, \hat{A}) are our (A', \hat{A}') . Provided that P_s is a diagonal matrix with ± 1 on the diagonal, the last formula in (A-113) amounts to multiplying each diagonal element of $A'^{-1} \hat{A}'$ by -1 if it is negative and by 1 , if it is positive. Following WZ's notation, let e_i denote the i -th column of I_k . Then the i -th diagonal element

²Note that using the notation in Proposition 4 in WZ, the minimizing function can be written as $\|BP_s - \hat{B}\|_{\hat{\Omega}^{-1}}$, where $\hat{\Omega} = \hat{A}^{-1} \hat{A}'^{-1} = \hat{B} \hat{B}'$ is the ML estimate of the covariance of the reduced form disturbances. The weighting function $\hat{\Omega}^{-1}$ is important for the distance to be invariant under changing the measurement units in the data. In particular, the solution to (A-112) remains the same if instead of the original data y_t we use Hy_t , where H is any nonsingular matrix - see Proposition 5 in WZ.

of $A^{-1}\hat{A}$ may be written as $e'_i A^{-1}\hat{A}e_i = e'_i A^{-1}\hat{a}_i$, where \hat{a}_i denotes the i -th column of \hat{A} . An A draw is LP normalized if $e'_i A^{-1}\hat{a}_i > 0$ for each $i = 1, \dots, k$. Multiplication of the i -th diagonal element of $A^{-1}\hat{A}$ by -1 is equivalent to multiplication of the i -th column of A by -1 since $-1 \cdot e'_i A^{-1}\hat{a}_i = e'_i I_k^* A^{-1}\hat{a}_i = e'_i (AI_k^*)^{-1}\hat{a}_i$, where I_k^* is the $k \times k$ identity matrix except its i -th diagonal element is set to -1 . This is exactly the LP normalization for Gaussian SVARs, rewritten in WZ's notation.

In our general setup, we need to compute the signed permutation matrix P_s that solves one of the equivalent problems in (A-113). To this end let us denote $G = \hat{A}A^{-1}$ and $g_{i,j}$ the (i,j) -th element of G . We focus on $\max_{P_s} \text{tr}\{P_s G\}$. The problem is to choose the permutation of rows of G (possibly multiplied by -1) such that $\text{tr}\{P_s G\}$ attains its maximum. At the maximum $\text{tr}\{P_s G\} = p_1 g_{\pi(1),1} + p_2 g_{\pi(2),2} + \dots + p_k g_{\pi(k),k}$, where $\pi(i)$ denotes permutation of the row index and each $p_i = \pm 1$. We first note that at the maximum each term $p_i g_{\pi(i),i}$ must be nonnegative. To realize it assume by contradiction that at least one $p_i g_{\pi(i),i} < 0$. Then by setting $p_i^* = -1 \cdot p_i$ we have $p_1 g_{\pi(1),1} + \dots + p_i g_{\pi(i),i} + \dots + p_k g_{\pi(k),k} < p_1 g_{\pi(1),1} + \dots + p_i^* g_{\pi(i),i} + \dots + p_k g_{\pi(k),k}$, i.e. contradiction. Hence at the maximum:

$$\begin{aligned} p_1 g_{\pi(1),1} + p_2 g_{\pi(2),2} + \dots + p_k g_{\pi(k),k} &= |p_1 g_{\pi(1),1}| + |p_2 g_{\pi(2),2}| + \dots + |p_k g_{\pi(k),k}|, \\ &= |p_1| |g_{\pi(1),1}| + |p_2| |g_{\pi(2),2}| + \dots + |p_k| |g_{\pi(k),k}|, \\ &= |g_{\pi(1),1}| + |g_{\pi(2),2}| + \dots + |g_{\pi(k),k}|. \end{aligned} \quad (\text{A-114})$$

This suggests the modified problem:

$$\max_P \text{tr}\{P \cdot |G|\} = \max_P \text{tr}\{P \cdot |\hat{A}A^{-1}|\}, \quad (\text{A-115})$$

where P is the usual permutation matrix and $|G|$ means absolute values taken element-wise for all entries in G matrix. At the maximum $\text{tr}\{P \cdot |G|\} = |g_{\pi(1),1}| + |g_{\pi(2),2}| + \dots + |g_{\pi(k),k}|$, hence though the space of permutation matrices is a subset of the space of signed permutation matrices, the maximum of the original optimization problem (A-113) is attained by the modified (i.e. constrained) optimization problem (A-115). When working with (A-115), finding the signed permutation corresponding to this maximum only requires picking $p_i = 1$ or $p_i = -1$ such that each $p_i g_{\pi(i),i}$ is positive (we omit considering the measure zero event such that $g_{\pi(i),i} = 0$ for some i , which does not appear in practice).

In practice, our problem

$$\max_P \text{tr}\{P \cdot |G|\} = \min_P \text{tr}\{P \cdot -|G|\}, \quad (\text{A-116})$$

could be addressed by evaluating the objective function of (A-112)-(A-113)-(A-115)-(A-116) for every P_s . This would be computationally infeasible for even medium-scaled models, since there exist $k!$ permutation matrices, and this step is required for every draw in the posterior sampler. However, the formulation in (A-116) shows that the maximization problem is a linear assignment problem from combinatorial optimization. The classic method to solve it is the so-called Hungarian algorithm and its modern refinements, which are computationally really fast. Trying many versions of this algorithm it turned out that the built-in MATLAB function ‘matchpairs’ is the fastest. To appreciate its computational efficiency, take a generic number of variables k . Set $\hat{A} = 10 \cdot I_k$, draw the entries of B from independent $N(0, 1)$, compute $G = \hat{A}B$, evaluate the time it takes to solve (A-116), and repeat. The following summarizes, on average over 100,000 repetitions, how long it took to solve (A-116): 0.000056 seconds for $k = 5$; 0.00007 seconds for $k = 10$; 0.0001 seconds for $k = 20$ and 0.0013 seconds for $k = 100$ (computation done on an Intel Xeon E5-1603 v4 and 2.80 GHz). As the illustration shows, the execution time of this optimization technique does not increase substantially even for large k , hence making the normalization rule practical even for large models.

With our method, once we find the permutation matrix P that solves (A-116), we have to consider the diagonal elements in $PG = P\hat{A}A^{-1} = P\hat{A}B$. If the i -th diagonal element in $P\hat{A}B$ is negative we change the 1 in the i -th row of P to -1 , otherwise we do nothing. Doing so we accomplish the task of finding the *signed* permutation matrix that solves (A-112). This completes the computation of P_s , which is required at every iteration of the posterior sampler.

The following algorithm summarizes the steps required for the generalized LP normalization:

Algorithm 1: generalized LP normalization:

Before starting the sampler compute the ML estimator \hat{A} . Then apply the following steps at each iteration of the sampler, after drawing (A, B) :

1. solve $\min_P \text{tr}\{P \cdot -|\hat{A}B|\}$ using a version of the Hungarian algorithm, where P is the usual permutation matrix and $|\hat{A}B|$ stands for the

matrix of absolute values of all entries in $\hat{A}B$ (taken element-wise);

2. if the i -th diagonal element in $P\hat{A}B$ is negative, change the 1 in the i -th row of P to -1 , otherwise do nothing. As a result, we get P_s that solves (A-112);
3. replace B with BP_s and A with $P'_s A$.

We stress that this normalization rule can be applied to any non-Gaussian SVAR, and not just to SVAR models with t -distributed shocks.

Recently, [Jarociński \(2024\)](#) proposed an alternative but similar rule to normalize the draws of a SVAR model with independent t -distributed structural shocks. He used the Gaussian approximation to the likelihood function of A as a criterion. In particular, using our notation, his method amounts to solving

$$\min_{P_s} (\text{vec}(A'P_s) - \text{vec}(\hat{A}'))'\hat{V}^{-1}(\text{vec}(A'P_s) - \text{vec}(\hat{A}')), \quad (\text{A-117})$$

where \hat{V} is the asymptotic variance of $\text{vec}(A')$ i.e. the corresponding block of the inverse of the (minus) Hessian of the likelihood evaluated at the mode. For better comparison let us write our criterion (A-112) as

$$\min_{P_s} [(\text{vec}(BP_s) - \text{vec}(\hat{B}))'(I_k \otimes \hat{\Omega})^{-1}(\text{vec}(BP_s) - \text{vec}(\hat{B}))], \quad (\text{A-118})$$

where $\hat{\Omega} = \hat{A}^{-1}\hat{A}'^{-1} = \hat{B}\hat{B}'$. Hence one difference between our method and his is that we normalize B draws, whereas [Jarociński \(2024\)](#) normalizes A draws. He then assumes large sample approximation of the covariance of $\text{vec}(A')$ as the weighting function, whereas we assume block diagonal covariance for $\text{vec}(B)$ with the same block $\hat{\Omega}$ (which however follows directly from the LP normalization approach). However, the main difference lies in how we solve the underlying minimization problem. We use the highly efficient Hungarian algorithm, whereas [Jarociński \(2024\)](#) evaluates all $k!$ permutation matrices, see Algorithm 2 in his Online Appendix. As documented above, the case of $k = 20$ requires about 0.0001 seconds to find the optimal permutation matrix. Using the approach by [Jarociński \(2024\)](#) requires computing $20! = 2.432902 \cdot 10^{18}$ permutation matrices to find the one that solves (A-117).

E Summary of the algorithm to sample from the posterior of a SVAR with t -distributed structural shocks

All in all, our Gibbs sampling method for sampling from the joint posterior distribution of a SVAR model with independent, t -distributed structural shocks can be summarized as follows:

Algorithm 2: Gibbs sampler for SVAR models with independent t -distributed structural shocks:

0. in a preliminary step to the sampler, estimate a target matrix \hat{B} using a numerical maximum likelihood estimator, and compute $\hat{A} = \hat{B}^{-1}$.³ Then, at each iteration of the sampler:
 1. draw ϕ from the Normal conditional posterior from equation (A-76);
 2. draw L from the Normal conditional posterior from equation (A-86);
 3. draw U from the Normal conditional posterior from equation (A-95);
 4. draw \boldsymbol{x} from the Gamma conditional posterior from equation (A-106);
 5. compute Λ by setting each i, i entry as $\pm x_i$, with equal probability;
 6. compute \bar{B} associated with (Λ, L, U) using equation (A-68) and $B = A^{-1}$;
 7. set $B = \bar{B}P_s$, with P_s computed using the generalized LP normalization from Algorithm 1, [Appendix D](#);
 8. draw D from the inverse Gamma conditional posterior from equation (A-60);
 9. draw \boldsymbol{v} from the discretized conditional posterior from equation (A-68);
 10. repeat from step 1.

Note that placing the normalization step 7. before drawing (D, \boldsymbol{v}) ensures that the ordering of the structural shocks is consistent with the ordering of the degrees of

³To estimate the preliminary target matrix \hat{B} we found it convenient to use the 3-step maximum likelihood estimator suggested by [Lanne et al. \(2017\)](#). This makes the procedure quite fast also for our ten variable application from [section 4](#) of the paper.

freedom. We note, in passing, that the stored values of B are associated with a matrix A that may not admit a decomposition via equation (A-68). This fact is without loss of generality, as the decomposition is only employed as an operational procedure to develop a Gibbs sampler.

F Additional material on the simulation exercise

The parameter values of the data generating process are

$$B = \begin{pmatrix} 0.60 & 0.40 \\ 0.70 & -0.70 \end{pmatrix} \quad (\text{A-119})$$

$$\mathbf{v} = \begin{pmatrix} 6 \\ 6 \end{pmatrix} \quad (\text{A-120})$$

$$\Pi_1 = \begin{pmatrix} 1.0612 & -0.0759 \\ -0.2502 & 1.1404 \end{pmatrix} \quad (\text{A-121})$$

$$\Pi_2 = \begin{pmatrix} -0.0660 & 0.0093 \\ -0.0253 & -0.0905 \end{pmatrix} \quad (\text{A-122})$$

$$\Pi_3 = \begin{pmatrix} -0.0641 & 0.0109 \\ 0.0286 & -0.0655 \end{pmatrix} \quad (\text{A-123})$$

$$\Pi_4 = \begin{pmatrix} -0.0530 & 0.0119 \\ 0.0639 & -0.0434 \end{pmatrix} \quad (\text{A-124})$$

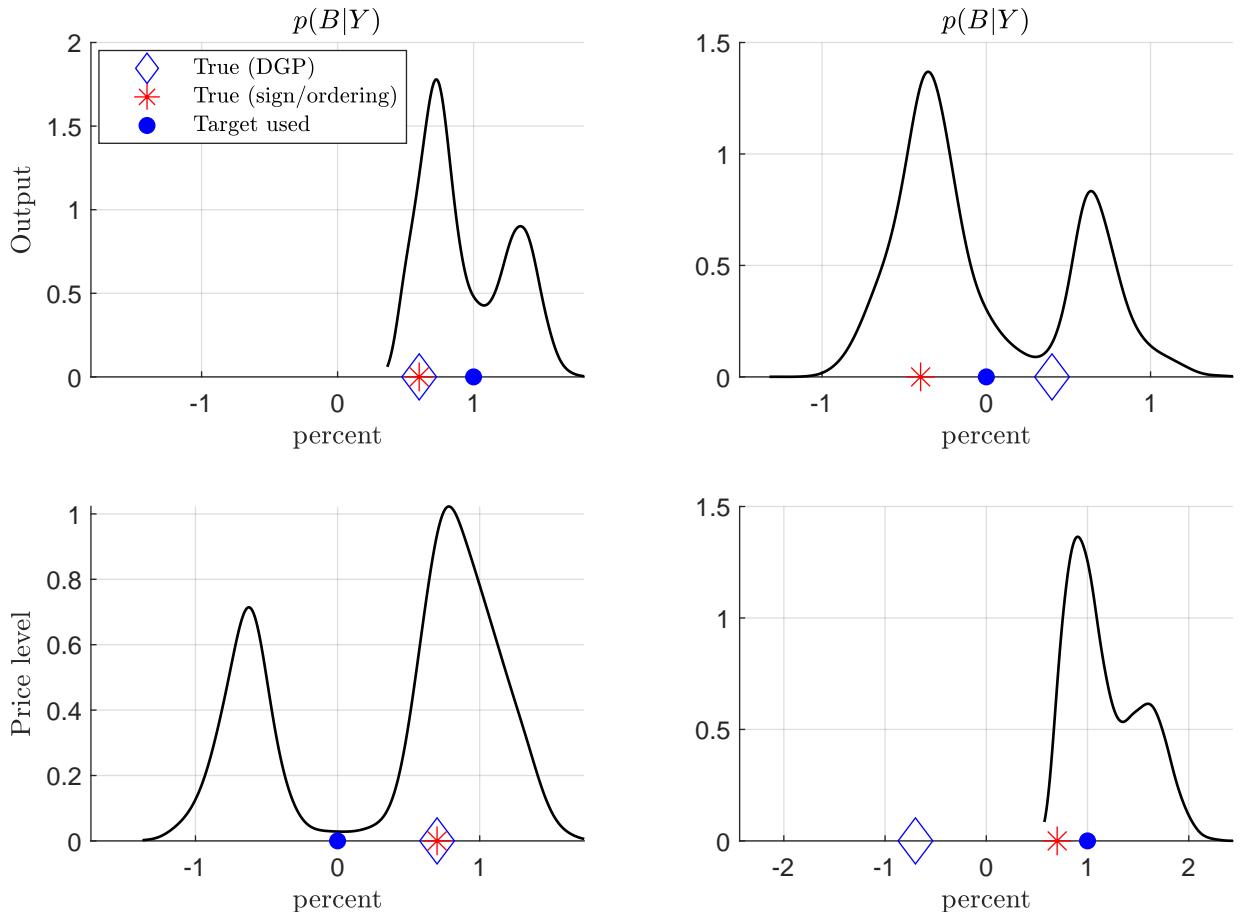
$$\Pi_5 = \begin{pmatrix} -0.0355 & 0.0113 \\ 0.0660 & -0.0304 \end{pmatrix} \quad (\text{A-125})$$

$$\Pi_6 = \begin{pmatrix} -0.0165 & 0.0084 \\ 0.0425 & -0.0230 \end{pmatrix} \quad (\text{A-126})$$

The target matrix used for the generalized LP normalization was estimated to

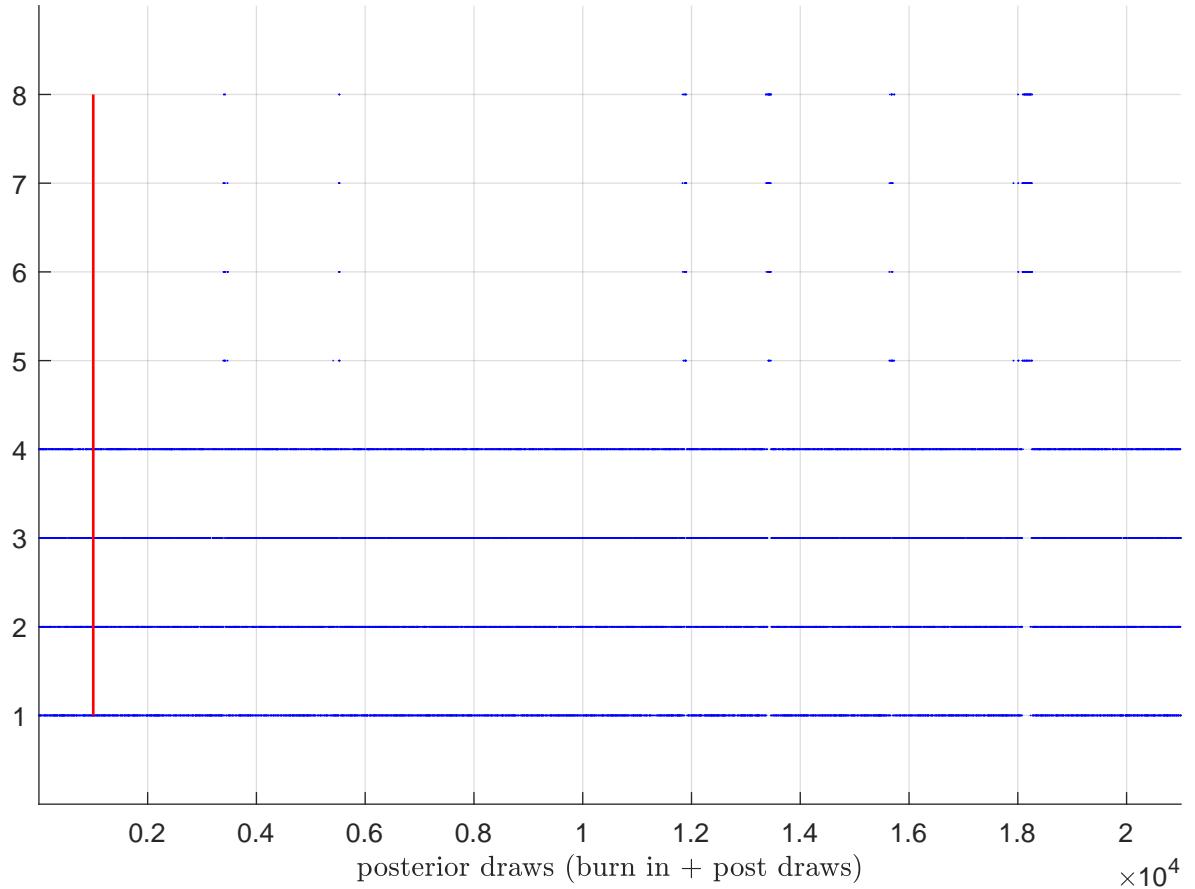
$$B_{target} = \begin{pmatrix} 0.7015 & -0.3279 \\ 0.7271 & 0.7236 \end{pmatrix} \quad (\text{A-127})$$

Figure A-1: Impact effect of the shocks (B) when the normalization targets an arbitrary matrix: $\hat{B} = I_2$



Note: The blue diamond indicates B_{true} . The blue dots indicate the target matrix \hat{B} used for the normalization. The red star indicates the sign/permutation of B_{true} that is the closest to \hat{B} . The continuous line shows the marginal posterior of the entries of B from 20,000 posterior draws when applying the generalized LP normalization.

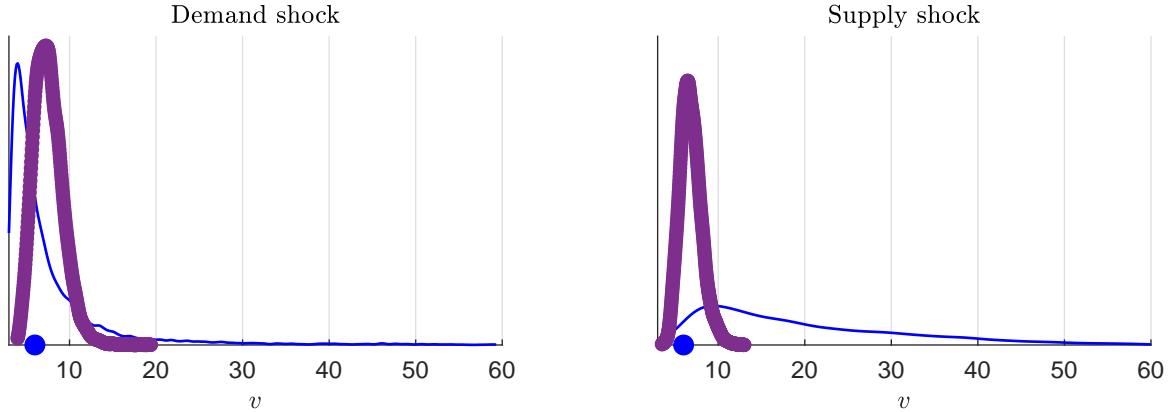
Figure A-2: Generalized LP normalization



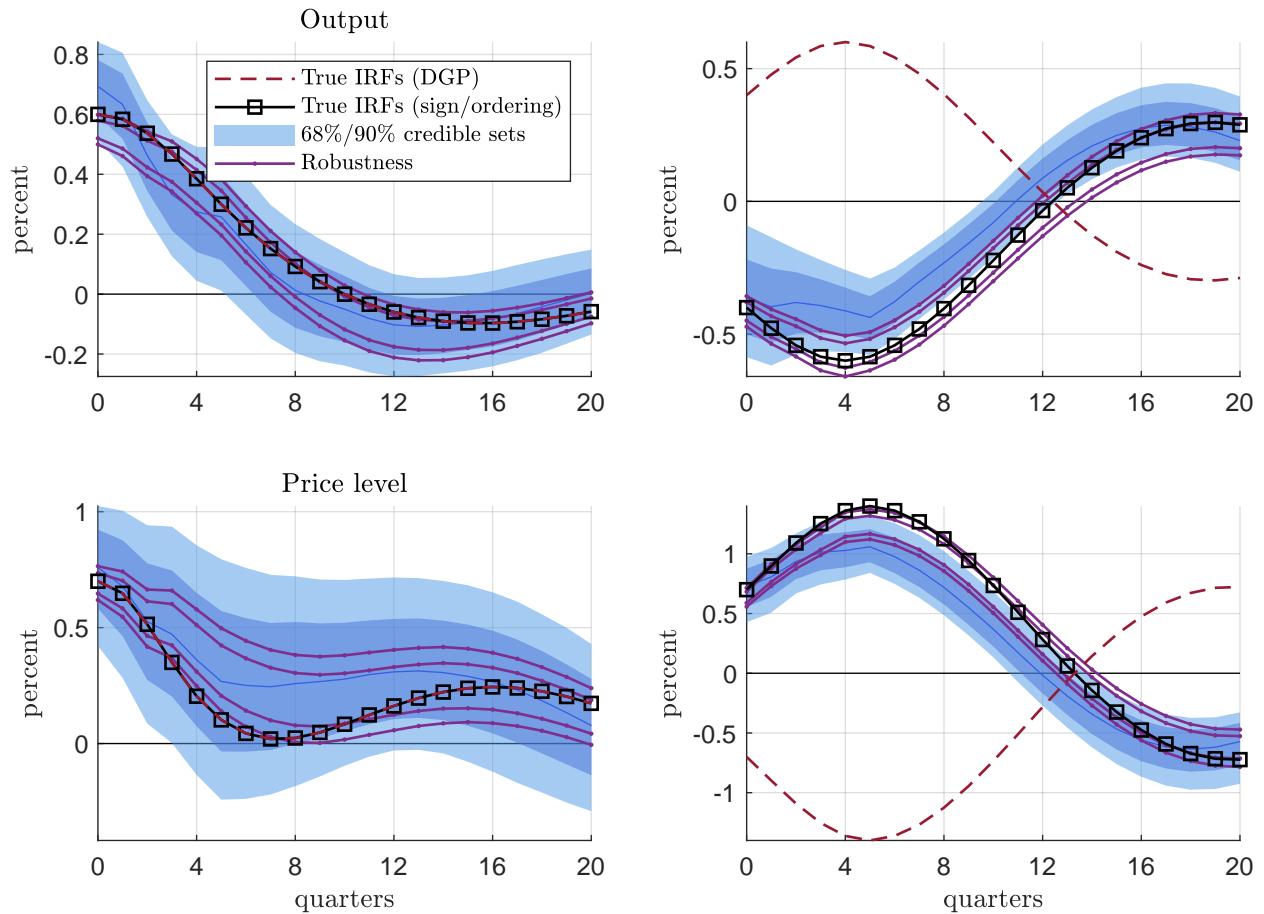
Note: The vertical red line shows when the burn-in draws end. For each of the 21,000 posterior draws, the figure indicates if the generalized LP normalization permutes the ordering of the columns of B (vertical values 5-8) or not (vertical values 1-4). It also indicates if the sign of the columns of B was not changed (values 1, 5), was changed for the first column only (values 2, 6), second column only (values 3, 7), or both columns (values 4, 8).

Figure A-3: Robustness for $T = 1,000$

A) Degrees of freedom



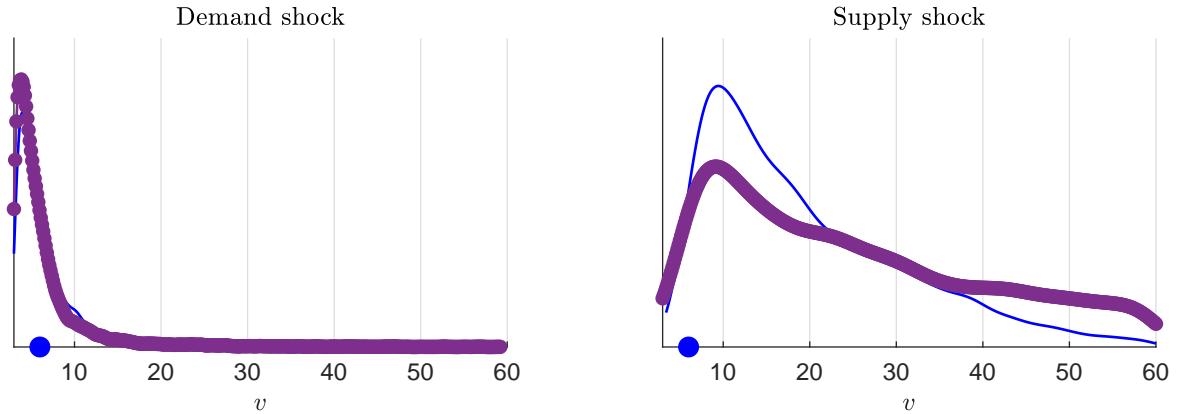
B) Impulse responses



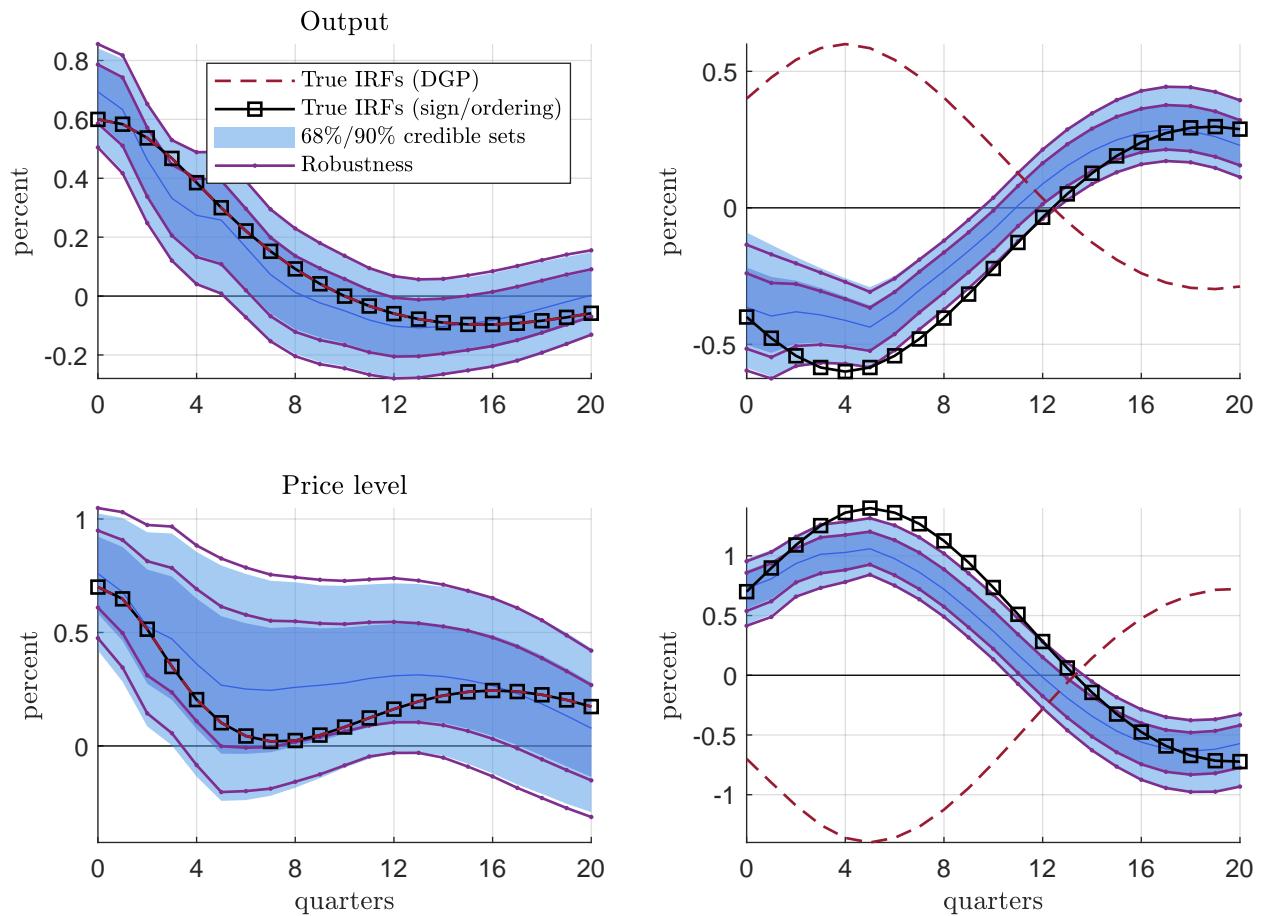
Note: In Panel A), the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B), the blue line and shaded areas show the pointwise median and credible sets corresponding to the baseline estimation, while the red dotted lines show the pointwise credible sets in the alternative specification.

Figure A-4: Robustness for flat prior on v

A) Degrees of freedom



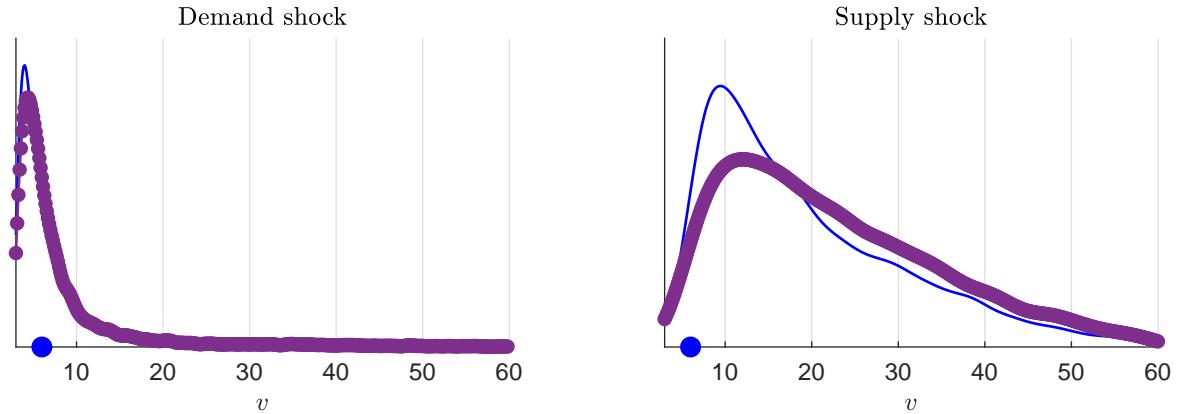
B) Impulse responses



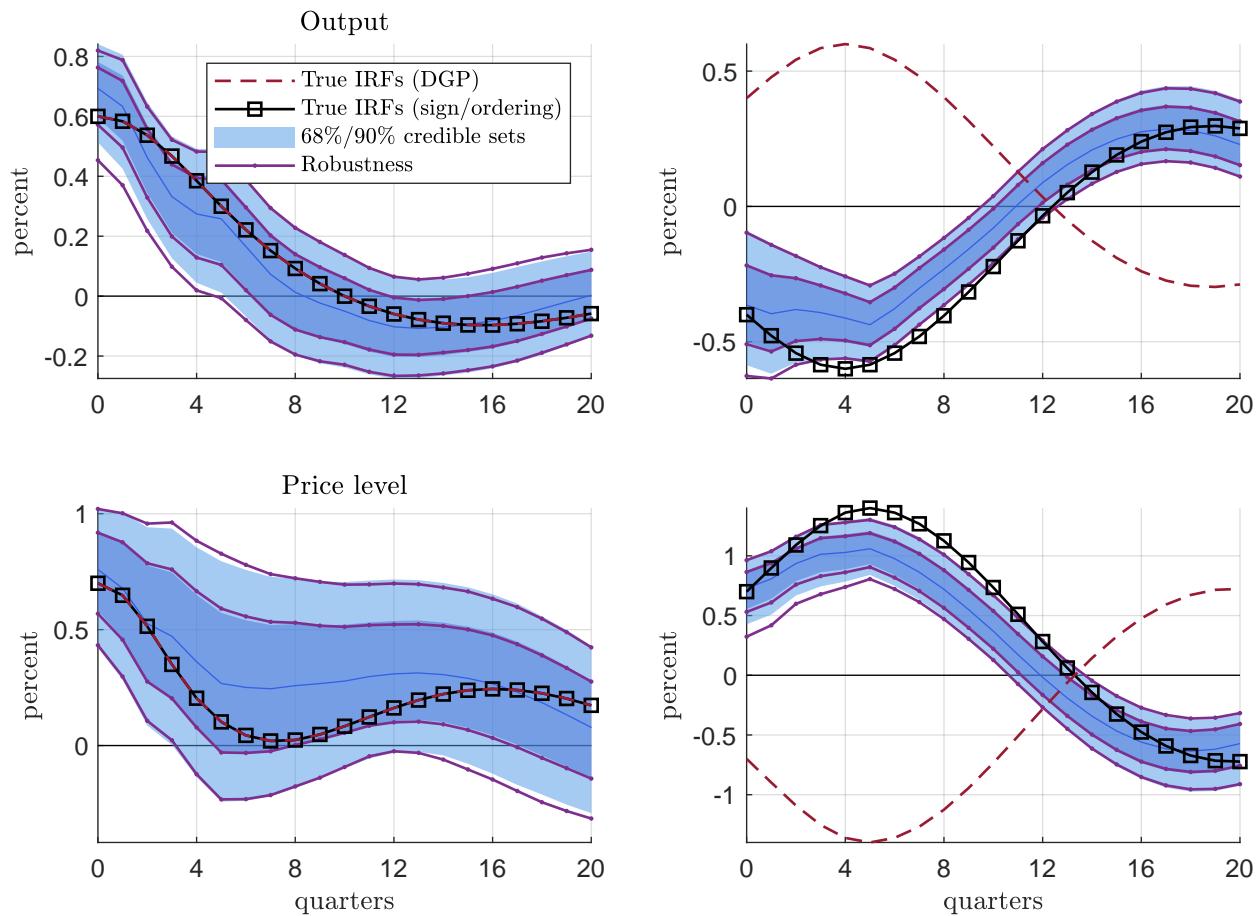
Note: In Panel A), the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B), the blue line and shaded areas show the pointwise median and credible sets corresponding to the baseline estimation, while the red dotted lines show the pointwise credible sets in the alternative specification.

Figure A-5: Robustness for flat prior on A

A) Degrees of freedom

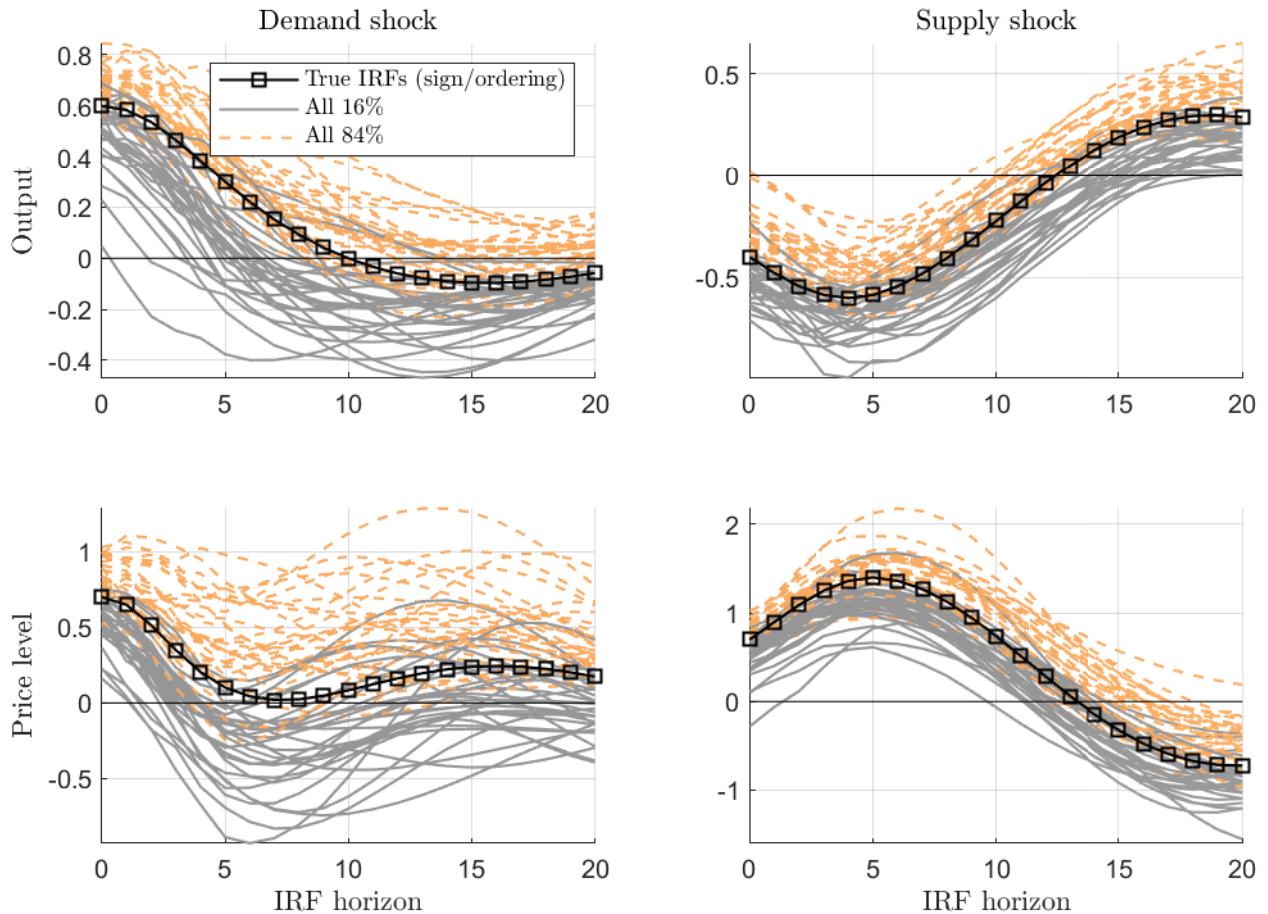


B) Impulse responses



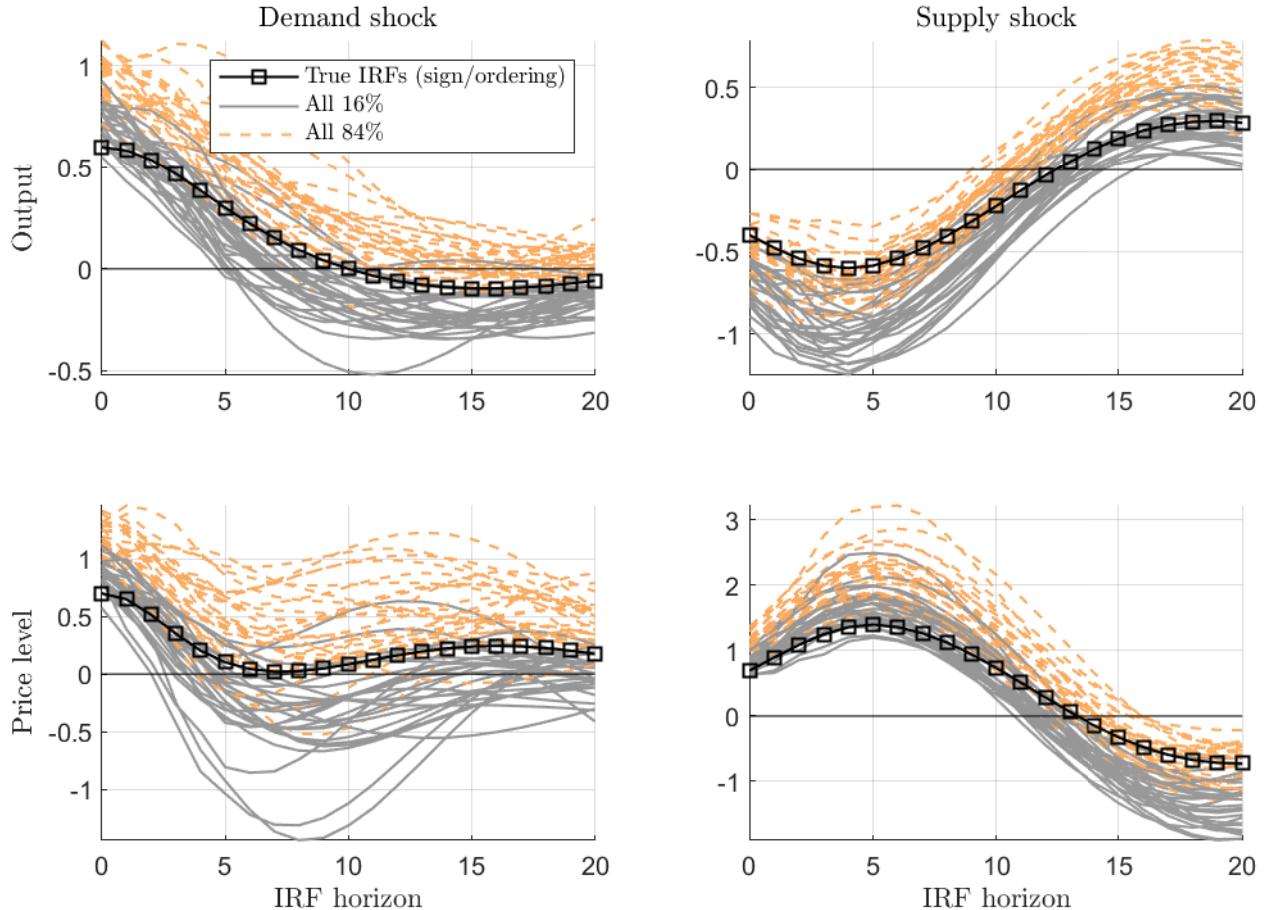
Note: In Panel A), the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B), the blue line and shaded areas show the pointwise median and credible sets corresponding to the baseline estimation, while the red dotted lines show the pointwise credible sets in the alternative specification.

Figure A-6: Impulse response functions, robustness across 100 samples, 68% credible sets



Note: The figure shows all 68% credible sets associated with 100 repetitions over newly drawn pseudo datasets. Each repetition applies the generalized LP normalization, as explained in the paper. To improve comparability, the sign and ordering of the shocks from each repetition was selected for this figure to improve comparability with the black dotted line, which corresponds to the true impulse response shown in [Figure 5](#)

Figure A-7: Impulse response functions, robustness across 100 samples and DGP featuring Laplace shocks, 68% credible sets

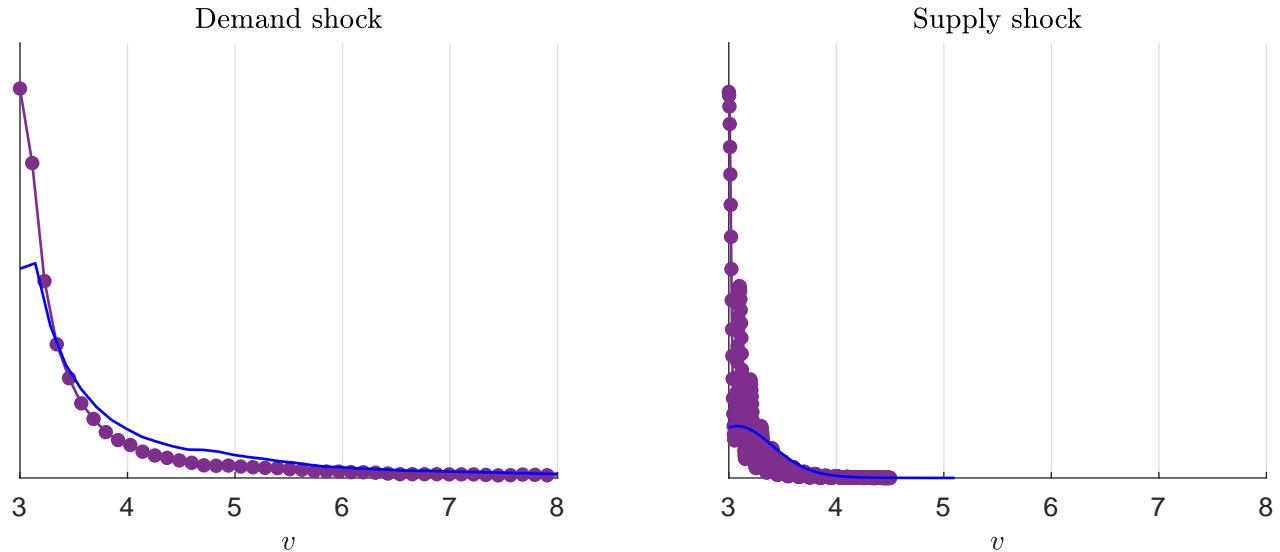


Note: The figure shows all 68% credible sets associated with 100 repetitions over newly drawn pseudo datasets. Each repetition applies the generalized LP normalization, as explained in the paper. To improve comparability, the sign and ordering of the shocks from each repetition was selected for this figure to improve comparability with the black dotted line, which corresponds to the true impulse response shown in [Figure 5](#).

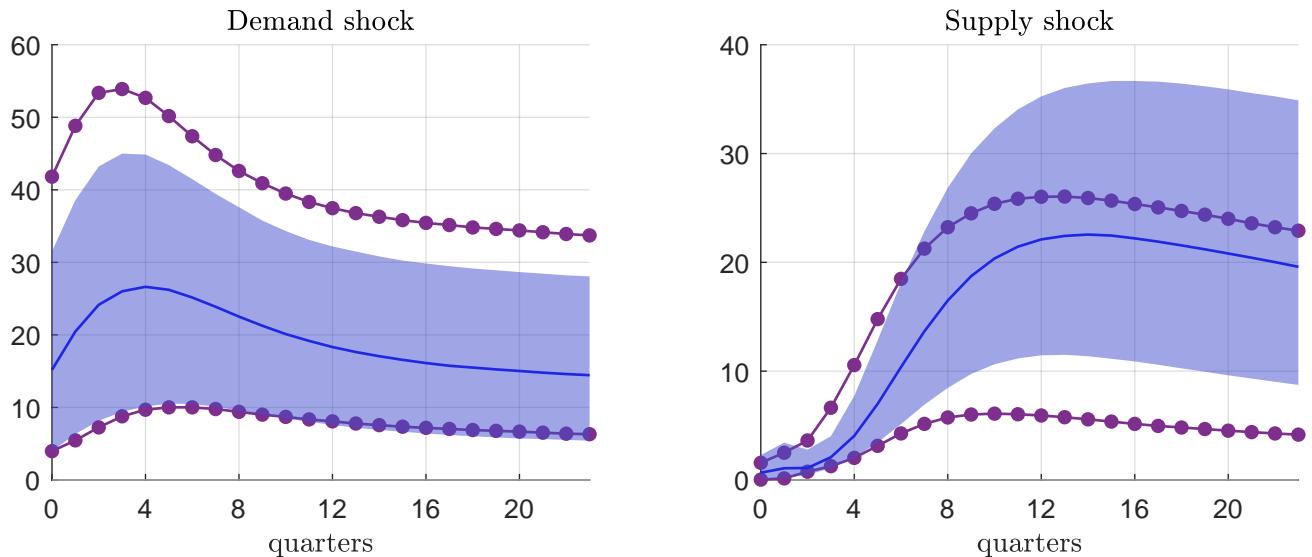
G Additional material on the application to the US GDP

Figure A-8: Robustness for $p = 4$ (1/2)

A) degrees of freedom

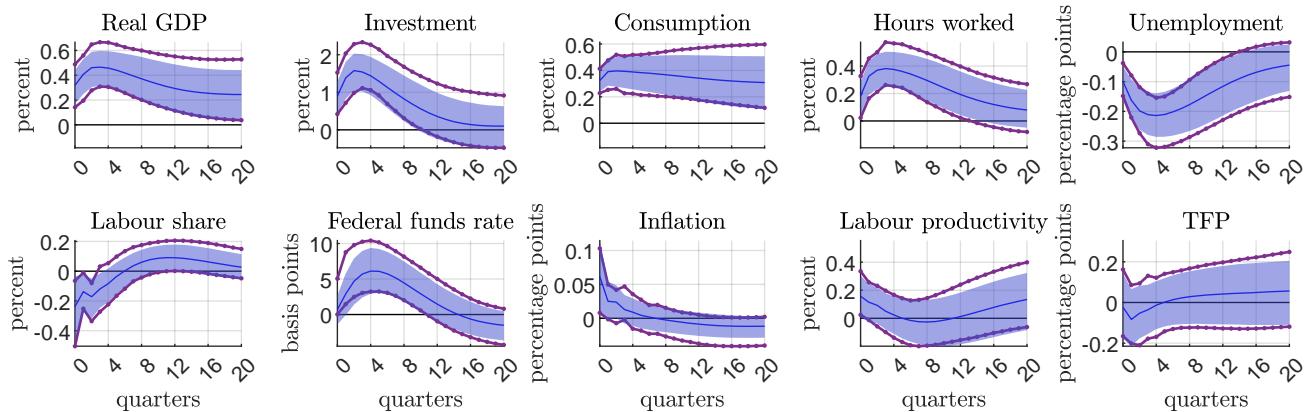


B) forecast error variance decomposition

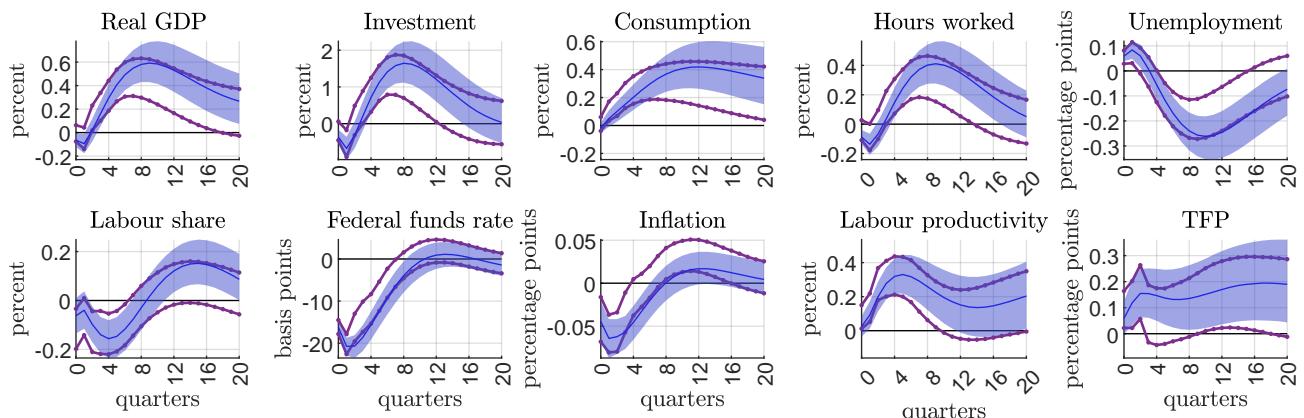


Note: The figure shows the results when the lag length of the model is increased to 4. In panel A the solid blue lines show the posterior distribution from the baseline specification, while the dotted purple line shows the posterior distribution under the alternative specification. In panel B the solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-9: Robustness for $p = 4$ (2/2)
IRFs to a demand shock

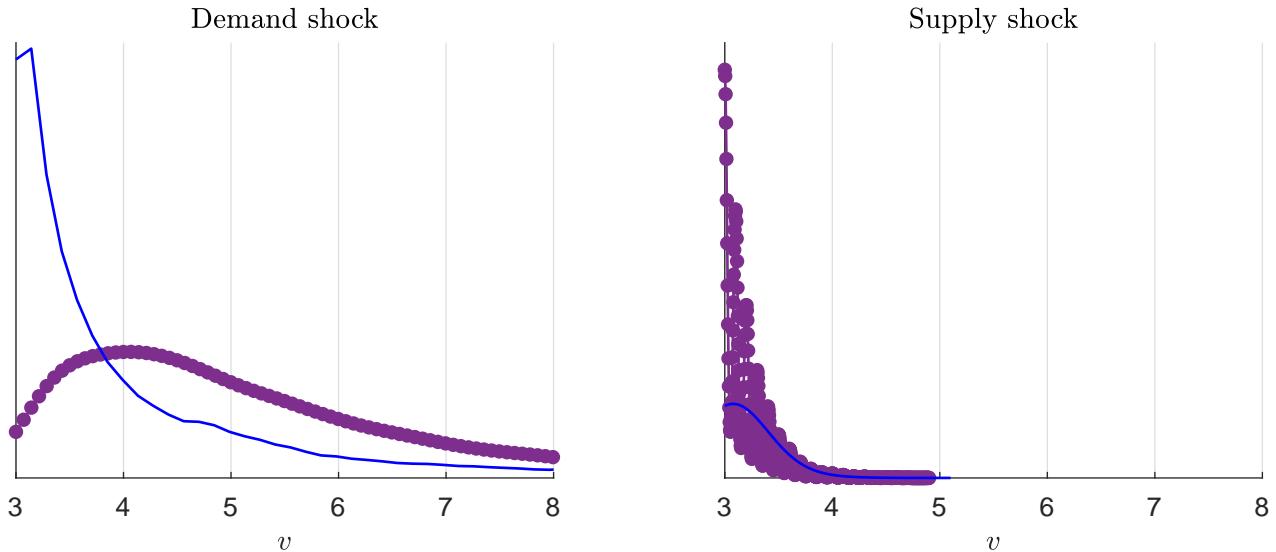


IRFs to a supply shock

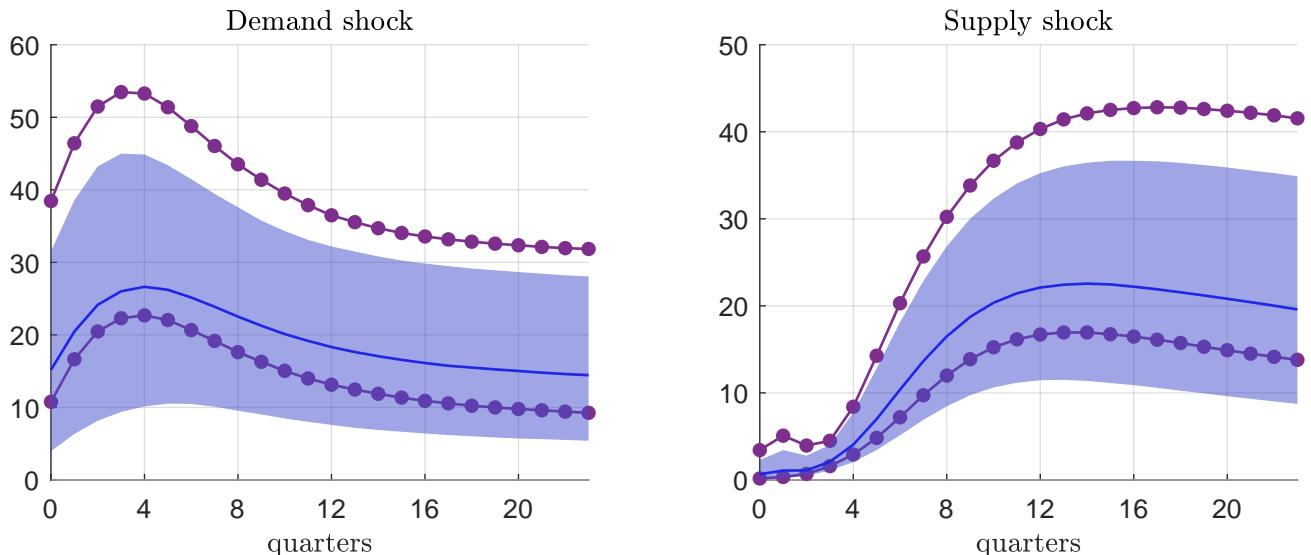


Note: The figure shows the results when the lag length of the model is increased to 4. The solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-10: Robustness for flat prior on A (1/2)
 A) degrees of freedom

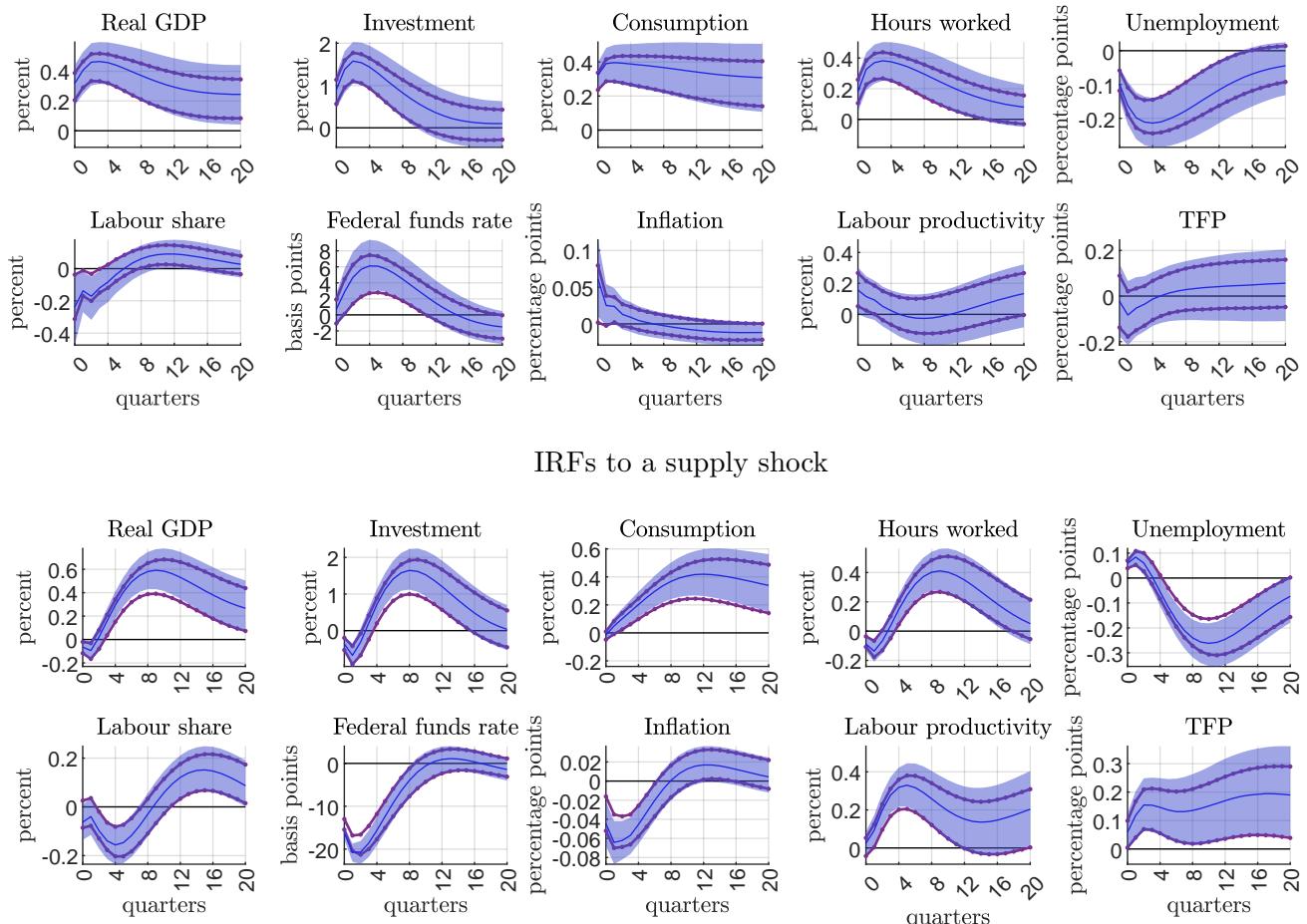


B) forecast error variance decomposition



Note: The figure shows the results when using a flat prior on A rather than B . In panel A the solid blue lines show the posterior distribution from the baseline specification, while the dotted purple line shows the posterior distribution under the alternative specification. In panel B the solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

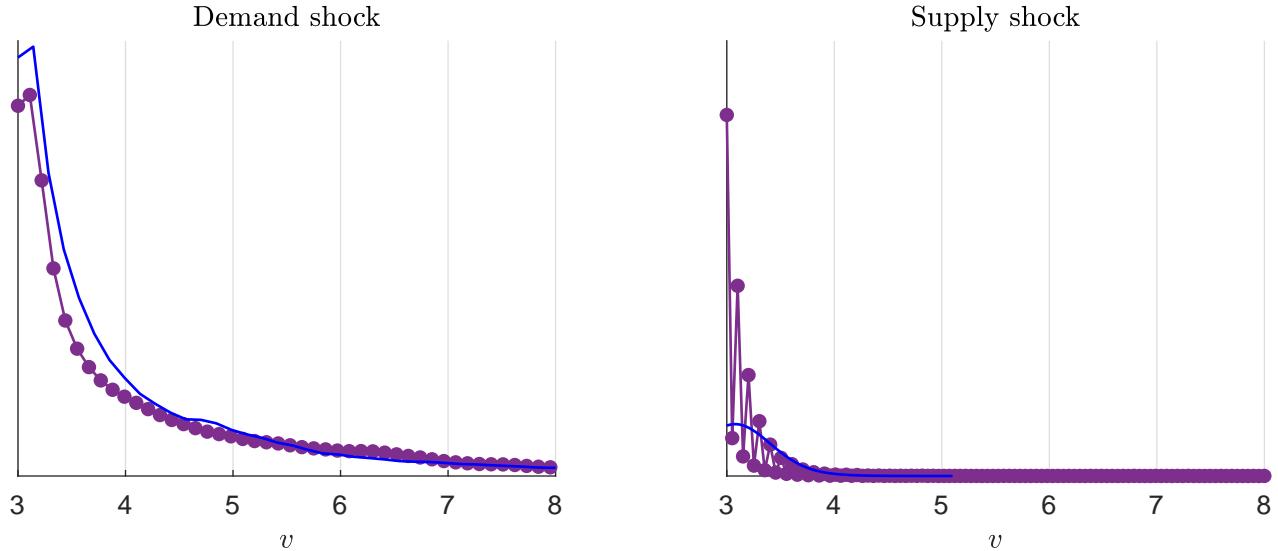
Figure A-11: Robustness for flat prior on A (2/2)
IRFs to a demand shock



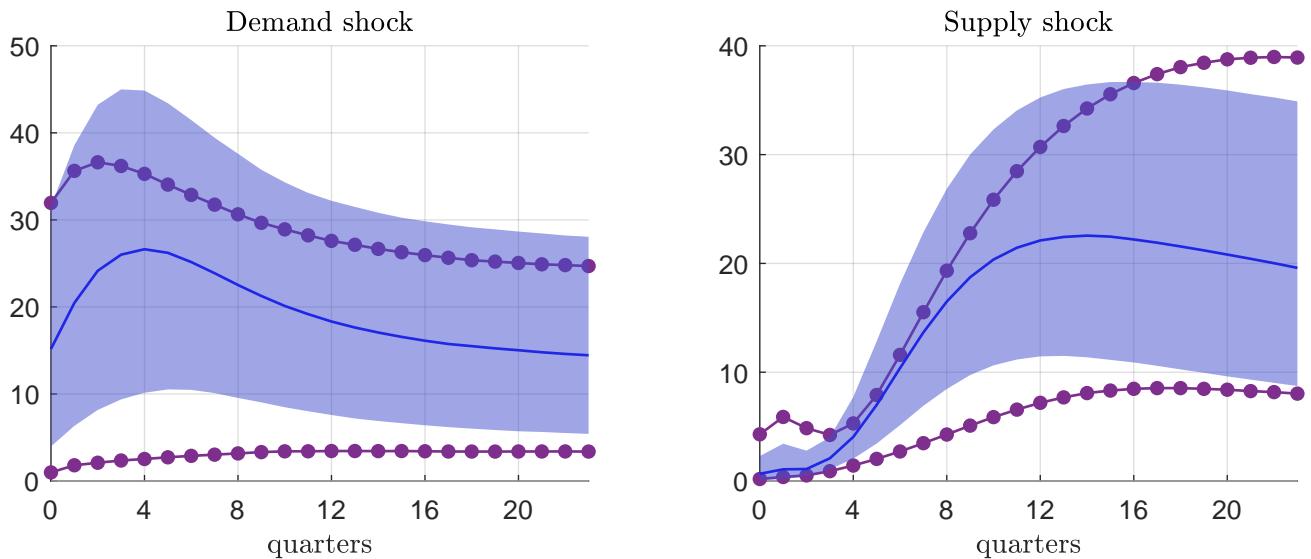
Note: The figure shows the results when using a flat prior on A rather than B . The solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-12: Robustness for 1948-2017 and 3m T-bill rate (1/2)

A) degrees of freedom

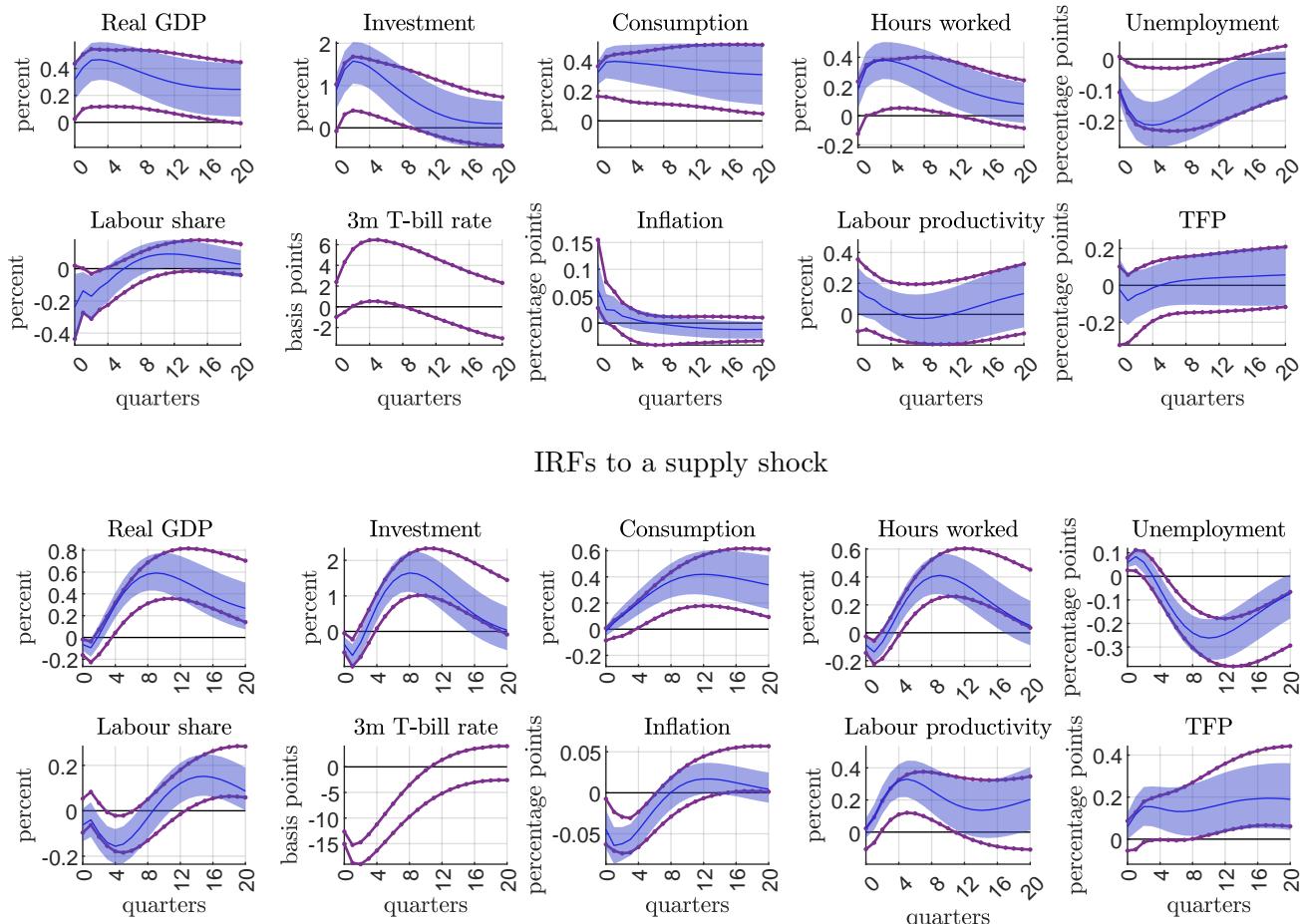


B) forecast error variance decomposition



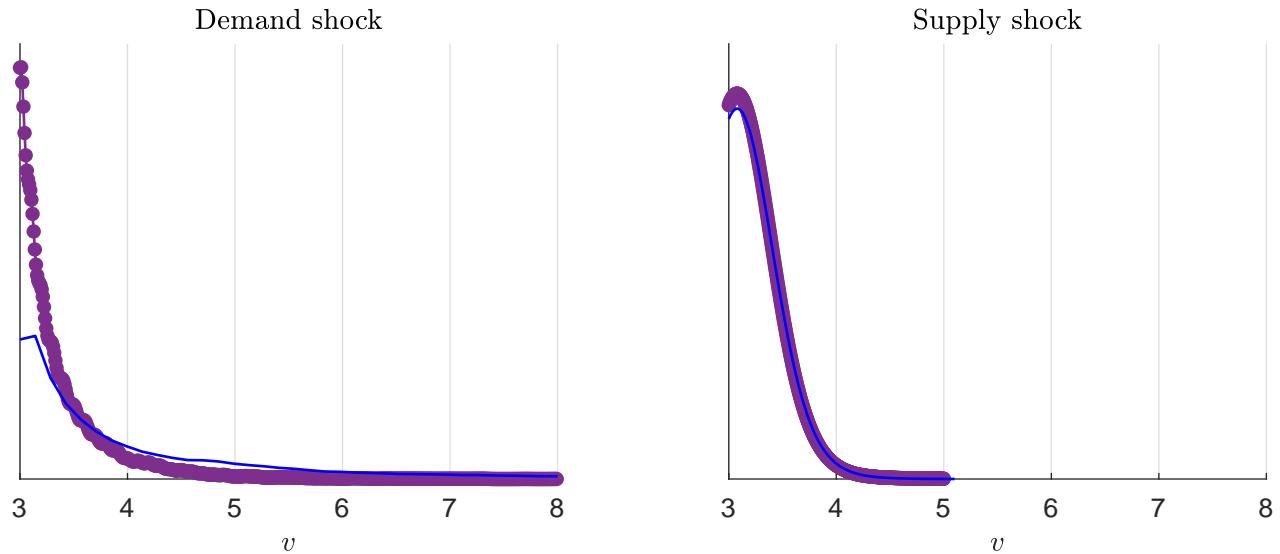
Note: The figure shows the results when the sample starts in 1948Q1 and the federal funds rate is replaced with the 3m T-bill rate. In panel A the solid blue lines show the posterior distribution from the baseline specification, while the dotted purple line shows the posterior distribution under the alternative specification. In panel B the solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-13: Robustness for 1948-2017 and 3m T-bill rate (2/2)
IRFs to a demand shock

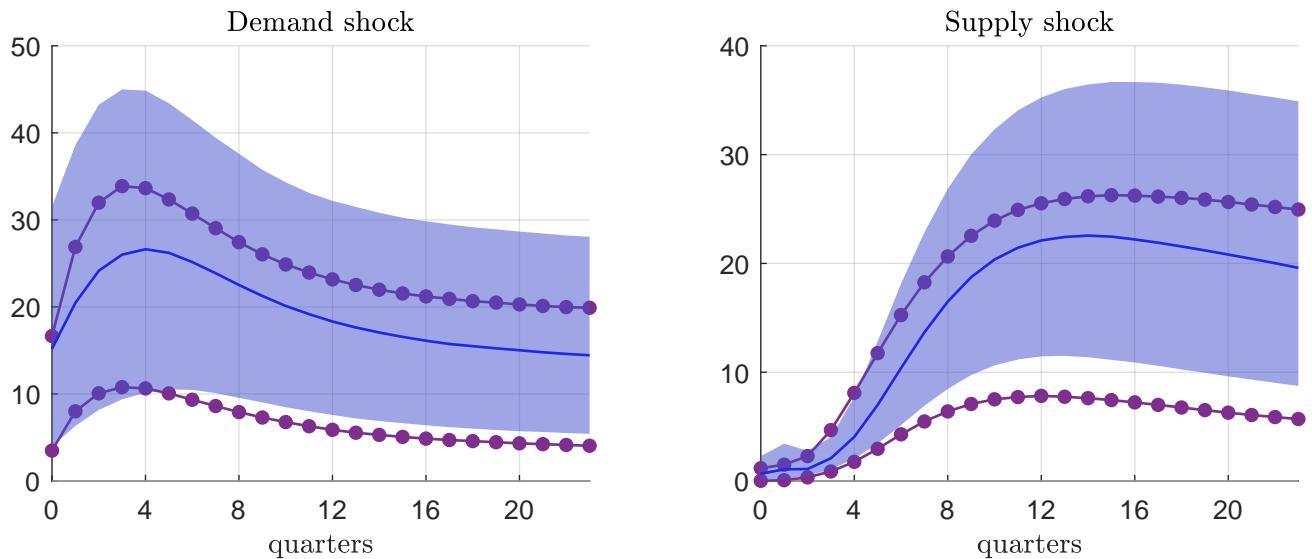


Note: The figure shows the results when the sample starts in 1948Q1 and the federal funds rate is replaced with the 3m T-bill rate. The solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-14: Robustness for adding credit spreads (1/2)
 A) degrees of freedom

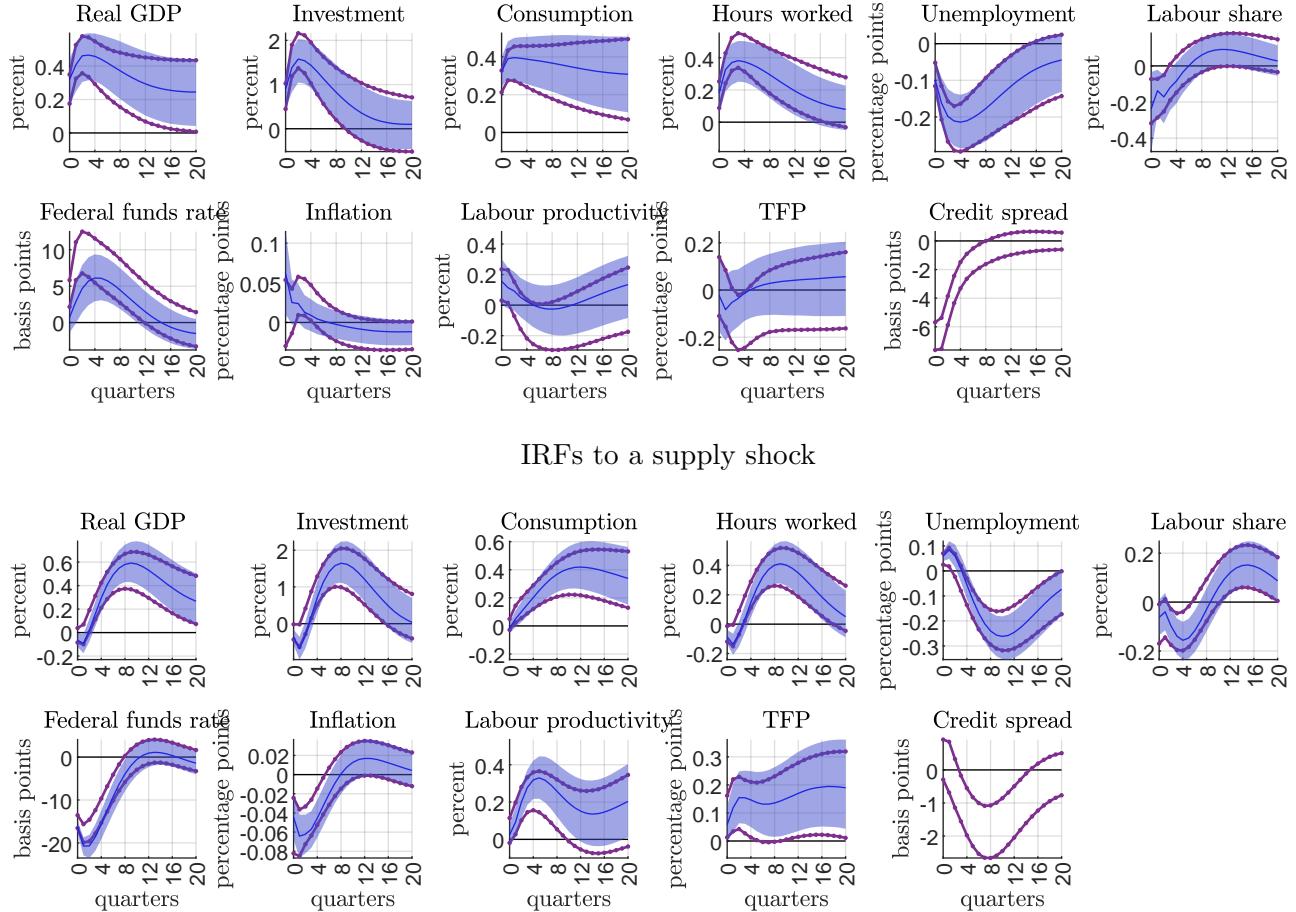


B) forecast error variance decomposition



Note: The figure shows the results when adding the difference between BAA-rated corporate bonds and the 10y US government bond rate. In panel A the solid blue lines show the posterior distribution from the baseline specification, while the dotted purple line shows the posterior distribution under the alternative specification. In panel B the solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.

Figure A-15: Robustness for adding credit spreads (2/2)
IRFs to a demand shock



Note: The figure shows the results when adding the difference between BAA-rated corporate bonds and the 10y US government bond rate. The solid lines and shaded areas show the pointwise posterior median and 68% credible sets, while the dotted purple lines show the 68% posterior credible sets under the alternative specification.