

A Scalable Framework for Statistical Identification in Structural VARs

Andrzej Kocięcki* Christian Matthes† Michele Piffer‡

March 12, 2026

Abstract

Non-Gaussian shocks can identify causal effects in structural VARs, but existing approaches face both conceptual and computational challenges. Conceptually, shocks are only identified up to sign and permutation, requiring careful normalization. Computationally, standard algorithms scale poorly, limiting applications to small models. We develop a Bayesian framework that resolves both issues. Our approach applies to all common departures from Gaussianity – such as heavy tails and stochastic volatility – and accommodates external instruments to sharpen identification. This expands the scope of non-Gaussian SVARs to the large-scale models increasingly used in economics.

JEL classification: C11, C32.

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

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

†University of Notre Dame, Notre Dame, Indiana, USA. e-mail: cmatthes@nd.edu

‡Bank of England and King’s Business School, King’s College London, UK. e-mail: m.b.piffer@gmail.com. We thank Luca Gambetti, Alessio Moneta, and Andrea Renzetti for helpful comments and suggestions. We also thank the participants at several seminars and conferences, including the Barcelona Summer Forum, the Cleveland and Philadelphia Feds, 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. Some of the results in this paper were previously circulated in a paper titled “A Non-Gaussian GDP Anatomy”.

1 Introduction

Since at least [Tinbergen \(1939\)](#), economists have understood that economic outcomes reflect both endogenous responses to economic conditions and unexpected changes—shocks. Identifying these shocks and measuring their effects remains a central challenge. In standard linear time series models, including vector autoregressions (VARs) ([Sims, 1980](#)), many different identification assumptions yield the same first and second moments for observables. Thus, Gaussianity offers no additional identifying information in such settings. To address this, researchers have turned to non-Gaussian models, including those with stochastic volatility (e.g., [Rigobon, 2003](#); [Lewis, 2021](#)) and t -distributed shocks (e.g., [Brunnermeier et al., 2021](#)). Both settings enable identification of structural shocks — though only up to sign and permutation.

Despite these advances, two key problems remain. First, existing approaches to statistical identification frequently rely on algorithms such as the Metropolis-Hastings algorithm, which scale poorly with model size.¹ This limits the number of observables that can be included. Second, the normalization problem — identification only up to sign and permutation — can lead to misleading posterior distributions if left unresolved ([Hamilton et al., 2007](#)). Existing solutions either do not scale ([Jarociński, 2024](#)), or can still yield multimodal posteriors ([Lanne et al., 2017](#); [Gouriéroux et al., 2020](#)), as we show in [Section 3](#).

This paper contributes three solutions. First, we develop a fast Gibbs sampler for non-Gaussian structural VARs that accommodates general forms of non-Gaussianity, including t -distributed shocks and stochastic volatility. Second, we propose a new normalization algorithm that handles sign and permutation indeterminacy by extending the likelihood-preserving approach of [Waggoner and Zha \(2003b\)](#). Importantly, our normalization algorithm works well in large systems. Third, we allow for additional identification restrictions, including zero restrictions, to incorporate information from external instruments ([Plagborg-Møller and Wolf, 2021](#)).

Our focus is Bayesian inference, which is common in VARs due to the regularization enabled by informative priors ([Bańbura et al., 2010](#); [Chan, 2020](#)). Maximum likelihood methods are impractical in large systems, and full-information frequentist approaches are rarely used. That said, our normalization algorithm can support GMM-type es-

¹[Arias et al. \(2025\)](#) recently proposed to use elliptical slice sampling to draw from the posterior of a structural VAR. Although this approach scales up well, it is restricted to applications in which either the likelihood or the prior for the relevant parameters is Gaussian (or can be reparameterized as Gaussian).

timation (Lanne and Luoto, 2021; Lanne et al., 2017; Keweloh, 2021), especially in combination with bootstrap methods that require many function evaluations. Methodologically, our approach extends the Gibbs sampling framework of Waggoner and Zha (2003a) from Gaussian to non-Gaussian settings.²

Our paper relates to a broader literature on non-Gaussian SVARs and hybrid identification strategies. For example, Braun (2023) and Herwartz (2018) apply non-parametric methods, while Drautzburg and Wright (2023) combine moment restrictions with sign restrictions to rule out dependent shocks. Like those papers, we assume shocks are independent, which is sufficient for identification.³ Andrade et al. (2023) use flexible moment restrictions alongside standard sign restrictions. Our assumption of t -distributed shocks delivers direct estimates of the degrees of freedom, which measure deviations from Gaussianity. This parametric approach is more parsimonious than non-parametric methods, but more vulnerable to misspecification. To address this, we provide Monte Carlo evidence that our method remains accurate even when the true shocks are not t -distributed. That will also be our specification in the application that follows.

Our work also connects to the literature on identification via stochastic volatility (e.g., Carriero et al., 2021; Lanne et al., 2010; Lewis, 2021). Recent work on Gibbs samplers for such models includes Chan et al. (2024) and Wu and Koop (2023). What distinguishes our method is its generality: we accommodate multiple sources of non-Gaussianity, allow additional identifying restrictions, and offer a new normalization algorithm that works even in large models. Such a normalization is crucial to solve the “label switching” problem, which, if not properly resolved, does not only slow down convergence of any posterior sampler, but can invalidate posterior consistency, as highlighted by Carriero et al. (2024).

In our applications, we focus on t -distributed shocks because they have recently become popular in structural VARs both as an identification device and as a way to model fat tails (Cúrdia et al., 2014; Chiu et al., 2017) and thus improve forecasting performance. Key papers that rely on t -distributions include Lanne and Luoto (2020) and Anttonen et al. (2024). In forecasting, researchers often assume independent t -distributed shocks with a recursive ordering, which allows equation-by-equation estimation even in large VARs (Clark and Ravazzolo, 2015; Chiu et al., 2017; Chan,

²The Gibbs sampler is more efficient than the Metropolis-Hastings algorithm, which has been used in smaller structural VARs (Baumeister and Hamilton, 2015).

³For alternatives to independence that still allow for identification, see Mesters and Zwiernik (2024).

2020).⁴ But recursive orderings impose strong and often unrealistic timing restrictions, making them often less useful for structural analysis. Our paper also relates to an important recent contribution by [Jarociński \(2024\)](#), who identified monetary policy shocks in the US using t -distributed shocks.

To illustrate our method, we revisit one of the fundamental questions in macroeconomics: Which shocks account for the bulk of business cycle fluctuations? In a key contribution, [Angeletos et al. \(2020\)](#) provide a contemporary answer.⁵ 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.⁶ [Dou et al. \(2025\)](#) formalize this concern, deriving necessary and sufficient conditions under which max-share identifies the correct shock. These conditions are typically strong, including, for example, orthogonal impulse response functions.

We take a different approach. Using the same data as [Angeletos et al. \(2020\)](#), we assume the data are non-Gaussian – specifically, that structural shocks follow independent t -distributions. This ensures full identification of the SVAR parameters ([Comon, 1994](#)), up to sign and column permutations, which our normalization resolves.⁷ We then compute forecast variance decompositions across horizons to assess which shocks explain most of GDP fluctuations. Crucially, non-Gaussianity is a testable assumption: the estimated degrees of freedom measure its strength. A model with higher degrees of freedom is closer to Gaussian, weakening identification and widening posterior intervals.⁸ This logic also applies to stochastic volatility: We can compare Gaussian and non-Gaussian specifications using the marginal likelihood, providing a natural measure

⁴Alternatively, one can introduce as many Gaussian shocks as observables, yielding a factor structure in forecast errors ([Prüser, 2024](#)).

⁵For an alternative approach focusing on the cyclical component of the data, see [Bianchi et al. \(2023\)](#).

⁶[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.

⁷If only some shocks are non-Gaussian, those remain identified ([Maxand, 2020](#)).

⁸For a frequentist alternative that allows for weak identification, see [Hoesch et al. \(2024\)](#).

of identification strength.

In our application, we uncover a shock whose impulse response resembles the “main business cycle shock” of [Angeletos et al. \(2020\)](#), yet it does not account for most of the variation in GDP. Still, it is one of the two most important shocks. The other, more influential at longer horizons, moves total factor productivity and resembles a traditional supply shock, in line with real business cycle (RBC) theory ([Kydland and Prescott, 1982](#)). This finding is consistent with previous work by [Forni et al. \(2025\)](#).

The paper is structured as follows: Section 2 presents details of our VAR model, while Section 3 reports the results on Monte Carlo simulations assessing our methodology. Section 4 presents our results on US GDP. Section 5 concludes.

2 Our Approach

In this section, we describe a class of time series models that is general enough to encompass all common approaches to exploit statistical properties of the data to identify structural shocks and their effects. We then present our proposed reparametrization of the matrix that governs the initial impact of these shocks, explain how it yields a Gibbs sampler, and then detail our approach to normalization to address the model’s identification issues due to sign and permutation switches of the impact responses. We refer to Appendix F of the Online Appendix for a full description of the posterior sampling procedure, and to Appendices B-C-D for the technical derivations.

2.1 The Model

We study models of the form

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

$$\mathbf{e}_t \sim N(\mathbf{0}, I).$$

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 $\phi = \text{vec}(\Pi)$ as the vector that stacks the columns of Π vertically. D_t is a diagonal matrix that contains the variances of the structural shocks on the main diagonal, $D_t = \text{diag}(d_{1t}, \dots, d_{it}, \dots, d_{kt})$.⁹ 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 (1) is Gaussian. It is exactly the time-varying nature of the elements of D_t that allows identification of the impact matrix B and the associated structural shocks up to sign and permutation. The exact conditions for identification depend on the specifics of D , but if all but one element of D_t are time-varying and the paths of volatility (i.e. the elements of D_t) are not proportional to each other over time, identification obtains (see [Lewis, 2021](#); [Bertsche and Braun, 2022](#) for the case of general stochastic volatility, [Maxand, 2020](#) for the case of non-Gaussian shocks with constant variance, and [Lewis, 2025](#) for a comparison).

At this point, we do not need to choose one specific law of motion for D_t , but it is useful to highlight ([Geweke, 1993](#)) that if we make the elements of D_t follow independent inverse gamma distributions with a choice of parameters we discuss below, the model is isomorphic to a model with t -distributed independent structural shocks (i.e. each element of $D_t^{\frac{1}{2}} \mathbf{e}_t$ then follows an independent t distribution). However, we want to emphasize that the mixture representation of the error component in (1) is very flexible - for a general approximation theorem of non-Gaussian densities by a mixture of Normal distributions, see, for example, [Goodfellow et al. \(2016\)](#); [Nguyen and McLachlan \(2019\)](#). [Braun \(2023\)](#) builds a Gibbs sampler for such a specification. On the other hand, more general persistent processes for the elements of D_t , e.g. smooth transition models ([Lütkepohl and Netsunajev, 2017](#)), Markov switching models ([Lanne et al., 2010](#); [Lütkepohl and Woźniak, 2020](#)), models with known volatility regimes ([Rigobon, 2003](#)), or models with autoregressive stochastic volatility of various forms ([Normandin and Phaneuf, 2004](#); [Lütkepohl and Netsunajev, 2017](#); [Bertsche and Braun, 2022](#)), are nested in our class of models as well.¹⁰

More broadly, our requirement for the law of motion for D_t is that (i) a Gibbs sampler is available for inference for this part of the model so that estimation is fast, and (ii) it delivers identification of B up to sign and permutation. Before we go on, two points are worth mentioning: In the case of $D_t = I; \forall t$, we return to the standard Gaussian model with constant variances where the statistical properties of

⁹Because D_t is a diagonal matrix with variances on the main diagonal, all elements of D_t are positive and the matrix square root is unique.

¹⁰Our approach can also be used in the case where t -distributed structural shocks have persistent stochastic volatility, as in [Chiu et al. \(2017\)](#); [Clark and Ravazzolo \(2015\)](#).

the data do not suffice to identify the impact of shocks B . As such, one natural way to test the strength of identification is to compare the fit of a Gaussian model versus the non-Gaussian alternative of choice, for example via the computation of marginal likelihoods, for which many algorithms are available. Second, since our framework is agnostic and can accommodate a large set of different modeling choices for D , it provides a natural laboratory to compare these different identification schemes.

Our approach exploits a Gibbs sampler, meaning that we need to characterize the posterior distribution of a given set of parameters conditional on not only the data but also all other parameters. We first focus here on the parts of the Gibbs sampler that are novel, in particular the drawing of the B matrix. For completeness, we also describe below how we draw D and associated parameters for the case where structural shocks follow a t -distribution, which is the specification we use in our application.

2.2 Prior Distributions

We define $p(\boldsymbol{\phi}, B, \boldsymbol{v})$ as the joint prior distribution for the parameters of model (1). \boldsymbol{v} denotes parameters that govern the D_t process. In our application, we follow Geweke (1993) and study a model with t -distributed errors, which leads to an inverse Gamma prior on the latent variables d_{it} parametrized as

$$p(d_{it}|h_i, r_i) = \frac{r_i^{h_i}}{\Gamma(h_i)} \cdot d_{it}^{-h_i-1} e^{-r_i \cdot \frac{1}{d_{it}}}, \quad (2a)$$

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

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

d_{it} is a-priori independent across (i, t) and a-priori independent of $(\boldsymbol{\phi}, B)$. 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). In this setting, v_i represents the degrees of freedom of shock i , where we define $\boldsymbol{v} = (v_1, \dots, v_i, \dots, v_k)$. We set the scale parameters of the t -distribution to normalize the variance of the structural shocks to unity and assume $v_i > 2, \forall i$.

More broadly, we work with the following decomposition of the joint prior distri-

bution:

$$p(\boldsymbol{\phi}, B, \boldsymbol{v}, D) = p(\boldsymbol{\phi}, B, \boldsymbol{v}) \cdot p(D|\boldsymbol{v}), \quad (3)$$

$$= p(\boldsymbol{\phi}) \cdot p(B) \cdot p(\boldsymbol{v}) \cdot p(D|\boldsymbol{v}) \quad (4)$$

Furthermore, in the application with t -distributed shocks, we also have that

$$p(D|\boldsymbol{v}) = \prod_{t=1}^T \prod_{i=1}^k p(d_{it}|v_i). \quad (5)$$

This decomposition relies on the variances d_{it} being *i.i.d.* in that setting. Instead, stochastic volatility specifications common in the literature assume persistence, but still imply a prior distribution $p(D|\boldsymbol{v})$. The prior $p(d_{it}|v_i)$ in our application is the inverse Gamma prior described in equation (2). We work with a Normal prior for $\boldsymbol{\phi}$, which nests priors already used for large VARs (Bańbura 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. Lastly, our method works with a wide range of candidate priors for \boldsymbol{v} . In our simulations and application we use a truncated Normal prior with a large upper bound so that our prior allows for shocks that are basically indistinguishable from Gaussian shocks.

2.3 Posterior Sampling

We want to explore the joint posterior distribution of model (1) 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 $p(\boldsymbol{\phi}|Y, B, D, \boldsymbol{v})$ is a Normal distribution. Furthermore, we can borrow existing approaches from the literature to draw these parameters efficiently when the model is large (Bańbura et al., 2010). $p(D|Y, \boldsymbol{\phi}, B, \boldsymbol{v})$ and $p(\boldsymbol{v}|Y, \boldsymbol{\phi}, B, D)$, the conditional posterior distributions of the structural shock variances and the parameters governing the evolution of those variances, depend on how exactly non-Gaussianity is modeled. For the case with t -distributed errors, Geweke (1993) shows that $p(D|Y, \boldsymbol{\phi}, B, \boldsymbol{v})$ is an inverse 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). We

show below how to draw from the distribution of the degrees of freedom $p(\mathbf{v}|Y, \boldsymbol{\phi}, B, D)$ in that case using a Gibbs sampler.

For a Gibbs sampler to be feasible for this model, it remains to derive a convenient way of sampling from the conditional distribution of B , which encodes the impact effects of structural shocks on the observables:

$$p(B|Y, \boldsymbol{\phi}, D, \mathbf{v}). \tag{6}$$

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

$$A = B^{-1}, \tag{7}$$

and achieve a computationally convenient sampling procedure for $p(B|Y, \boldsymbol{\phi}, D, \mathbf{v})$ 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 frequently resorted to less efficient methods, employing a Metropolis-Hastings step on $p(B|Y, \boldsymbol{\phi}, \mathbf{v}, D)$ (as in [Lanne and Luoto, 2020](#), [Brunnermeier et al., 2021](#)), or employing a more involved MCMC procedure on the full joint posterior distribution $p(\boldsymbol{\phi}, B, \mathbf{v}, D|Y)$ ([Anttonen et al., 2024](#)). One of our contributions is to develop a way of drawing from the conditional posterior distributions (6) and (10) 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 (6) hinges on a new reparametrization, which, to our knowledge, has not been used in Bayesian econometrics. Although our benchmark approach sets a flat prior to B , here we parameterize A instead, introducing the decomposition

$$A = \Lambda L U, \tag{8}$$

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 restricted to have ones on their main diagonals. We assume $\lambda_i \neq 0$ for all i , which is a necessary and sufficient condition for

(8) to exist and be unique.¹¹

The choice between using the A or B parametrization is inconsequential for our 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 (instead, our applications and simulations in the main text impose a flat prior on B , taking into account the Jacobian of the mapping from A to B). 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. Hence, essentially (8) does not have restrictions other than the non-singularity of B (or A).¹² As a consequence, we can think of (8) much like other reparameterizations employed in the SVAR literature.¹³

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 (8) conveniently yields

$$|\det(A)| = \prod_{i=1}^k |\lambda_i|, \quad (9)$$

i.e., 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 (i.e., the non-normalized density) of a Gamma distribution. These two properties make a Gibbs sampler feasible, because they imply a Gamma conditional posterior for Λ and a Normal conditional posterior for both L and U . In summary, by drawing L, U and Λ separately, we can derive a Gibbs sampler for A (or B). Details of those posterior distributions can be found in the Online Appendix. To the best of our knowledge, this insight has never been exploited in other existing

¹¹Under this reparametrization, $B = U^{-1} L^{-1} \Lambda^{-1}$.

¹²The LU decomposition can be computed for any invertible matrix A , though it may require permuting its rows. In our structural VAR, this amounts to nothing more than changing the order of the equations.

¹³For example, [Arias et al. \(2018\)](#) apply sign restrictions by setting $A = Q h(\Sigma)$, where Σ is the reduced-form covariance matrix of a Gaussian VAR, $h(\cdot)$ is a unique decomposition, and Q is an orthogonal matrix. In contrast, [Wu and Koop \(2023\)](#) use the eigenvalue-eigenvector decomposition of Σ to simplify sampling in large-scale VAR models.

contributions that construct Gibbs samplers for SVAR models (see [Waggoner and Zha, 2003a](#); [Villani, 2009](#); [Rubio-Ramirez et al., 2010](#); [Arias et al., 2025](#)).

One key fundamental issue that remains with regard to B or A is that the likelihood function has many peaks that correspond to impact matrices that have the same economic interpretation, but differ in the sign and ordering of their columns. We tackle that normalization problem below. Our Gibbs sampler requires a flat prior on either B or A . In a 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 the necessary identification restrictions ([Baumeister and Hamilton, 2015](#); [Arias et al., 2019](#); [Inoue and Kilian, 2020](#)). Nonetheless, in [Section 2.5](#) we show how to move away from uniform priors to incorporate additional information for identification.

To complete our Gibbs procedure for the case with t -distributed shocks, we must also show that the conditional posterior for the degrees of freedom has a tractable form:

$$p(\boldsymbol{v}|Y, \boldsymbol{\phi}, B, D). \tag{10}$$

We accomplish this via a Griddy-Gibbs sampler ([Ritter and Tanner, 1992](#)), which discretizes \boldsymbol{v} . In our simulations and empirical application, we let \boldsymbol{v} range over a fine grid between 3 and 60, specifying a truncated Normal prior over this interval.¹⁴ Further details appear in the Online Appendix.

2.4 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: For any $k \times k$ matrix B , there are $k! \cdot 2^k$ 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

¹⁴In principle, we could set the lower bound to 2, which is a bound we need to impose because we want to study the response to a one standard deviation shock. However, when doing so, we encountered the problems identified by [Fernandez and Steel \(1999\)](#). As shown by their Theorem 5, when \boldsymbol{v} approaches 2, since the scale parameter of the t -distribution, defined in [Online Appendix B](#), tends to zero in such a case, the t -Student likelihood function can become arbitrarily large and ultimately lead to an improper posterior.

different types, multimodality of posterior distributions (Hamilton et al., 2007), and slow convergence of the posterior sampler. We first show this issue in an example before turning to our solution.

2.4.1 An Illustration

To highlight the need for normalization, we focus on a numerical example where the non-Gaussianity comes from t -distributed structural shocks. Details and derivations for this example can be found in Appendix A of the Online Appendix.

Consider the following bivariate data generating process:

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

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})$ are independently t -distributed structural shocks with common degrees of freedom \boldsymbol{v} . We normalize the variance of these shocks to 1, which means that we assume the variance is finite. Furthermore, for this example, we find it useful to illustrate identification by computing fourth-order moments, so we assume $\boldsymbol{v} = 6$ in this numerical example. Model (11) can be rewritten as

$$\boldsymbol{y}_t = B_c Q \boldsymbol{\epsilon}_t, \quad (12)$$

where we define $\Sigma = BB'$ as the covariance matrix of \boldsymbol{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, \boldsymbol{v}) , or equivalently, (B_c, Q, \boldsymbol{v}) . We reparametrize the model and set

$$Q = Q(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad (13)$$

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 that we will come back to in the following), meaning that the number of modes of the likelihood function is cut in half. We treat B_c as known (since B_c is identified by second moments alone). We also assume that the degrees of freedom are known for now (but study

the identification of those parameters in Appendix A). We fix numerical values for all parameters and assess how the fourth moments of the data vary as we vary the parameters. Checking whether these fourth moments identify the model parameters amounts to asking if more than one value of θ implies the same fourth moments. We compute fourth moments in this section to assess identification, but the parameter values where implied fourth moments are equal to the true population moments are exactly those parameter values associated with peaks of the likelihood function, so in this specific example this approach is without loss of generality. By construction, any θ is consistent with the second moments alone, and because we study symmetric distributions, all third moments are zero.

Each panel in Figure 1 shows the fourth moments for this model evaluated on a grid of θ . The moments are shown in percentage deviation from the true moments. Panel A) reports the case of Gaussianity. In that case, the moments are constant in θ , showing that fourth moments do not provide any information when the model is Gaussian. Since the likelihood function of the model under Gaussianity is uniquely pinned down by the first and second moments, higher moments provide no additional information. With t -distributed errors instead, we see that there are 4 distinct parameter values that give the true fourth moments.¹⁵ As we confirm in Appendix A, the B matrices implied by those values of θ only differ in signs and order of columns, highlighting the need for normalization.

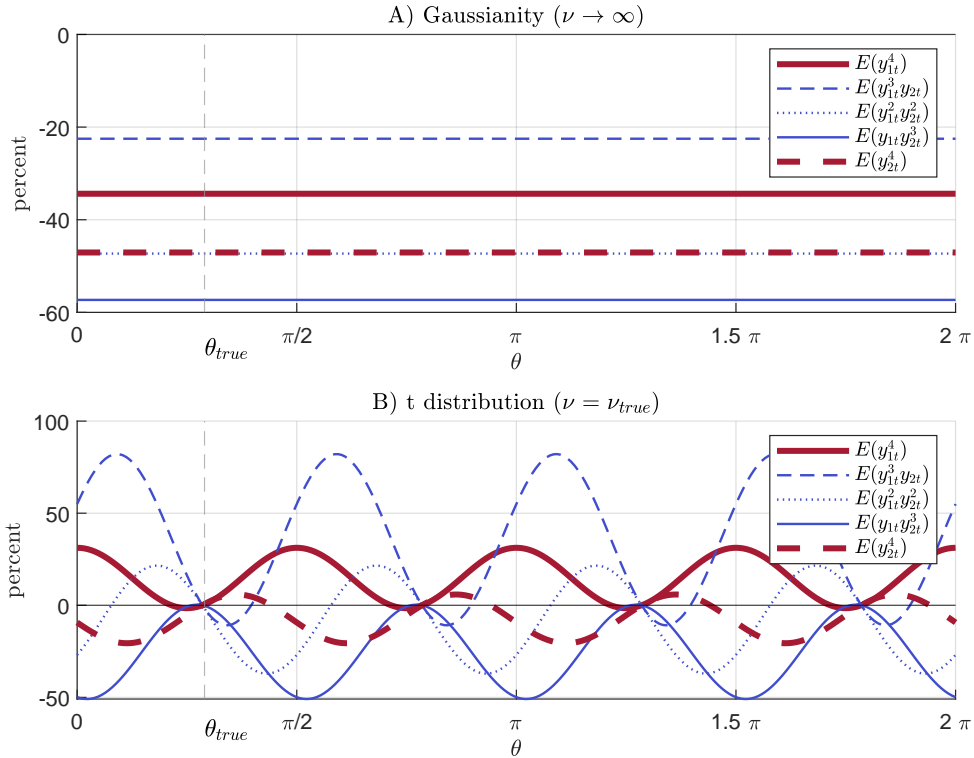
2.4.2 Our Normalization

In light of this issue, a second methodological contribution of our paper is a new, computationally efficient way to handle sign and permutation indeterminacies in VARs with non-Gaussian shocks.¹⁶ Before starting the Gibbs sampler iterations, we first compute a matrix \hat{B} that approximates one of the model's modes, which serves as a

¹⁵There are actually 8 peaks of the likelihood function, but our use of Givens rotation matrices in this example rules out 4 of those because the rotation matrix Q in those cases has a determinant of -1 , which is ruled out by Givens rotation matrices. This issue does not arise in our estimation algorithm since we do not use Givens rotations there.

¹⁶A researcher can use a normalization strategy to remove this multiplicity during the sampling process, as we do, or sample the full unnormalized posterior. In the latter case, a normalization routine such as the one proposed in this paper will have to be applied ex post to be able to interpret results. We focus here on normalization issues related to the matrix B - depending on the choice of model for the volatility matrix D_t , researchers might need to solve additional, well known, normalization issues in the estimation of that block when considering discrete regime shifts in variances governed by a Markov chain, which can lead to a label switching problem.

Figure 1: Illustrative example



target in the Gibbs sampler. For example, \hat{B} could be the maximum-likelihood estimator of B .¹⁷ In each iteration of the Gibbs sampler, we then apply our normalization scheme, which we describe now.

Let P_s be a signed permutation matrix of dimension $k \times k$. A signed permutation matrix is a square matrix $P_s \in \mathbb{R}^{k \times k}$ that has exactly one nonzero element in each row and each column, where each nonzero element equals either $+1$ or -1 . Equivalently,

¹⁷One approach is to estimate \hat{B} by running the MLE procedure by [Lanne et al. \(2017\)](#) before starting the sampler. This approach works efficiently in small- and medium-scale models. For larger models, an alternative is to apply no normalization in the burn-in part of the Gibbs sampler, and set \hat{B} equal to the value of B associated with the highest value of the posterior among burn-in draws. The remainder of the paper considers both approaches. In practice, we use either of the described approaches to obtain an initial estimate \tilde{B} , and then compute \hat{B} as the sign and column permuted version of that initial estimate that minimizes the distance (measured as we discuss below) from the identity matrix. This last step is not necessary for our algorithm (nor does it change the properties of our algorithm), but makes comparison between specifications easier because the target \hat{B} is always chosen to be as close as possible to the identity matrix within the set of sign and column permuted versions of the initial estimate. The results are not affected by this choice.

it can be written as

$$P_s = DP,$$

where P is a (standard) permutation matrix and D is a diagonal matrix with diagonal entries equal to ± 1 . In our context, postmultiplying a matrix B by P_s changes the order and the signs of the columns of B .

In each sampler iteration, we draw (Λ, L, U) , thus obtaining A and B . We store BP_s , where P_s solves

$$\min_{P_s} \Delta; \quad \text{where} \quad \Delta = \text{trace} \left[(BP_s - \hat{B})' (\hat{B}\hat{B}')^{-1} (BP_s - \hat{B}) \right]. \quad (14)$$

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 *generalized LP normalization*. The most important feature of our approach is computational. Beyond interpretability, applying the normalization within each Gibbs iteration mitigates mode-switching/label-switching and improves mixing and convergence. Naively, one might compute the distance (14) for all $k! \cdot 2^k$ possible signed permutation matrices in each Gibbs iteration, which is an infeasible option even for moderate k . Instead, as we discuss in Appendix D of the Online Appendix, we can show that the minimization problem from (14) amounts to solving

$$\min_P \text{tr} \{ P \cdot -|\hat{B}^{-1}B| \}, \quad (15)$$

where P denotes the usual permutation matrix, tr signifies 'trace', and $|\hat{B}^{-1}B|$ means absolute values for all entries in $\hat{B}^{-1}B$ matrix (taken element-wise). Being able to rewrite (14) as (15) is a crucial advantage because it rewrites the key minimization problem as a linear assignment problem, for which the literature on combinatorial optimization has already derived solutions. In particular, the Online Appendix shows how to apply the Hungarian algorithm from combinatorial optimization to solve (15) and transforming the optimal P into the optimal P_s . Importantly, all these operations can be executed 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) and Gouriéroux

et al. (2020), 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 Section 3 for a comparison of alternative normalization procedures in simulation, and the Online Appendix for details and additional speed comparisons.

In a recent contribution, Keweloh and Wang (2026) use a refinement of the Lanne et al. (2017) normalization scheme that exploits a preliminary estimate of B , just as our approach. This normalization will likely often work well in practice and improve upon the original Lanne et al. (2017) normalization, but there is no guarantee that it will always normalize the parameter draw to be closest to the preliminary estimate, which our normalization guarantees by construction. In Appendix E, we show this by constructing an example. Computationally, there is no gain in speed by using the Keweloh and Wang (2026) approach instead of ours.

Although our approach is new to the VAR literature, it has some precedents. The distance measure used in our generalized LP normalization is closely related to the minimum distance index (MDI) proposed by Ilmonen et al. (2010). In fact, when structural shocks have normalized variances, the MDI and the distance measure for our generalized LP normalization become essentially equivalent. However, their intended use differs. The MDI was designed to evaluate the performance of independent component analysis (ICA) methods relative to the true model, where the main object of interest is the value of the distance itself. By contrast, we use the distance to implement a normalization procedure, not to evaluate estimator quality. The MDI is the standard benchmark for comparing ICA algorithms, as discussed in Matteson and Tsay (2017) and Moneta and Pallante (2022).

Because their goal is different from ours, the algorithm of Ilmonen et al. (2010) does not yield the signed permutation matrix that minimizes the distance, which is essential for our purposes. Their motivation for the MDI was to construct a distance metric that is invariant to scaling, sign changes, and permutations of the mixing matrix (corresponding to our B matrix if variances are not normalized). In contrast, our

distance measure is directly derived from the LP normalization. Interestingly, [Ilmonen et al. \(2010\)](#) also recognized that minimizing the MDI leads to a linear assignment problem, and proposed the Hungarian algorithm to solve it, as we do. This algorithm also appears in related ICA work on matching signal components across time windows, see [Tichavsky and Koldovsky \(2004\)](#), and in the popular LiNGAM algorithm of [Shimizu et al. \(2006\)](#). [Risk et al. \(2014\)](#) also rely on both the MDI and the algorithm of [Tichavsky and Koldovsky \(2004\)](#), and thus extensively use the Hungarian algorithm. In the statistics literature, the Hungarian algorithm has been used in addressing the label switching problem in finite mixture models. [Stephens \(2000\)](#) observed its potential to improve MCMC-based algorithms but did not pursue it further. [Cron and West \(2011\)](#) were the first to explicitly apply the Hungarian algorithm in that context, demonstrating its scalability for models with many components. See also [Lin et al. \(2013, 2016\)](#) for applications in hierarchical and classification mixture models.

Finally, the normalization procedure we develop here is not limited to our Bayesian framework. It can be applied with other priors on the impact matrix, or in frequentist contexts such as GMM estimation of non-Gaussian SVARs ([Lanne et al., 2017](#); [Lanne and Luoto, 2021](#); [Keweloh, 2021](#)).

2.5 Adding Identification Restrictions

We now show how to incorporate linear identification restrictions into our model. Strictly speaking, identifying restrictions are not needed in a non-Gaussian model, and introduce overidentification. However, identifying restrictions can still help inference, especially when the sample size is relatively small.

We focus on restrictions placed on A , which makes it straightforward to include instrument-based restrictions of the type proposed by [Plagborg-Møller and Wolf \(2021\)](#).¹⁸ More generally, any zero restriction – such as requiring A or B to be lower triangular – can be accommodated within our framework. To keep the conditional posterior distribution tractable, we only impose restrictions that operate on one row of A at a time. However, restrictions can be stacked and imposed jointly. In practice, we implement these restrictions when sampling U , one of the matrices in the decomposition of A . One prototypical restriction is given by:

¹⁸[Keweloh et al. \(2025\)](#) also combine identification via non-Gaussianity with instruments for structural shocks. In contrast to their approach, which builds a pseudo-likelihood for the moment conditions used for identification, we propose a fully Bayesian approach.

$$\mathbf{z}A\mathbf{s} = \mathbf{z}\Lambda LU\mathbf{s} = r, \quad (16)$$

where \mathbf{z} is a $1 \times k$ vector of known constants, \mathbf{s} is a $k \times 1$ vector of known constants, and r is a known scalar. Then all j linear restrictions imposed on the i -th row of A may be written:

$$\mathbf{z}_i A[\mathbf{s}_1 \ \mathbf{s}_2 \dots \mathbf{s}_j] = \mathbf{z}_i \Lambda LU[\mathbf{s}_1 \ \mathbf{s}_2 \dots \mathbf{s}_j] = [r_1 \ r_2 \dots r_j], \quad (17)$$

where $\mathbf{z}_i = [0 \dots 0 \ 1 \ 0 \dots 0]$, a vector of zeros except for a unit value in the i -th entry. Vectorizing this equation gives

$$([\mathbf{s}_1 \ \mathbf{s}_2 \dots \mathbf{s}_j]' \otimes \mathbf{z}_i \Lambda L) \text{vec}(U) = \text{vec}([r_1 \ r_2 \dots r_j]). \quad (18)$$

If the restrictions involve more than one row in A then we can just stack them one after another to get:

$$R_{lr} \cdot \text{vec}(U) = \mathbf{r}_{lr}, \quad (19)$$

where R_{lr} is a function of Λ, L . U in our algorithm is unit upper triangular. To be consistent with this assumption, we explicitly model these additional restrictions:

$$R_{up} \cdot \text{vec}(U) = \mathbf{r}_{up}, \quad (20)$$

where R_{up} is not a function of Λ, L .

We next stack all restrictions to arrive at:

$$\underbrace{\begin{bmatrix} R_{lr} \\ R_{up} \end{bmatrix}}_R \text{vec}(U) = \underbrace{\begin{bmatrix} \mathbf{r}_{lr} \\ \mathbf{r}_{up} \end{bmatrix}}_{\mathbf{r}}. \quad (21)$$

Since $\text{vec}(U)$ in our algorithm is conditionally multivariate Gaussian and we are imposing linear restrictions ($R \cdot \text{vec}(U) = \mathbf{r}$) on a Gaussian random vector, which is a well-studied problem in probability theory – the resulting constrained distribution is still Gaussian (see [Cong et al., 2017](#), and the references therein), and as such can be embedded in a computationally straightforward way in our Gibbs sampler.

3 Validation Using Simulated Data

We illustrate our methodology using a simulation based on a bivariate VAR(6) without a constant, where the structural shocks follow a t -distribution with unit variance. Specifically,

$$\mathbf{y}_t = \sum_{l=1}^6 \tilde{\Pi}_l \mathbf{y}_{t-l} + \tilde{B} \boldsymbol{\epsilon}_t, \quad (22)$$

$$p(\boldsymbol{\epsilon}_t) = \prod_{i=1}^2 p(\epsilon_{it}), \quad (23)$$

$$\epsilon_{it} \sim t(\tilde{v}_i). \quad (24)$$

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 parameterize 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

$$\tilde{\psi}_{ij,h} = a_{ij} \cdot \exp \left[- \left(\frac{(h - b_{ij})^2}{c_{ij}^2} \right) + \frac{b_{ij}^2}{c_{ij}^2} \right], \quad (25)$$

where a_{ij} governs the impact effect of shock j on variable i , b_{ij} is an integer that marks 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 up to horizon H coincide with the responses from equation (25). In our baseline analysis, we set $\tilde{v}_i = 6$ for $i = 1, 2$. See Appendix G of the Online Appendix for additional details on the data-generating process.

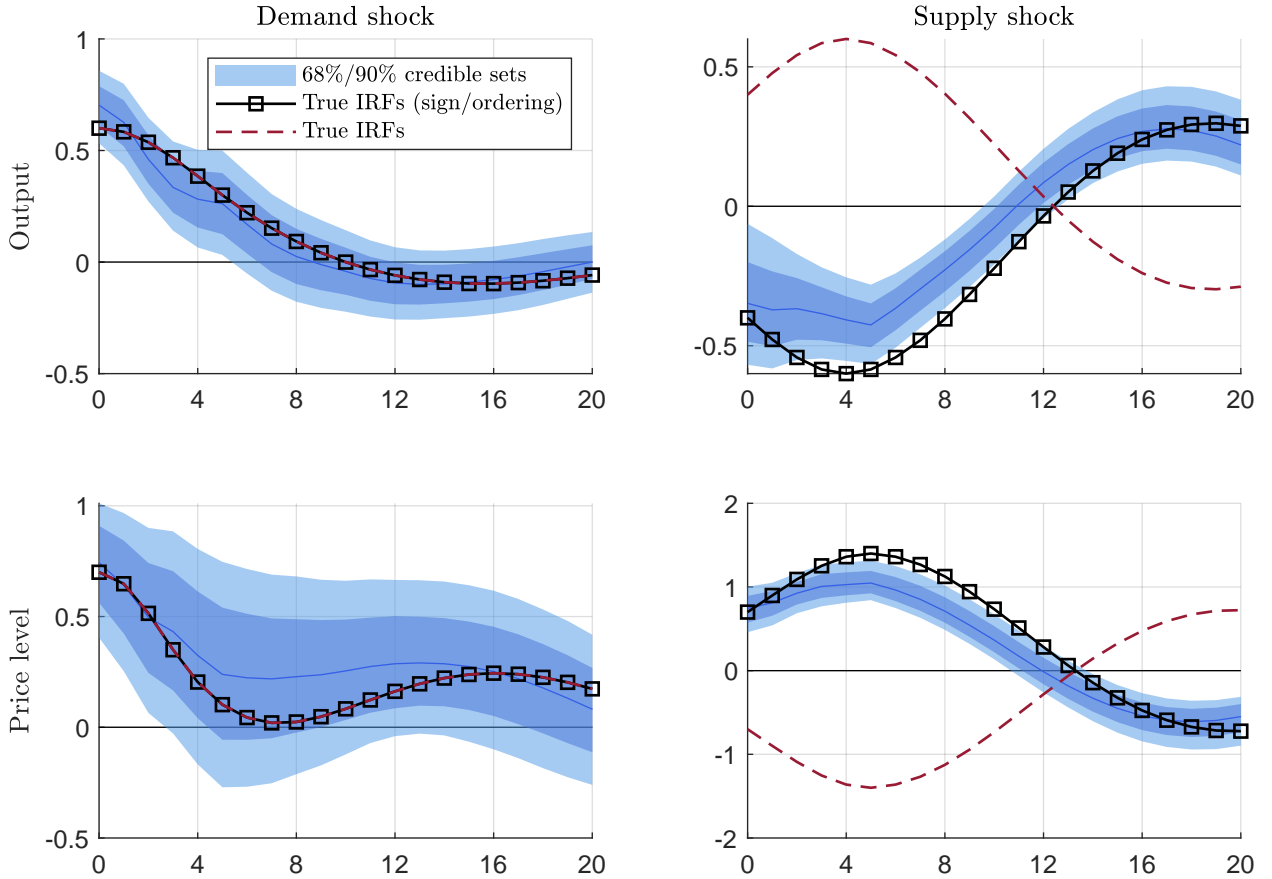
We first generate a single dataset from our specified data-generating process. We initialize the simulation at the unconditional mean (zero) and create 350 observations. We discard the first 100, retain the next 50 as a training sample, and use the remaining 200 for estimation. We then estimate a VAR with 6 lags (no constant) on the simulated data, applying the following prior. 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 to set the hyperparameters, in the spirit of the Minnesota prior). 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 (pre-truncation) variance 20 for each degrees-of-freedom parameter.

In the baseline simulations, we compute the target matrix \hat{B} required for the generalized LP normalization using the preliminary maximum likelihood estimator by [Lanne et al. \(2017\)](#) (see footnote 17) of the sampler. We initialize our Gibbs sampler as described in Appendix F of the Online Appendix. The sampler runs for 25,000 draws, discarding the first 5,000 as burn-in. On a standard computer with an Intel i7-7700K 4.2GHz Quad Core processor and 64GB RAM, this takes about eight minutes.

Figure 2 presents our methodology’s performance in estimating impulse responses. The solid blue line and shaded regions represent 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 once we decide on a sign normalization (such a normalization can always be changed ex post to aid interpretation). 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 precisely estimated.

Figure 3 illustrates how our sampler succeeds in recovering the true impact re-

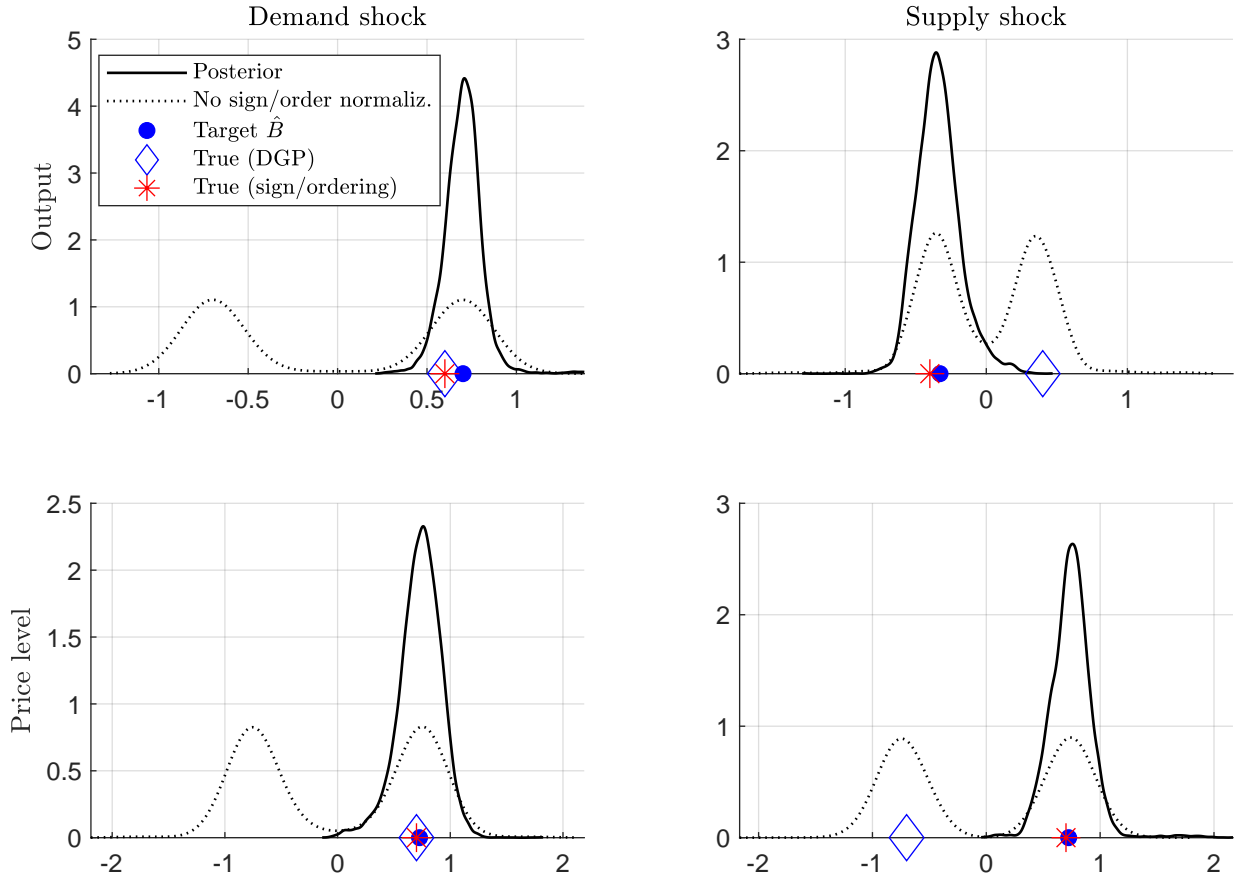
Figure 2: Impulse response functions



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/permutation of the true impulse responses that is closest to the impulse responses estimated in the posterior.

sponses. The solid line represents the marginal posterior distribution of each entry of B under our normalization procedure, which addresses sign and permutation indeterminacies. The dotted line shows the posterior when no normalization is applied. 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. In contrast, \hat{B} interprets a positively signed supply shock as *contracting* output, while the true positive supply shock actually increases output. The red star shows the true impact effect of the supply shock, but with its

Figure 3: Impact effect of the shocks: $p(B|Y)$

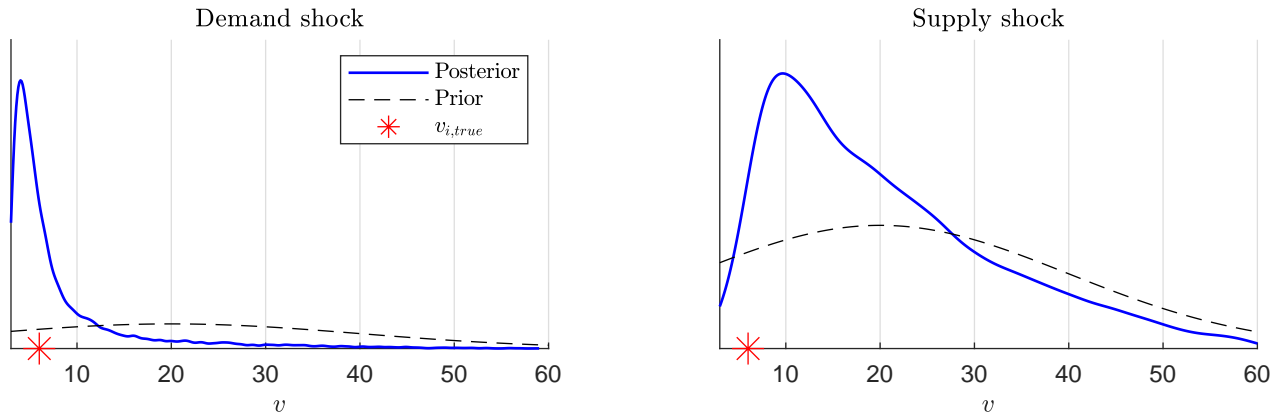


Note: The blue diamond indicates B_{true} . The blue dot indicates the target matrix \hat{B} used for the normalization. The red star indicates the sign/permutation of B_{true} that is the closest to \hat{B} . The continuous line shows the marginal posterior of the entries of B from 20,000 posterior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

sign flipped to match \hat{B} . After correcting for this sign, the model recovers the true impact of the shock.

Figure 4 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, hence ensuring that B is identified. The degrees of freedom of the demand shock are estimated more precisely than those of the supply shock. However, our estimated impulse responses show that the impulse responses are precisely estimated for

Figure 4: Degrees of freedom



Note: The red star 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.

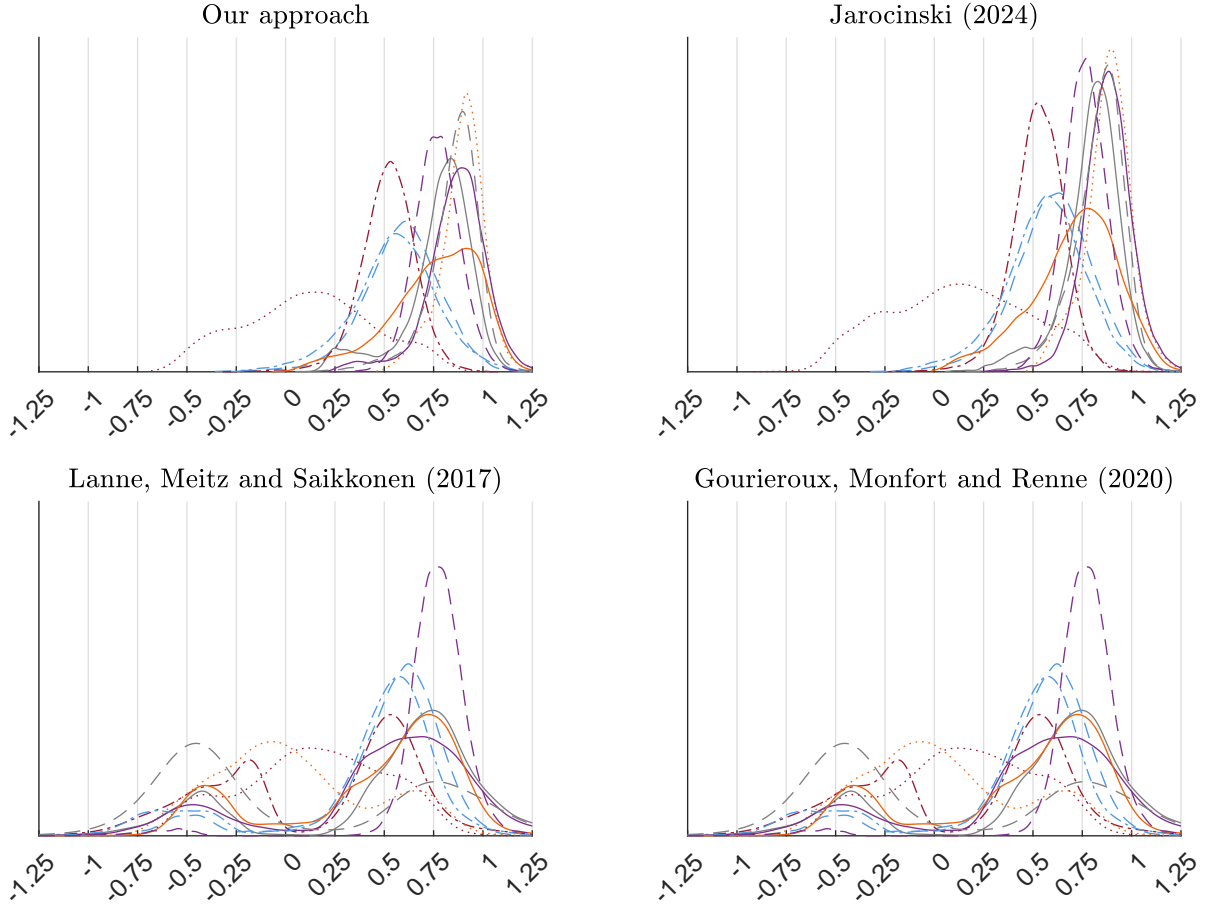
both shocks. Figure G-9 in the Online Appendix confirms that our ability to estimate \mathbf{v}_{true} improves with larger samples, as expected.

So far, we have illustrated our Bayesian procedure using a single dataset. We now illustrate an exercise that compares the performance of our generalized LP normalization with alternative methods. We keep the data-generating process fixed and generate ten datasets. For each dataset, we run the sampler four times, using either our normalization, or the normalizations by Lanne et al. (2017), Gouriéroux et al. (2020) or Jarociński (2024). The results are shown in Figure 5, which compares the estimated marginal posterior distributions associated with the effect of a positive demand shock on the price level (this corresponds to entry (2,1) of matrix B , see Figure G-3 – Figure G-6 in the Online Appendix for the full analysis). For each dataset, our approach and the approach by Jarociński (2024) are implemented using the same numerical ML estimate of \hat{B} , for comparability. The two methods give very similar results, displaying only minimal bimodality. Yet, our method can be used for large models, as discussed before, which is not true for the method by Jarociński (2024). In contrast, the remaining two methods occasionally deliver strong bimodality, an issue that can arise with some normalization schemes (Hamilton et al., 2007).¹⁹

We then extend the analysis to 100 datasets, and explore the ability of our model

¹⁹In this application, we found that the results from our approach are similar to those obtained when using the normalization by Keweloh and Wang (2026). We refer to Appendix E for a discussion of their approach.

Figure 5: Comparing alternative normalizations: (2,1) entry of matrix B :
Effect of a demand shock on the price level

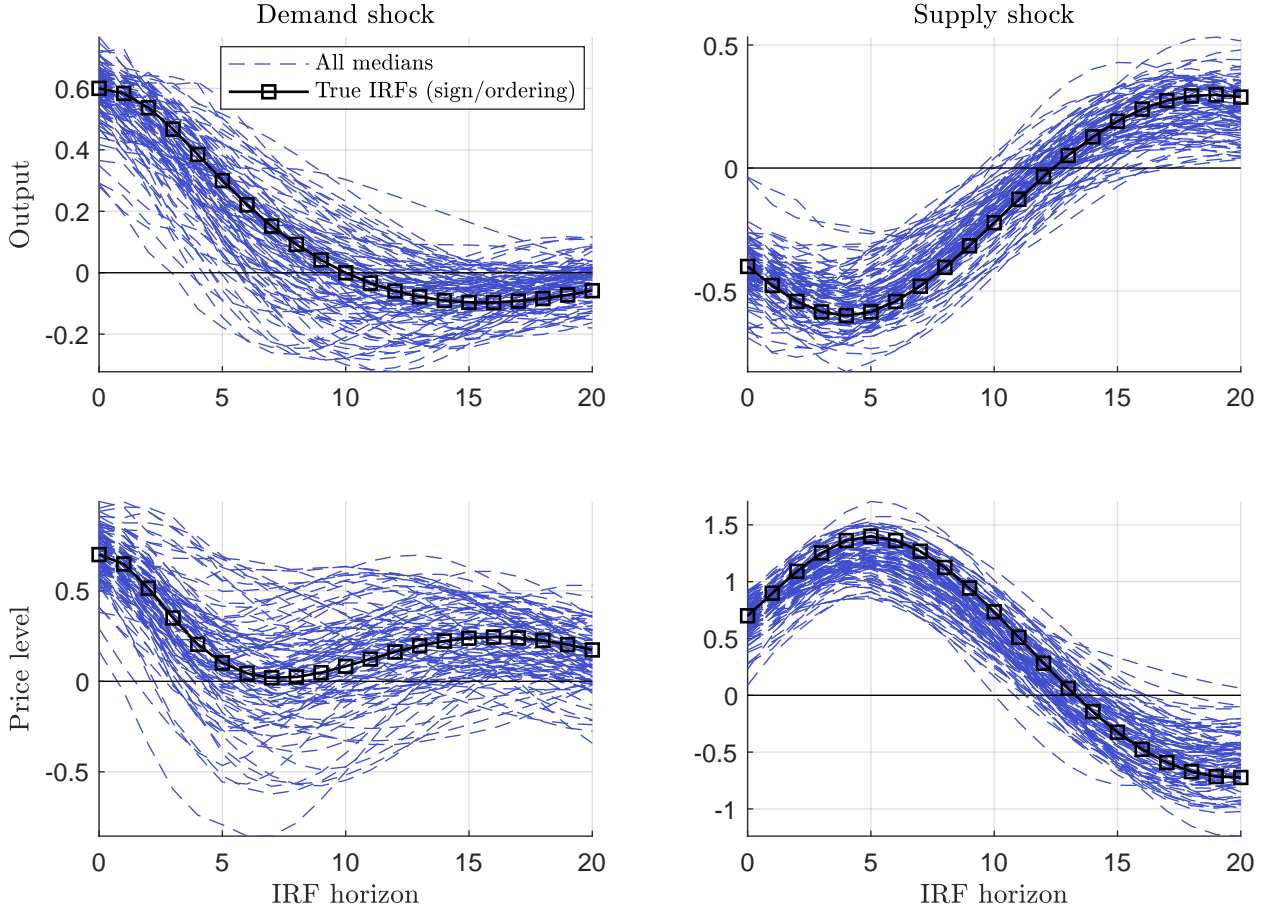


Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation. For each dataset, the sampler was run four times, applying each of the four normalizations. See Figure G-3 – Figure G-6 in the Online Appendix for the full set of results.

to capture the correct impulse responses. The black squared line in Figure 6 shows the true impulse response, matching the sign and ordering of the shocks in Figure 2. For each dataset, when necessary, we flip the order and signs of the target matrix \hat{B} so that they match the sign and ordering of the black squared line, to improve comparability. The blue lines in Figure 6 represent the pointwise medians of the estimated responses for each of the 100 replications. The figure indicates that our method reliably recovers the true dynamics.

We find that our approach accurately recovers impulse responses even when the

Figure 6: Impulse response functions
Robustness across 100 samples

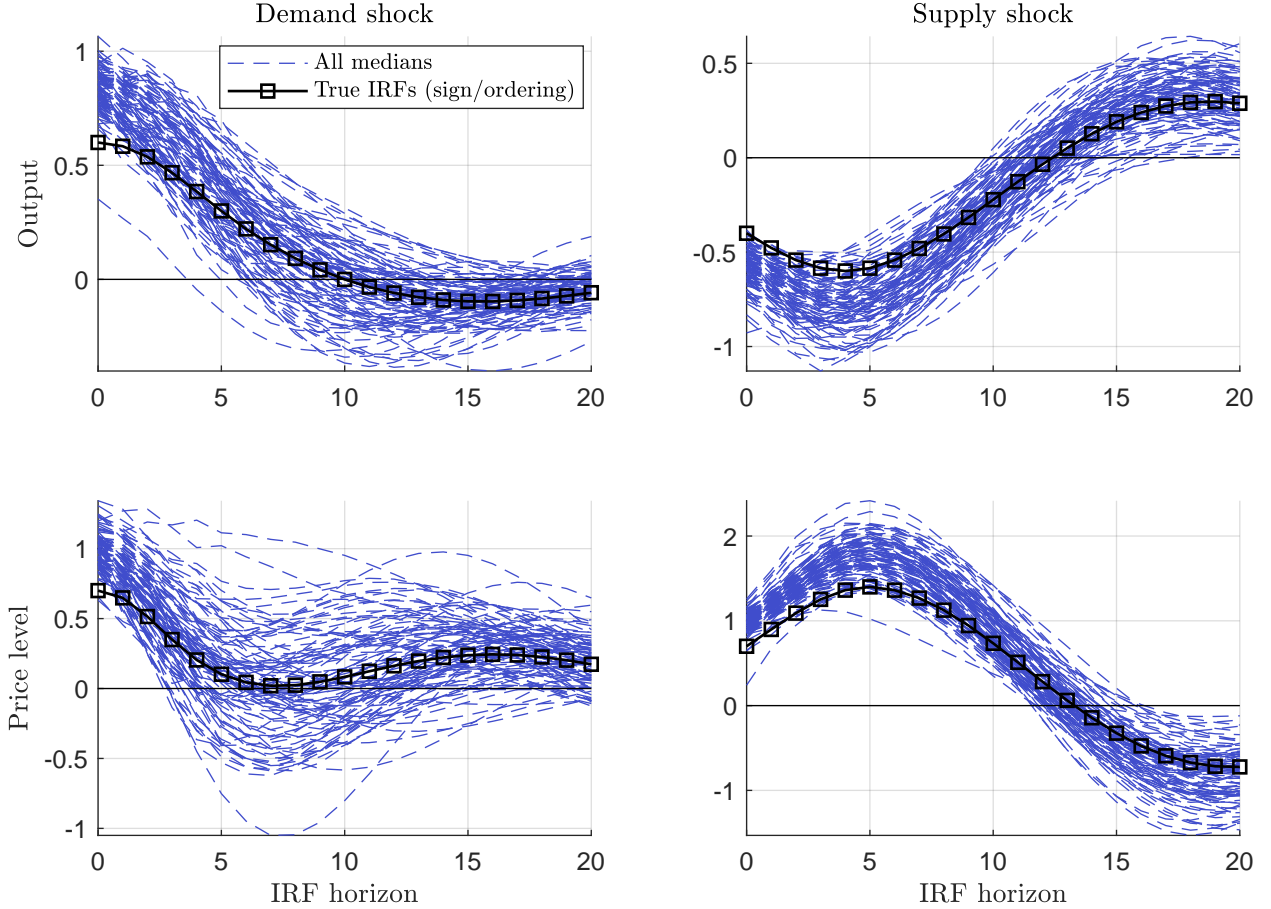


Note: The black squared line shows the true impulse responses, after changing the sign/ordering of the shocks as in Figure 2. The dashed blue lines show the median impulse responses across 100 datasets, with sign/ordering of the shocks changed when needed to improve comparability to the true impulse responses.

model is misspecified. Figure 7 replicates Figure 6, but now the data-generating process features Laplace-distributed shocks with unit standard deviation, while the estimated model still assumes t -distributed shocks. This exercise tests robustness to the specific distributional form. For comparability, we re-order and re-sign the columns of \hat{B} to align with the true impulse responses, as indicated by the black squared line. As the figure shows, the estimated impulse responses remain close to the truth. This result supports claims by Sims (2021) and Brunnermeier et al. (2021) (see their footnote 8): if shocks are independent, symmetric, and fat-tailed – even if not t -distributed – a

t -based structural VAR can still uncover the correct dynamics.

Figure 7: Impulse response functions
Robustness across 100 samples and DGP featuring Laplace shocks



Note: The black squared line shows the true impulse responses, after changing the sign/ordering of the shocks as in Figure 2. The dashed blue lines show the median impulse responses across 100 datasets, with sign/ordering of the shocks changed when needed to improve comparability to the true impulse responses.

We refer to Appendix G 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 (Figure G-7). We then show that the results are qualitatively unchanged relative to our benchmark if the target matrix is computed within the burn-in part of the sampler rather than with a preliminary ML estimation (Figure G-8), see footnote 17. We also increase the sample size (Figure G-9), use a flat prior on B^{-1} instead of B (Figure G-10), vary the

prior on the degrees of freedom (Figure G-11), and increasing the prior variance on ϕ (Figure G-12).

4 What Shocks Drive Real GDP?

We now apply our framework to assess which shocks matter most for U.S. real GDP. Angeletos et al. (2020) use a structural VAR with Gaussian shocks to identify a “Main Business Cycle” shock that explains much of the volatility in real activity at business-cycle frequencies. They find that (i) a single shock can account for nearly 80 percent of GDP volatility when GDP is used to measure real activity,²⁰ (ii) the shock resembles a weakly inflationary demand shock based on its impulse responses, (iii) it is largely uncorrelated with TFP, and (iv) it does not drive 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 follows the baseline Angeletos et al. (2020) model, using their exact dataset. This dataset includes 10 variables (GDP, investment, consumption, hours worked, unemployment rate, labor share, Federal funds rate, inflation, labor productivity, and TFP), all in log levels except for the Federal funds rate. All variables are real and in per capita terms where applicable. The sample is quarterly from 1955Q1 to 2017Q4, and, as in the original study, we include two lags.

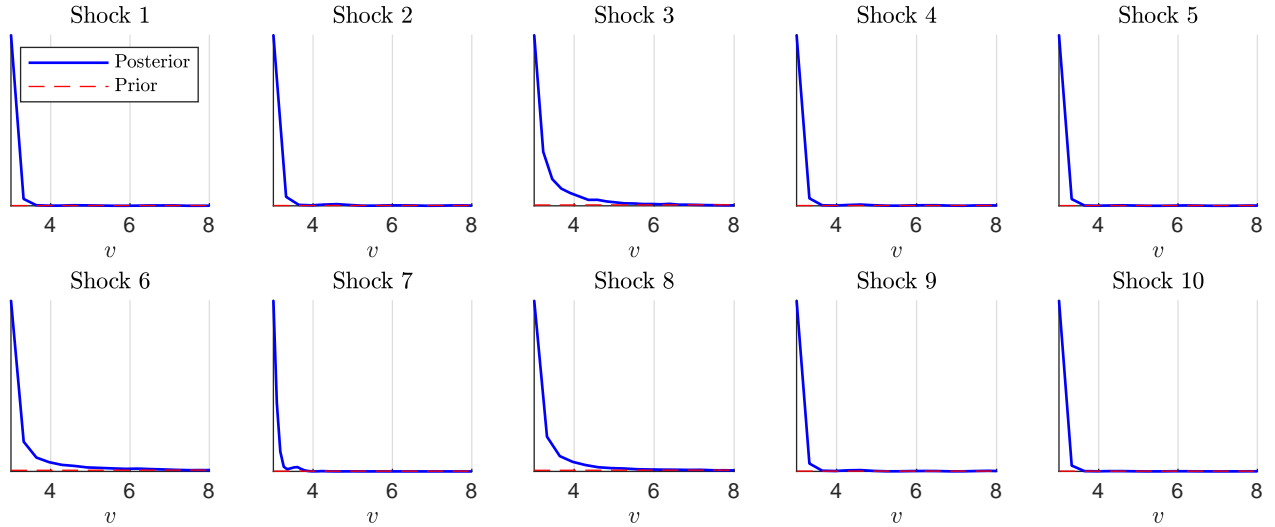
We specify the baseline prior for the analysis as follows. For the autoregressive parameters, we use the traditional Minnesota prior. For contemporaneous impulse responses B , we use a flat prior. For degrees of freedom, we use a Normal distribution centered at 20 and with variance 20, truncated to be positive between (3, 60). Posterior sampling is carried out using the method discussed in Section 2, generating 25,000 draws and discarding the first 10,000 draws. The marginal posterior distributions of B display unimodality, suggesting that the our generalized LP normalization is successfully avoiding the label switching problem (Figure H-13 in the Appendix).

Before interpreting our results, we first check the credibility of our identification assumption. In our framework, shocks are t -distributed, and we estimate their degrees of freedom as a natural measure of the identification strength. Lower degrees of freedom imply stronger deviations from Gaussianity. This contrasts with previous

²⁰When targeting unemployment, instead, their identified shock still explains more than half of the business-cycle variation in GDP.

studies, such as [Brunnermeier et al. \(2021\)](#), that fix degrees of freedom. Note that the degrees of freedom are generally identified in our model. [Figure 8](#) reports the posterior distributions for the degrees of freedom for all 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 8: Evidence of non-Gaussianity: prior and posterior on the degrees of freedom



Note: The dashed red 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 15,000 posterior draws.

Next, we turn to our main question: Is there one key shock driving 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:

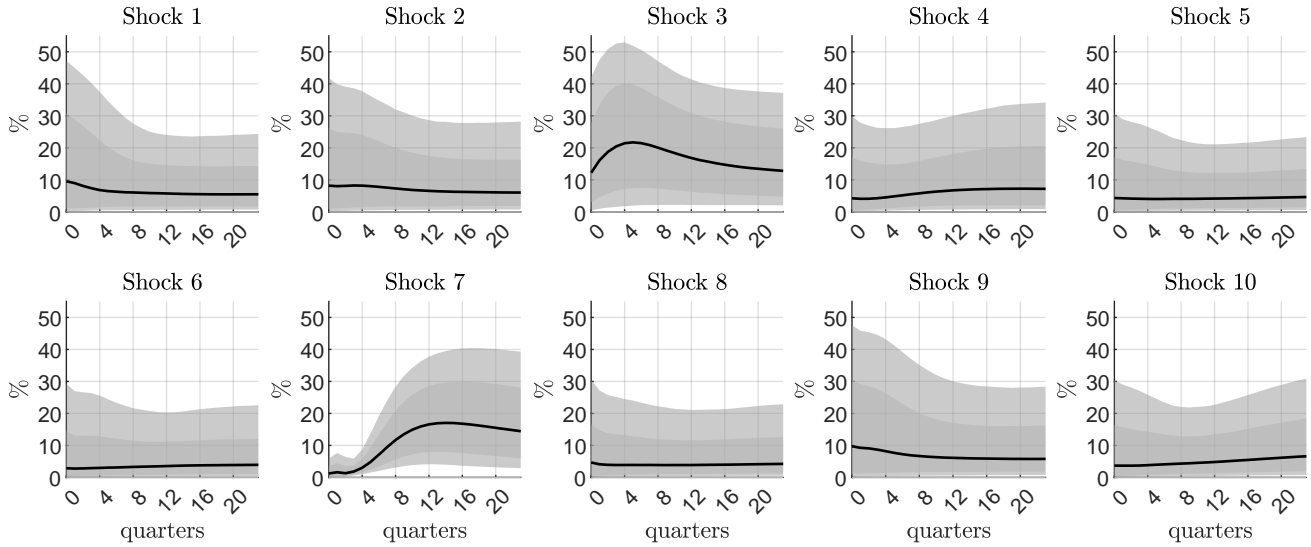
$$\text{FEVD}_{i,k}(h) = \frac{\sum_{m=0}^{h-1} (\mathbf{s}'_i \Theta_m \mathbf{s}_k)^2}{\sum_{m=0}^{h-1} \mathbf{s}'_i \Theta_m \Theta'_m \mathbf{s}_i}, \quad (26)$$

where Θ_m are the impulse response coefficients that can be obtained by repeated substitution into the VAR model.²¹ We report this time domain variance decomposition,

²¹For details on this derivation, see [Lütkepohl \(2005\)](#).

as it is arguably more common in the literature than the frequency domain version used for identification in [Angeletos et al. \(2020\)](#). However, it should be noted that [Angeletos et al. \(2020\)](#) report the time-domain decomposition in their paper when they discuss results, so our way of reporting results is directly comparable to theirs. Also, they find that in their model the frequency domain 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 in [Figure 9](#) and find that *no shock* explains more than 25% of the (posterior median) relative forecast error variance of real GDP at *any horizon*.²² 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 9: 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.

[Figure 10](#) 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 increases 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

²²In contrast, their [Figure 4](#) shows that their benchmark shock, which 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.

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. The full forecast error variance decompositions associated with the shocks from [Figure 10](#) is reported in [Figure H-14](#) of the Online Appendix.

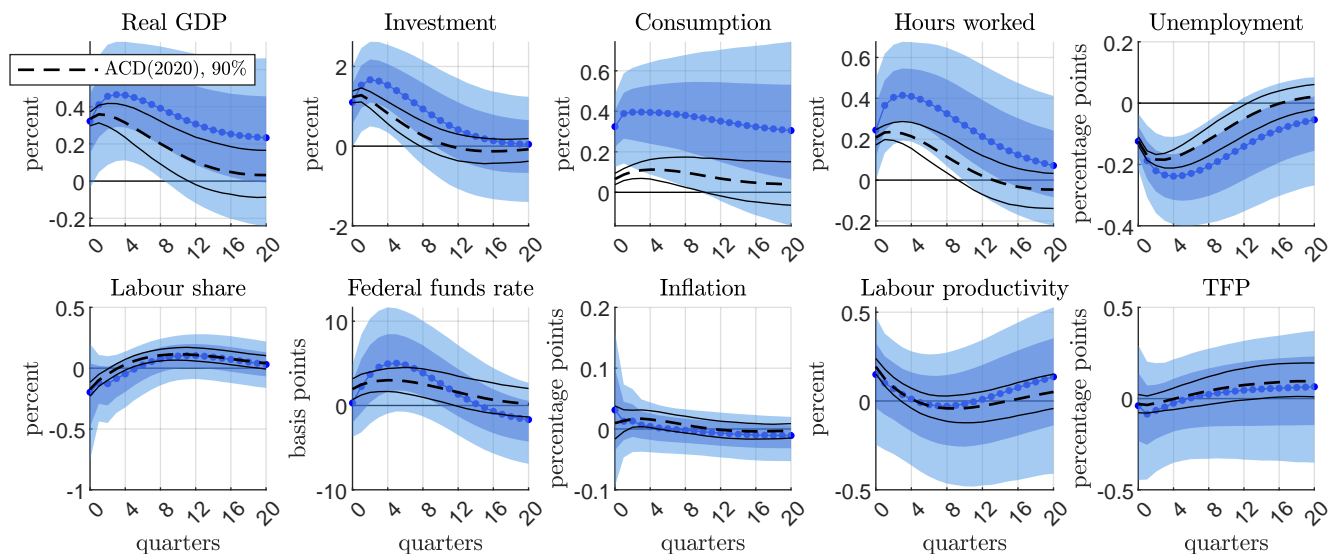
Before turning to robustness checks, one might wonder why our conclusions about the demand shock (shock 3) differ from those of [Angeletos et al. \(2020\)](#), despite using the same data and lag length, and recovering similar impulse responses. The key difference lies in our assumption of t -distributed shocks, which alters the shape of the likelihood and thus affects the entire posterior distribution. We find clear evidence of non-Gaussianity in the data, which leads to different parameter estimates. While we still recover a shock that resembles their main business cycle shock, the improved fit of our model reveals a second shock that also plays an important role in explaining GDP fluctuations.²³

In the Appendix, we show that the results do not change when computing the target matrix \hat{B} in the burn-in phase of the sampler rather than using a preliminary ML estimation ([H-15–H-16](#)). We also show that our findings are robust to several alternative specifications: a flat prior on A instead of a flat prior on $B = A^{-1}$ ([H-17–H-18](#)), a flat prior on the degrees of freedom ([H-19–H-20](#)), a looser prior on the autoregressive parameters ([H-21–H-22](#)), and a model with four lags ([H-23–H-24](#)).

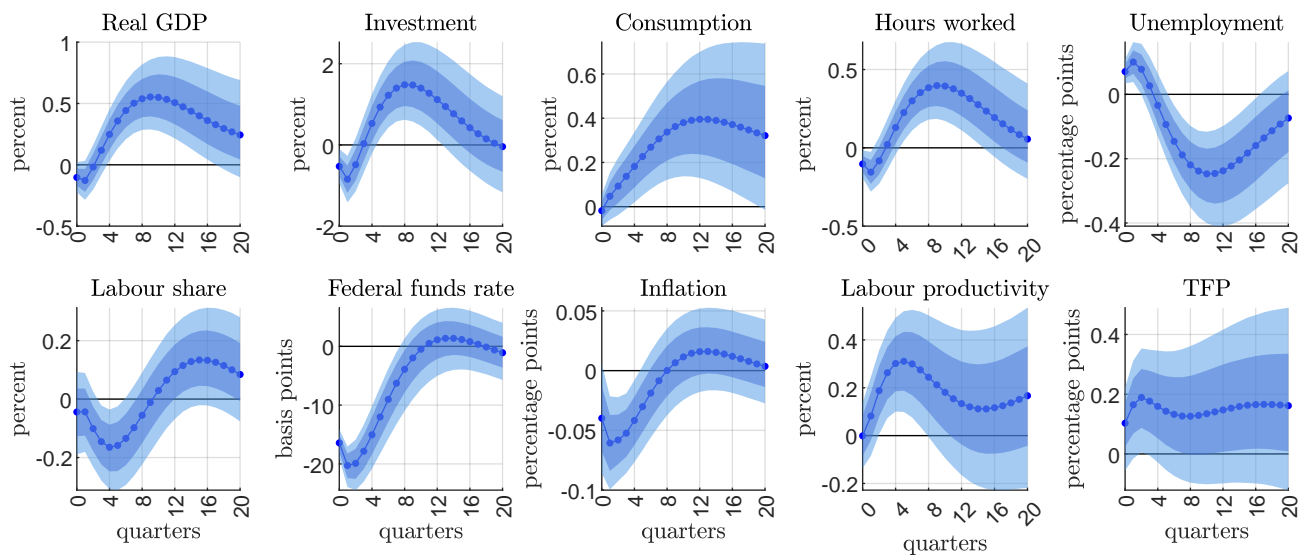
²³Our estimated degrees of freedom are low, confirming that the t -distribution improves model fit. If the data favored Gaussianity, the posterior would push degrees of freedom higher. Since our prior is weakly informative and leans towards Gaussianity, this result reflects genuine evidence from the data.

Figure 10: Impulse responses

Shock 3: demand shock



Shock 7: supply shock



Note: The blue dotted lines show the pointwise posterior median from our estimation, with corresponding 68% and 90% pointwise posterior credible sets shown with shaded areas. The black dashed and solid lines in the top figure show the pointwise median and 90% credible sets estimated by [Angeletos et al. \(2020\)](#) for the Main Business Cycle shock. To improve visibility, these were scaled to ensure that the median impact response of unemployment is the same as in our estimation.

4.1 Max-Share Identification Revisited

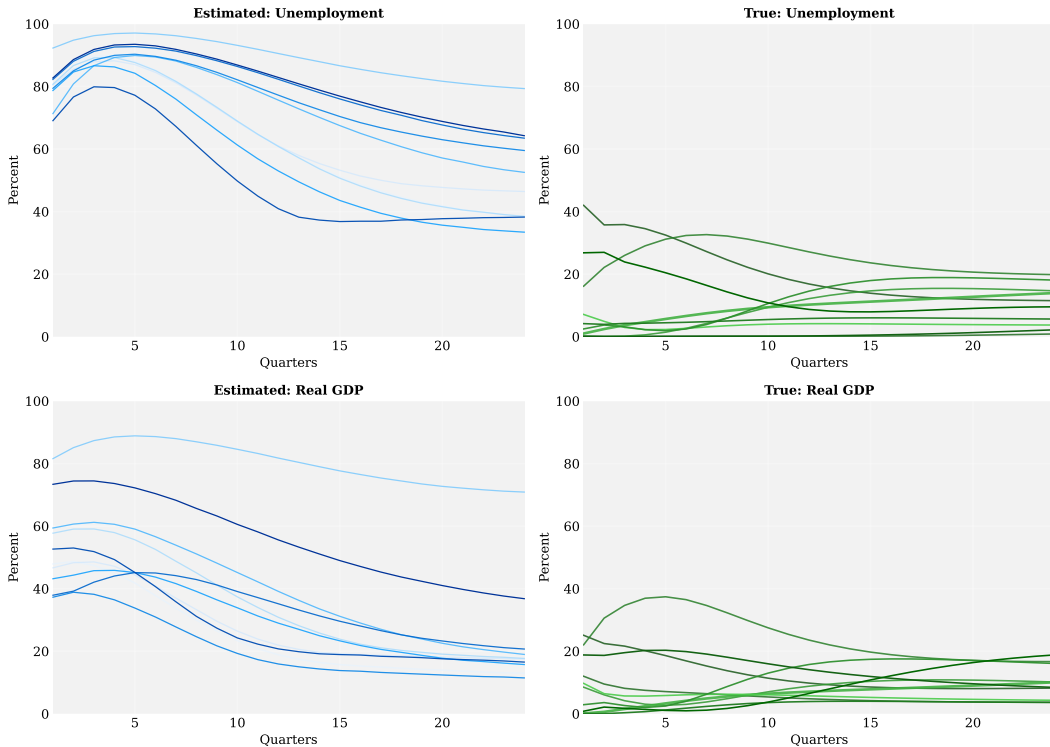
Finally, we ask what the [Angeletos et al. \(2020\)](#) approach would recover if our estimated model were the truth. To do so, we select the parameter draw from our benchmark specification that yields a forecast error variance decomposition (FEVD) closest to our median estimate,²⁴ using a quadratic loss. We treat these parameters as the data-generating process and simulate ten artificial datasets of the same length as the actual U.S. sample. We then estimate the [Angeletos et al. \(2020\)](#) model on each simulated dataset. As shown in the left panels of [Figure 11](#), their method identifies a dominant business cycle shock with properties similar to the one found in actual U.S. data.²⁵ Yet by construction, no such dominant shock exists in the true model, as can be seen in the right panels, which plot the FEVDs for the data-generating process. Importantly, the shock that delivers the largest true FEVD for unemployment is not the same shock that gives the largest true FEVD for real GDP.²⁶ This illustrates the masquerading problem discussed in the introduction.

²⁴To be precise, we choose the posterior draw as our data generating process that yields a FEVD for GDP for all shocks from horizon 1 to 24 that is as close as possible to our median estimate.

²⁵This is not surprising, given that our data-generating process is based on parameters estimated from the same dataset.

²⁶While we set the data-generating process by choosing the one posterior draw that yields FEVDs that most closely resemble the posterior median FEVDs, the fit is not perfect in that the panels on the right hand side of [Figure 11](#) do not exactly match the median FEVDs displayed in [Figure 9](#). This is expected because the posterior medians are estimated horizon-by-horizon. Similar issues have been pointed out for impulse responses by [Inoue and Kilian \(2022\)](#).

Figure 11: The [Angeletos et al. \(2020\)](#) approach in a non-Gaussian setting



Note: Left panels - Median FEVDs across 10 simulated datasets based on benchmark results, [Angeletos et al. \(2020\)](#) methodology, using a data-generating process with t -distributed shocks. Different shades of blue denote different simulated datasets. Right panels - True FEVDs of the relevant variable for all shocks.

5 Conclusions

This paper develops a new Gibbs sampler for structural VARs with non-Gaussian shocks. By reparameterizing the model, we show that all conditional posteriors take standard forms, removing the computational bottlenecks that have limited non-Gaussian SVARs to small systems. Our framework nests common specifications, including t -distributed shocks and stochastic volatility, and avoids strong theoretical restrictions by relying solely on statistical properties for identification. It also accommodates additional identifying assumptions, such as zero restrictions or external instruments. To address the sign and ordering indeterminacy inherent to non-Gaussian models, we introduce a fast and scalable normalization procedure.

Applied to U.S. data, our method shows that no single shock dominates real GDP volatility. Both demand and supply shocks matter, highlighting the importance of allowing for multiple sources of business cycle fluctuations.

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-Ramirez, J. F. and Shin, M. (2025), ‘A Gibbs sampler for efficient Bayesian inference in sign-identified SVARs’.
URL: <https://arxiv.org/abs/2505.23542>
- 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.
- Bañbura, 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.
- Bertsche, D. and Braun, R. (2022), ‘Identification of Structural Vector Autoregressions by Stochastic Volatility’, *Journal of Business & Economic Statistics* **40**(1), 328–341.

- Bianchi, F., Nicolò, G. and Song, D. (2023), ‘Inflation and real activity over the business cycle’.
- 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*.
- Carriero, A., Clark, T. and Marcellino, M. (2021), ‘Using time-varying volatility for identification in vector autoregressions: An application to endogenous uncertainty’, *Journal of Econometrics* **225**, 47–73.
- Carriero, A., Marcellino, M. and Tornese, T. (2024), ‘Blended identification in structural VARs’, *Journal of Monetary Economics* **146**, 103581.
URL: <https://www.sciencedirect.com/science/article/pii/S0304393224000345>
- 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.
- Cong, Y., Chen, B. and Zhou, M. (2017), ‘Fast Simulation of Hyperplane-Truncated Multivariate Normal Distributions’, *Bayesian Analysis* **12**(4), 1017 – 1037.
URL: <https://doi.org/10.1214/17-BA1052>
- Cron, A. J. and West, M. (2011), ‘Efficient Classification-Based Relabeling in Mixture Models’, *The American Statistician* **65**(1), 16–20.
- Cúrdia, V., Del Negro, M. and Greenwald, D. L. (2014), ‘Rare shocks, great recessions’, *Journal of Applied Econometrics* **29**(7), 1031–1052.
- Dou, L., Ho, P. and Lubik, T. A. (2025), Max-share misidentification. Working paper.
- Drautzburg, T. and Wright, J. H. (2023), ‘Refining set-identification in VARs through independence’, *Journal of Econometrics* **235**(2), 1827–1847.
- Eriksson, J. and Koivunen, V. (2004), ‘Identifiability, separability, and uniqueness of linear ICA models’, *IEEE Signal Processing Letters* **11**(7), 601–604.
- Fernandez, C. and Steel, M. (1999), ‘Multivariate Student-t regression models: Pitfalls and inference’, *Biometrika* **86**(1), 153–167.
- Forni, M., Gambetti, L., Granese, A., Sala, L. and Soccorsi, S. (2025), ‘An American Macroeconomic Picture: Supply and Demand Shocks in the Frequency Domain’, *American Economic Journal: Macroeconomics* **17**(3), 311–341.

- 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>
- Geweke, J. (1993), ‘Bayesian treatment of the independent student-t linear model’, *Journal of Applied Econometrics* **8**(S1), S19–S40.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, MIT Press. <http://www.deeplearningbook.org>.
- Gouriéroux, C., Monfort, A. and Renne, J.-P. (2020), ‘Identification and estimation in non-fundamental structural VAR models’, *The Review of Economic Studies* **87**(4), 1915–1953.
- 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>
- Ilmonen, P., Nordhausen, K., Oja, H. and Ollila, E. (2010), A new performance index for ICA: properties, computation and asymptotic analysis, in V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval and E. Vincent, eds, ‘Latent Variable Analysis and Signal Separation’, Springer, pp. 229–236.
- Inoue, A. and Kilian, L. (2020), ‘The role of the prior in estimating VAR models with sign restrictions’.
- Inoue, A. and Kilian, L. (2022), ‘Joint bayesian inference about impulse responses in VAR models’, *Journal of Econometrics* **231**(2), 457–476. Special Issue: The Econometrics of Macroeconomic and Financial Data.
URL: <https://www.sciencedirect.com/science/article/pii/S0304407621002475>

- 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.
- Kagan, A., Linnik, Y. V. and Rao, C. R. (1973), *Characterization Problems in Mathematical Statistics*, John Wiley & Sons Ltd.
- 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.
- Keweloh, S. A. (2021), ‘A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments’, *Journal of Business & Economic Statistics* **39**, 772–782.
- Keweloh, S. A., Klein, M. and Prüser, J. (2025), ‘Estimating Fiscal Multipliers by Combining Statistical Identification with Potentially Endogenous Proxies’, *Econometrics Journal*, *forthcoming* .
- Keweloh, S. A. and Wang, S. (2026), ‘Uncertain Short-Run Restrictions and Statistically Identified Structural Vector Autoregressions’, *Journal of Applied Econometrics* **41**(1), 12–25.
- 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. and Luoto, J. (2021), ‘GMM Estimation of Non-Gaussian Structural Vector Autoregression’, *Journal of Business & Economic Statistics* **39**, 69–81.
- Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010), ‘Structural vector autoregressions with Markov switching’, *Journal of Economic Dynamics and Control* **34**(2), 121–131.
- Lanne, M., Meitz, M. and Saikkonen, P. (2017), ‘Identification and estimation of non-gaussian structural vector autoregressions’, *Journal of Econometrics* **196**(2), 288–304.
- Lewis, D. J. (2021), ‘Identifying shocks via time-varying volatility’, *Review of Economic Studies* **88**, 3086–3124.
- Lewis, D. J. (2025), ‘Identification Based on Higher Moments in Macroeconometrics’, *Annual Review of Economics* **17**, 665–693.
- Lin, L., Chan, C., Hadrup, S. R., Froesig, T. M., Wang, Q. and West, M. (2013), ‘Hierarchical Bayesian mixture modelling for antigen-specific T-cell subtyping in combinatorially encoded flow cytometry studies’, *Statistical Applications in Genetics and Molecular Biology* **12**(3), 309–331.
- Lin, L., Chan, C. and West, M. (2016), ‘Discriminative variable subsets in Bayesian classification with mixture models, with application in flow cytometry studies’, *Biostatistics* **17**(1), 40–53.
- 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>
- Lütkepohl, H. and Netsunajev, A. (2017), ‘Structural vector autoregressions with smooth transition in variances’, *Journal of Economic Dynamics and Control* **84**, 43–57.

- Lütkepohl, H. and Netsūnajev, A. (2017), ‘Structural vector autoregressions with smooth transition in variances’, *Journal of Economic Dynamics and Control* **84**, 43–57.
- Lütkepohl, H. and Woźniak, T. (2020), ‘Bayesian inference for structural vector autoregressions identified by markov-switching heteroskedasticity’, *Journal of Economic Dynamics and Control* **113**, 103862.
- Matteson, D. S. and Tsay, R. S. (2017), ‘Independent Component Analysis via Distance Covariance’, *Journal of the American Statistical Association* **112**(518), 623–637.
- 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>
- Moneta, A. and Pallante, G. (2022), ‘Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study’, *Journal of Economic Dynamics and Control* **144**, 104530.
- Nguyen, H. D. and McLachlan, G. (2019), ‘On approximations via convolution-defined mixture models’, *Communications in Statistics - Theory and Methods* **48**(16), 3945–3955.
URL: <https://doi.org/10.1080/03610926.2018.1487069>
- Normandin, M. and Phaneuf, L. (2004), ‘Monetary policy shocks: Testing identification conditions under time-varying conditional volatility’, *Journal of Monetary Economics* **51**(6), 1217–1243.
- Plagborg-Møller, M. and Wolf, C. K. (2021), ‘Local Projections and VARs Estimate the Same Impulse Responses’, *Econometrica* **89**(2), 955–980.
- Prüser, J. (2024), ‘A large non-Gaussian structural VAR with application to monetary policy’.
URL: <https://arxiv.org/abs/2412.17598>

- Rigobon, R. (2003), ‘Identification through heteroskedasticity’, *The Review of Economics and Statistics* **85**(4), 777–792.
- Risk, B. B., Matteson, D. S., Ruppert, D., Eloyan, A. and Caffo, B. S. (2014), ‘An evaluation of independent component analyses with an application to resting-state fMRI’, *Biometrics* **70**(1), 224–236.
- Ritter, C. and Tanner, M. A. (1992), ‘Facilitating the Gibbs sampler: the Gibbs stopper and the gridy-Gibbs sampler’, *Journal of the American Statistical Association* **87**(419), 861–868.
- Rubio-Ramirez, J. F., Waggoner, D. F. and Zha, T. (2010), ‘Structural vector autoregressions: Theory of identification and algorithms for inference’, *The Review of Economic Studies* **77**(2), 665–696.
- Shimizu, S., Hoyer, P. O., Hyvarinen, A. and Kerminen, A. (2006), ‘A Linear Non-Gaussian Acyclic Model for Causal Discovery’, *Journal of Machine Learning Research* **7**(72), 2003–2030.
- 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.
- Stephens, M. (2000), ‘Dealing With Label Switching in Mixture Models’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **62**(4), 795–809.
- Tichavsky, P. and Koldovsky, Z. (2004), ‘Optimal pairing of signal components separated by blind techniques’, *IEEE Signal Processing Letters* **11**(2), 119–122.
- 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>

- Villani, M. (2009), ‘Steady-state priors for vector autoregressions’, *Journal of Applied Econometrics* **24**(4), 630–650.
- 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 Scalable Framework for Statistical
Identification in Structural VARs”**

A	Example of identification under non-Gaussianity	A-2
B	The model	B-9
C	Derivations of the conditional posterior distributions	C-11
C.1	Conditional posterior $p(\phi Y, B, D, v)$	C-11
C.2	Conditional posterior $p(D Y, \phi, B, v)$	C-12
C.3	Conditional posterior $p(v Y, \phi, B, D)$	C-13
C.4	Conditional posterior $p(B Y, \phi, D, v)$	C-14
D	Identification up to sign and ordering	D-18
E	A Comparison with the Keweloh and Wang (2026) Normalization	E-22
F	Summary of the algorithm to sample from the posterior of a SVAR with t-distributed structural shocks	F-24
G	Additional material on the simulation exercise	G-26
H	Additional material on the application to the US GDP	H-38

A Example of identification under non-Gaussianity

This section of the Online Appendix provides the details of the illustration in Section 2.4.1 of the paper, showing how non-Gaussianity affects the identification of SVAR models. We use the illustration not only to show that higher moments help identify the parameters of the model, but also to help visualize why, in this class of model, identification is achieved only up to the sign and ordering of the shock, a point that plays an important role in the paper. See also [Hoesch et al. \(2024\)](#) and [Lewis \(2025\)](#) for a related discussion of identification under non-Gaussianity.

Consider the following bivariate data generating process:

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

where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$ 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 implies that we assume the variance to be finite and thus $v_1, v_2 > 2$). Model (A-1) can be rewritten as

$$\mathbf{y}_t = B_c Q \epsilon_t, \quad (\text{A-2})$$

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 $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{11} \end{pmatrix}$. Our goal is to identify (B, \mathbf{v}) , or equivalently, (B_c, Q, \mathbf{v}) .

Equation $\Sigma = BB'$ implies the following second moment restrictions on the elements of B :

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

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

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

Because these three equations involve four unknowns, B_c is identified, but B is not ([Kilian and Lütkepohl, 2018](#)). While these restrictions hold under both Gaussianity and more generally with a t -distribution, these are the *only* model restrictions that can be used for the identification of B under Gaussianity. In fact, the Gaussian special

case of the model (which holds for $v_i = \infty$, $i = 1, 2$) implies that higher moments are constant in B , given that the full distribution of a Gaussian process is pinned down by the first two moments. Thus, in the Gaussian setting, only second moment restrictions are available from the model, and we would need to impose further restrictions to identify B (e.g., $b_{12} = 0$ to obtain $B = B_c$).

We now illustrate how non-Gaussianity helps with the identification of B . To do so, in this example we assume that higher moments exist, and show that information in these moments can be exploited. It is important to note, however, that identification via non-Gaussianity does *not* rely on the existence of higher moments, see e.g. [Kagan et al. \(1973\)](#), ch. 10, [Eriksson and Koivunen \(2004\)](#). We assume their existence only to streamline the exposition in this example. Since we assume symmetric distributions (Gaussian or t), third moments vanish, so any arguments based on moments has to use at least fourth moments.

In this bivariate example, the relevant expressions for the fourth moments (derived at the end of this section) are:

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

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

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

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

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

with $\kappa_i = \frac{3(v_i-2)}{v_i-4}$ (note that κ_i coincides with both the fourth moment and kurtosis of ϵ_{it} , namely $E(\epsilon_{it}^4)$, thanks to $E(\epsilon_{it}) = 0$ and $E(\epsilon_{it}^2) = 1$). When the data generating process is Gaussian, $v_1 = v_2 = \infty$ implies that $\kappa_1 = \kappa_2 = 3$, and all equations (A-4) become constant in the orthogonal matrix Q (i.e., in alternative choices of B such that $BB' = \Sigma$). Hence, the fourth moments provide no additional information to identify B . This is because under Gaussianity, the right-hand sides of (A-4) can be written exclusively as functions of the right-hand sides of the second-moment conditions in (A-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 left hand side from (A-4) vary with Q , given B_c . 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

$$B_{\text{true}} = \begin{pmatrix} 1 & -1.25 \\ 2 & 0.5 \end{pmatrix}. \quad (\text{A-5})$$

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 .¹

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

$$Q = Q(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad (\text{A-6})$$

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 use (A-3)-(A-4) to compute the fourth moments implied by $(\theta_{\text{true}}, v_{\text{true}})$.² Checking whether these fourth moments identify the model parameters amounts to asking if the system (A-4) has a solution other than $(\theta, v) = (\theta_{\text{true}}, v_{\text{true}}) = (0.29\pi, 6)$. By construction, any θ is consistent with the second moments alone.

Each of the three panels in Figure A-1 show the fourth moments from equation (A-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 (A-4) for $v = \infty$, hence $\kappa_i = 3$, $i = 1, 2$ (Gaussianity). Two results emerge from Panel A. First, that the fourth moments are constant in θ , showing that fourth moments do not provide

¹In Section 3 of the paper we examined our approach in finite samples, under a more realistic data-generating process.

²These are $E(y_{1t}^4) = 30.01$, $E(y_{1t}^3 y_{2t}) = 13.64$, $E(y_{1t}^2 y_{2t}^2) = 27.84$, $E(y_{1t} y_{2t}^3) = 41.06$ and $E(y_{2t}^4) = 102.38$.

any information when the estimated model is restricted to be Gaussian. Second, that a Gaussian model fails to match the fourth moments in the data. 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 (A-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 likelihood function of model under Gaussianity is uniquely pinned down by the first and second moments, higher moments provide no additional information, except for uncovering that the true fourth moments cannot be reconciled with a Gaussian model.

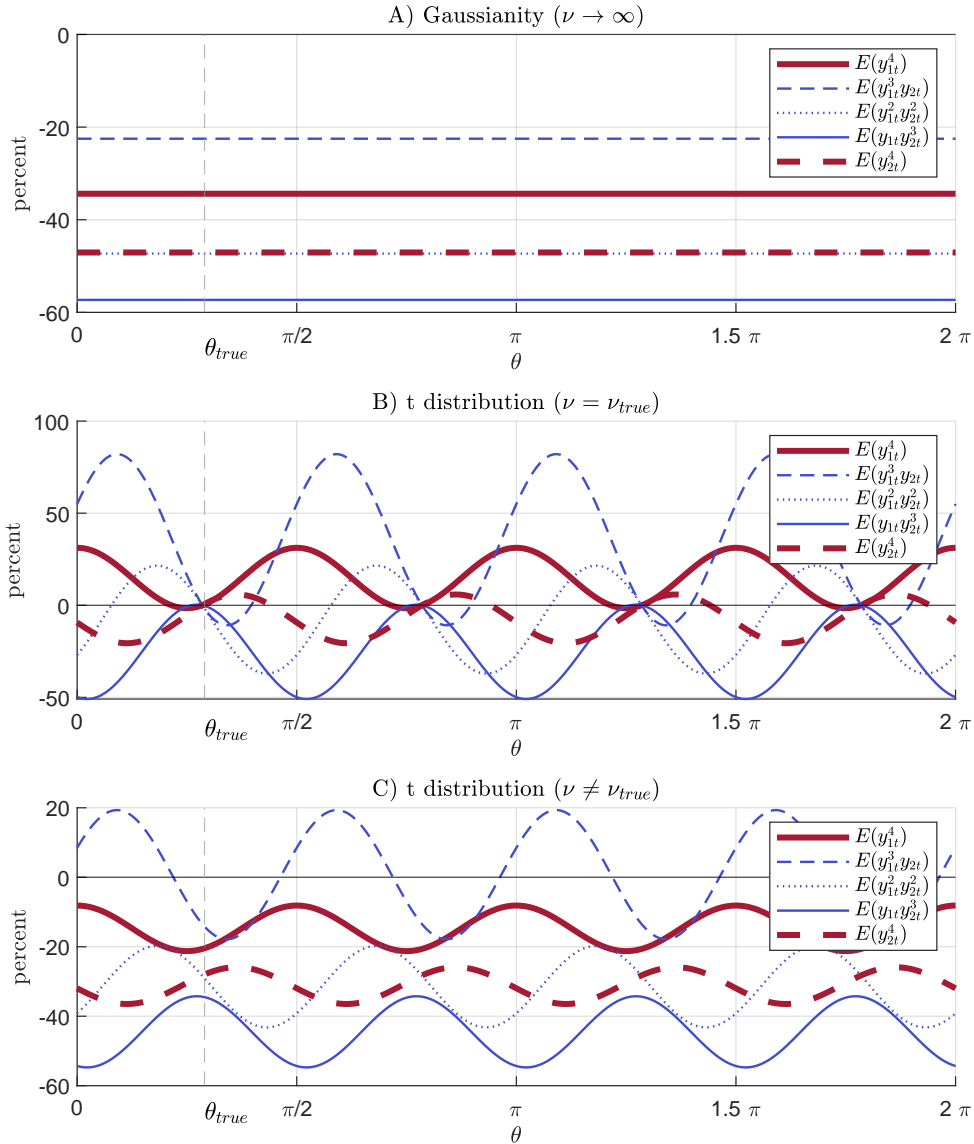
Things change dramatically when evaluating (A-4) using $\kappa_i = 6$ for $i = 1, 2$, which coincides with the true value. The middle panel of Figure A-1 shows that the fourth moments in (A-4) are no longer constant in θ , hence fourth moments provide additional information for identification. In addition, it shows that there are now exactly four values of θ that match the five moments from (A-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 A-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 A-1 clarifies how the four values of θ identified in Panel B of Figure A-1 relate to matrix 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 A-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 A-1 correspond to matrices a, f, g , and d in Table A-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

Figure A-1: Illustrative example



Note: Fourth moments are computed using equations (A-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 (A-6) over a grid on θ . The figure shows the moments in percentage deviation from the true values associated with (B_{true}, ν_{true}) . The top panel evaluates equations (A-4) for $\kappa_i = 3$, $i = 1, 2$, which is associated with a Gaussian model ($\nu_1 = \nu_2 = \infty$). The middle and lower panel set κ_i implied by $\nu = \nu_{true} = 6$ or $\nu = 9$.

(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 A-1 also shows why it is important to address the inference challenges posed

Table A-1: Modes of the likelihood of the model in Panel B from [Figure A-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.

by identification up to the sign and ordering of shocks. Although the eight modes in [Table A-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 ([Waggoner and Zha, 2003b](#); [Jarociński, 2024](#)). In the 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 large-scale systems.

We conclude this section by providing the derivations for the population moments ([A-4](#)) used above. The model is given by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}, \quad (\text{A-7})$$

where

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \quad (\text{A-8})$$

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

$i = 1, 2$. Then

$$E(y_{1t}^4) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})^4, \quad (\text{A-9})$$

$$= E(b_{11}^4\epsilon_{1t}^4 + 4b_{11}^3\epsilon_{1t}^3b_{12}\epsilon_{2t} + 6b_{11}^2\epsilon_{1t}^2b_{12}^2\epsilon_{2t}^2 + 4b_{11}\epsilon_{1t}b_{12}^3\epsilon_{2t}^3 + b_{12}^4\epsilon_{2t}^4), \quad (\text{A-10})$$

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

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

$$E(\epsilon_{it}^4) = \kappa_i = \frac{3(\nu_i - 2)}{\nu_i - 4}. \quad (\text{A-12})$$

Hence

$$E(y_{1t}^4) = b_{11}^4\kappa_1 + 6b_{11}^2b_{12}^2 + b_{12}^4\kappa_2, \quad (\text{A-13})$$

while similar derivations give

$$E(y_{2t}^4) = b_{21}^4\kappa_1 + 6b_{21}^2b_{22}^2 + b_{22}^4\kappa_2. \quad (\text{A-14})$$

Next:

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

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

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

$$= E(b_{11}^3\epsilon_{1t}^4b_{21} + 3b_{11}^2\epsilon_{1t}^2b_{12}^2\epsilon_{2t}^2b_{21} + 3b_{11}^2\epsilon_{1t}^2b_{12}\epsilon_{2t}^2b_{22} + b_{12}^3\epsilon_{2t}^4b_{22}), \quad (\text{A-18})$$

$$= b_{11}^3b_{21}\kappa_1 + 3b_{11}b_{12}^2b_{21} + 3b_{11}^2b_{12}b_{22} + b_{12}^3b_{22}\kappa_2. \quad (\text{A-19})$$

$$E(y_{1t}y_{2t}^3) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t})^3, \quad (\text{A-20})$$

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

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

$$= E(b_{11}b_{21}^3\epsilon_{1t}^4 + 3b_{11}\epsilon_{1t}^2b_{21}b_{22}^2\epsilon_{2t}^2 + 3b_{12}\epsilon_{2t}^2b_{21}^2\epsilon_{1t}^2b_{22} + b_{12}b_{22}^3\epsilon_{2t}^4), \quad (\text{A-23})$$

$$= b_{11}b_{21}^3\kappa_1 + 3b_{11}b_{21}b_{22}^2 + 3b_{12}b_{21}^2b_{22} + b_{12}b_{22}^3\kappa_2, \quad (\text{A-24})$$

$$E(y_{1t}^2y_{2t}^2) = E(b_{11}\epsilon_{1t} + b_{12}\epsilon_{2t})^2(b_{21}\epsilon_{1t} + b_{22}\epsilon_{2t})^2, \quad (\text{A-25})$$

$$= E(b_{11}^2\epsilon_{1t}^2 + 2b_{11}\epsilon_{1t}b_{12}\epsilon_{2t} + b_{12}^2\epsilon_{2t}^2)(b_{21}^2\epsilon_{1t}^2 + 2b_{21}\epsilon_{1t}b_{22}\epsilon_{2t} + b_{22}^2\epsilon_{2t}^2), \quad (\text{A-26})$$

$$= E(b_{11}^2\epsilon_{1t}^2b_{21}^2\epsilon_{1t}^2 + b_{12}^2\epsilon_{2t}^2b_{21}^2\epsilon_{1t}^2 + \\ + 4b_{11}\epsilon_{1t}b_{12}\epsilon_{2t}b_{21}\epsilon_{1t}b_{22}\epsilon_{2t} + \\ + b_{11}^2\epsilon_{1t}^2b_{22}^2\epsilon_{2t}^2 + b_{12}^2\epsilon_{2t}^2b_{22}^2\epsilon_{2t}^2), \quad (\text{A-27})$$

$$= b_{11}^2b_{21}^2\kappa_1 + b_{12}^2b_{21}^2 + 4b_{11}b_{12}b_{21}b_{22} + b_{11}^2b_{22}^2 + b_{12}^2b_{22}^2\kappa_2. \quad (\text{A-28})$$

B The model

The model is given by

$$\mathbf{y}_t = \sum_{l=1}^p \Pi_l \mathbf{y}_{t-l} + \mathbf{c} + U_t, \quad (\text{B-29})$$

$$= \Pi \mathbf{x}_t + U_t, \quad (\text{B-30})$$

$$U_t = B\epsilon_t, \quad (\text{B-31})$$

$$p(\epsilon_t | \boldsymbol{\sigma}, \mathbf{v}) = \prod_{i=1}^k p(\epsilon_{it} | \sigma_i, v_i), \quad (\text{B-32})$$

$$\epsilon_{it} \sim t(\sigma_i, v_i), \quad (\text{B-33})$$

$$p(\epsilon_{it} | \sigma_i, v_i) = \sigma_i^{-1} \cdot v_i^{-\frac{1}{2}} \cdot \frac{\Gamma\left(\frac{v_i+1}{2}\right)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{v_i}{2}\right)} \cdot \left(1 + \frac{\epsilon_{it}^2}{v_i \cdot \sigma_i^2}\right)^{-\frac{v_i+1}{2}}. \quad (\text{B-34})$$

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 (B-34), with (σ_i, v_i) the shock-specific scale and degrees of freedom, and $\boldsymbol{v} = (v_1, \dots, v_i, \dots, v_k)'$, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_i, \dots, \sigma_k)'$. As in Brunnermeier et al. (2021) (see their footnote 11), we set the scale parameter to:

$$\sigma_i = \sqrt{\frac{v_i - 2}{v_i}}, \quad (\text{B-35})$$

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

$$V(\epsilon_{it}) = 1. \quad (\text{B-36})$$

Under this normalization, the $k \times k$ matrix B captures the impact effect of a one standard deviation shocks.³

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

$$\boldsymbol{y}_t = \sum_{l=1}^p \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{c} + U_t, \quad (\text{B-37})$$

$$= \Pi \boldsymbol{x}_t + U_t, \quad (\text{B-38})$$

$$U_t = B \sqrt{D_t} \boldsymbol{e}_t, \quad (\text{B-39})$$

$$\boldsymbol{e}_t \sim N(\mathbf{0}, I), \quad (\text{B-40})$$

$$D_t = \text{diag}(d_{1t}, \dots, d_{it}, \dots, d_{kt}), \quad (\text{B-41})$$

$$p(d_{it} | h_{d,i}, r_{d,i}) = \frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}}, \quad (\text{B-42})$$

$$h_{d,i} = \frac{v_i}{2}, \quad r_{d,i} = \frac{v_i - 2}{2}. \quad (\text{B-43})$$

The k stochastic terms in the new specification, \boldsymbol{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

³This 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 2.4 of the paper and Appendix D.

(B-42), 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 (B-43), which comply with the normalization $V(\epsilon_{it}) = 1$ made in the original model specification.

Rewrite the parametrization of the model as

$$\boldsymbol{\phi} = \text{vec}(\Pi), \tag{B-44}$$

$$D = \text{diag}(D_1, \dots, D_t, \dots, D_T). \tag{B-45}$$

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

$$p(\boldsymbol{\phi}, B, D, \mathbf{v}) = p(\boldsymbol{\phi}) \cdot p(B) \cdot p(D|\mathbf{v}) \cdot p(\mathbf{v}). \tag{B-46}$$

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 (B-42), with hyperparameters calibrated according to (B-43). By contrast, the prior on (B, \mathbf{v}) can be more flexibly selected by the researcher, as discussed below.

C Derivations of the conditional posterior distributions

Posterior sampling is achieved by means of a Gibbs sampler. The conditional posteriors for the parameters $(\boldsymbol{\phi}, D, \mathbf{v})$ 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 $p(\boldsymbol{\phi}|Y, B, D, \mathbf{v})$

The likelihood function of model (B-37)-(B-40) can be written as

$$p(Y|\boldsymbol{\phi}, B, D) = (2\pi)^{-\frac{Tk}{2}} \cdot |B|^{-T} \cdot |D|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}} - W\boldsymbol{\phi})' \left((I_T \otimes B) D (I_T \otimes B') \right)^{-1} (\tilde{\mathbf{y}} - W\boldsymbol{\phi})}, \tag{C-47}$$

with

$$Y = [\mathbf{y}_1, \dots, \mathbf{y}_t, \dots, \mathbf{y}_T], \quad (\text{C-48})$$

$$\tilde{\mathbf{y}} = \text{vec}(Y), \quad (\text{C-49})$$

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T], \quad (\text{C-50})$$

$$W = (X' \otimes I_k). \quad (\text{C-51})$$

Using (C-47), we can derive

$$p(\boldsymbol{\phi}|Y, B, D, \mathbf{v}) \propto p(\boldsymbol{\phi}) \cdot p(Y|\boldsymbol{\phi}, B, D), \quad (\text{C-52})$$

$$\propto e^{-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\mu})'V^{-1}(\boldsymbol{\phi}-\boldsymbol{\mu})} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}}-W\boldsymbol{\phi})'((I_T \otimes B)D(I_T \otimes B'))^{-1}(\tilde{\mathbf{y}}-W\boldsymbol{\phi})}, \quad (\text{C-53})$$

$$= e^{-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\mu})'V^{-1}(\boldsymbol{\phi}-\boldsymbol{\mu})} \cdot e^{-\frac{1}{2}(\tilde{\mathbf{y}}-W\boldsymbol{\phi})'\Omega^{-1}(\tilde{\mathbf{y}}-W\boldsymbol{\phi})}, \quad (\text{C-54})$$

$$\propto e^{-\frac{1}{2}(\boldsymbol{\phi}'V^{-1}\boldsymbol{\phi}-2\boldsymbol{\phi}'V^{-1}\boldsymbol{\mu}+\boldsymbol{\phi}'W'\Omega^{-1}W\boldsymbol{\phi}-2\boldsymbol{\phi}'W'\Omega^{-1}\tilde{\mathbf{y}})}, \quad (\text{C-55})$$

$$\boldsymbol{\phi}|Y, B, D, \mathbf{v} \sim N(\boldsymbol{\mu}^*, V^*), \quad (\text{C-56})$$

$$V^* = (V^{-1} + W'\Omega^{-1}W)^{-1}, \quad (\text{C-57})$$

$$\boldsymbol{\mu}^* = V^*[V^{-1}\boldsymbol{\mu} + W'\Omega^{-1}\tilde{\mathbf{y}}], \quad (\text{C-58})$$

$$\Omega = (I_T \otimes B)D(I_T \otimes B'). \quad (\text{C-59})$$

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

The conditional posterior for D is also standard. After defining

$$\mathbf{g}_t = B^{-1}[\mathbf{y}_t - \Pi\mathbf{x}_t], \quad (\text{C-60})$$

model (B-37)-(B-40) coincides with $\mathbf{g}_t \sim N(\mathbf{0}, D_t)$, hence

$$p(D|Y, \phi, B, \mathbf{v}) = \prod_{t=1}^T \prod_{i=1}^k p(d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, v_i), \quad (\text{C-61})$$

$$p(d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, v_i) \propto p(d_{it}|h_{d,i}, r_{d,i}) \cdot p(\mathbf{g}_t|\Pi, B, \mathbf{v}, D_t), \quad (\text{C-62})$$

$$\propto \left[\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right] \cdot d_{it}^{-\frac{1}{2}} \cdot e^{-\frac{1}{2} \frac{g_{it}^2}{d_{it}}}, \quad (\text{C-63})$$

$$\propto d_{it}^{-(h_{d,i} + \frac{1}{2})-1} \cdot e^{-(r_{d,i} + \frac{g_{it}^2}{2}) \frac{1}{d_{it}}}, \quad (\text{C-64})$$

$$= d_{it}^{-h_{d,i}^* - 1} \cdot e^{-r_{d,it}^* \frac{1}{d_{it}}}, \quad (\text{C-65})$$

$$d_{it}|\mathbf{y}_t, \mathbf{x}_t, \Pi, B, \mathbf{v} \sim i\Gamma(h_{d,i}^*, r_{d,it}^*), \quad (\text{C-66})$$

$$h_{d,i}^* = h_{d,i} + \frac{1}{2}, \quad (\text{C-67})$$

$$r_{d,it}^* = r_{d,i} + \frac{g_{it}^2}{2}. \quad (\text{C-68})$$

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

The conditional posterior for \mathbf{v} is

$$p(\mathbf{v}|Y, \Pi, B, D) \propto p(D|\mathbf{v}) \cdot p(\mathbf{v}), \quad (\text{C-69})$$

$$\propto \left[\prod_{t=1}^T \prod_{i=1}^k \frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \cdot d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right] \cdot p(\mathbf{v}), \quad (\text{C-70})$$

$$= \left[\prod_{i=1}^k \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \prod_{t=1}^T \left(d_{it}^{-h_{d,i}-1} e^{-r_{d,i} \cdot \frac{1}{d_{it}}} \right) \right] \cdot p(\mathbf{v}), \quad (\text{C-71})$$

$$= \left[\prod_{i=1}^k \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \left(\prod_{t=1}^T d_{it} \right)^{-h_{d,i}-1} e^{-r_{d,i} \cdot \left(\sum_{t=1}^T \frac{1}{d_{it}} \right)} \right] \cdot p(\mathbf{v}). \quad (\text{C-72})$$

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:

$$p(\mathbf{v}|Y, \Pi, B, D) = \prod_{i=1}^k p(v_i|Y, \Pi, B, D), \quad (\text{C-73})$$

$$p(v_i|Y, \Pi, B, D) \propto \left(\frac{r_{d,i}^{h_{d,i}}}{\Gamma(h_{d,i})} \right)^T \cdot \left(\prod_{t=1}^T d_{it} \right)^{-h_{d,i}-1} e^{-r_{d,i} \cdot \left(\sum_{t=1}^T \frac{1}{d_{it}} \right)} \cdot p(v_i). \quad (\text{C-74})$$

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

$$\mathbf{z}_t = \mathbf{y}_t - \Pi \mathbf{x}_t, \quad (\text{C-75})$$

model (B-37)-(B-40) coincides with $\mathbf{z}_t \sim N(\mathbf{0}, A^{-1}D_tA^{-1})$, hence the likelihood function is

$$p(Z|\Pi, A, D) = (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot \left(\prod_{t=1}^T |D_t|^{-\frac{1}{2}} \right) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' A' D_t^{-1} A \mathbf{z}_t}, \quad (\text{C-76})$$

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

We rewrite A as

$$A = \Lambda L U. \quad (\text{C-77})$$

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 (\mathbf{l}, \mathbf{u}) , respectively. We use the notation

$$\text{vec}(L) = \mathbf{s} + S_L \mathbf{l}, \quad (\text{C-78})$$

$$\text{vec}(U) = \mathbf{s} + S_U \mathbf{u}, \quad (\text{C-79})$$

with (\mathbf{l}, \mathbf{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

$$|A| = \prod_{i=1}^k |\lambda_i|. \quad (\text{C-80})$$

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:

$$p_A(A) \propto 1, \quad (\text{C-81})$$

$$p_A(A) \propto |A|^{-2k}. \quad (\text{C-82})$$

(C-81) is the case in which the researcher expresses an uninformative prior directly on A (hence implicitly the SVAR uses A as its basic parameter). (C-82) 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) , (C-82) is only a function of λ_i 's via equation (C-80).

Combining (C-76)-(C-81)-(C-82) gives the conditional posterior for A :

$$p(A|Y, \phi, D, \mathbf{v}) \propto p_A(A) \cdot |A|^T \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t A' D_t^{-1} A \mathbf{z}_t}. \quad (\text{C-83})$$

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

$$J(A \rightarrow \Lambda, L, U) = \prod_{i=1}^k |\lambda_i|^{k-1}. \quad (\text{C-84})$$

Hence, (C-83) implies the following conditional posterior jointly for (Λ, L, U) :

$$p(\Lambda, L, U|Y, \phi, D, \mathbf{v}) \propto \left(\prod_{i=1}^k |\lambda_i|^{k-1} \right) \cdot p_A(\Lambda L U) \cdot \left(\prod_{i=1}^k |\lambda_i|^T \right) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{C-85})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda L U) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}. \quad (\text{C-86})$$

Our approach to sample from $p(\Lambda, L, U|Y, \phi, D, \mathbf{v})$ consists in showing that the conditional posteriors of Λ , L and U all have a common form. Starting with L , under either (C-81) or (C-82) we can derive

$$p(L|Y, \phi, D, \mathbf{v}, \Lambda, U) \propto e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{C-87})$$

$$= e^{-\frac{1}{2} \text{vec}(L)' \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}'_t U' \otimes \Lambda D_t^{-1} \Lambda) \text{vec}(L)}, \quad (\text{C-88})$$

$$= e^{-\frac{1}{2} \text{vec}(L)' \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}'_t U' \otimes \Lambda^2 D_t^{-1}) \text{vec}(L)}, \quad (\text{C-89})$$

$$p(\mathbf{l}|Y, \phi, D, \mathbf{v}, \Lambda, U) \propto e^{-\frac{1}{2} (\mathbf{s} + S_L \mathbf{l})' W_L (\mathbf{s} + S_L \mathbf{l})}, \quad (\text{C-90})$$

$$\propto e^{-\frac{1}{2} (\mathbf{l}' S'_L W_L S_L \mathbf{l} + 2 \mathbf{l}' S'_L W_L \mathbf{s})}, \quad (\text{C-91})$$

hence,

$$\mathbf{l}|Y, \phi, D, \mathbf{v}, \Lambda, U \sim N(\boldsymbol{\mu}_L^*, V_L^*), \quad (\text{C-92})$$

$$V_L^* = (S'_L W_L S_L)^{-1} \quad (\text{C-93})$$

$$\boldsymbol{\mu}_L^* = -V_L^* S'_L W_L \mathbf{s}, \quad (\text{C-94})$$

$$W_L = \sum_{t=1}^T (U \mathbf{z}_t \mathbf{z}'_t U' \otimes \Lambda^2 D_t^{-1}), \quad (\text{C-95})$$

The derivations for U are similar. Under either (C-81) or (C-82) we can derive

$$p(U|Y, \phi, D, \mathbf{v}, \Lambda, L) \propto e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}'_t U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{C-96})$$

$$= e^{-\frac{1}{2} \text{vec}(U)' \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}'_t \otimes L' \Lambda D_t^{-1} \Lambda L) \text{vec}(U)}, \quad (\text{C-97})$$

$$= e^{-\frac{1}{2} \text{vec}(U)' \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}'_t \otimes L' \Lambda^2 D_t^{-1} L) \text{vec}(U)}, \quad (\text{C-98})$$

$$p(\mathbf{u}|Y, \phi, D, \mathbf{v}, \Lambda, U) \propto e^{-\frac{1}{2} (\mathbf{s} + S_U \mathbf{u})' W_U (\mathbf{s} + S_U \mathbf{u})}, \quad (\text{C-99})$$

$$\propto e^{-\frac{1}{2} (\mathbf{u}' S'_U W_U S_U \mathbf{u} + 2 \mathbf{u}' S'_U W_U \mathbf{s})}, \quad (\text{C-100})$$

hence

$$\mathbf{u}|Y, \boldsymbol{\phi}, D, \mathbf{v}, \Lambda, L \sim N(\boldsymbol{\mu}_U^*, V_U^*) \quad (\text{C-101})$$

$$V_U^* = (S_U' W_U S_U)^{-1}, \quad (\text{C-102})$$

$$\boldsymbol{\mu}_U^* = -V_U^* S_U' W_U \mathbf{s}, \quad (\text{C-103})$$

$$W_U = \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}_t' \otimes L' \Lambda^2 D_t^{-1} L). \quad (\text{C-104})$$

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

$$p(\Lambda|Y, \boldsymbol{\phi}, D, \mathbf{v}, L, U) \propto \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda LU) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' U' L' \Lambda D_t^{-1} \Lambda L U \mathbf{z}_t}, \quad (\text{C-105})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda LU) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{z}_t' U' L' D_t^{-0.5} \Lambda^2 D_t^{-0.5} L U \mathbf{z}_t}, \quad (\text{C-106})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda LU) \cdot e^{-\frac{1}{2} \sum_{t=1}^T \mathbf{c}_t' \Lambda^2 \mathbf{c}_t}, \quad (\text{C-107})$$

$$= \left(\prod_{i=1}^k |\lambda_i|^{T+k-1} \right) \cdot p_A(\Lambda LU) \cdot e^{-\frac{1}{2} \sum_{i=1}^k \sum_{t=1}^T c_{it}^2 \lambda_i^2}, \quad (\text{C-108})$$

$$= \prod_{i=1}^k |\lambda_i|^{T+k+\alpha-1} \cdot e^{-\frac{1}{2} \lambda_i^2 \sum_{t=1}^T c_{it}^2}, \quad (\text{C-109})$$

where $\alpha = 0$ if we adopt (C-81), or $\alpha = -2k$ if we use (C-82).

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

$$p(x_1, x_2, \dots, x_k|Y, \boldsymbol{\phi}, D, \mathbf{v}, L, U) \propto \prod_{i=1}^k x_i^{-\frac{1}{2}} |x_i^{\frac{1}{2}}|^{T+k+\alpha-1} \cdot e^{-\frac{1}{2} x_i \sum_{t=1}^T c_{it}^2}, \quad (\text{C-110})$$

$$= \prod_{i=1}^k x_i^{\frac{T+k+\alpha}{2}-1} \cdot e^{-\frac{1}{2} x_i \sum_{t=1}^T c_{it}^2}, \quad (\text{C-111})$$

hence

$$x_i|Y, \phi, D, \mathbf{v}, L, U \sim \Gamma(h_{\lambda,i}^*, r_{\lambda,i}^*), \quad (\text{C-112})$$

$$r_{\lambda,i}^* = \frac{\sum_{t=1}^T c_{it}^2}{2}, \quad (\text{C-113})$$

$$\mathbf{c}_t = D_t^{-0.5} L U \mathbf{z}_t, \quad (\text{C-114})$$

where it holds

$$h_{\lambda,i}^* = \frac{T + k}{2}, \quad (\text{C-115})$$

if prior (C-81) is used, and

$$h_{\lambda,i}^* = \frac{T - k}{2}, \quad (\text{C-116})$$

if prior (C-82) is used. Note that we are using the following shape-rate parametrization of the Gamma distribution

$$p(x|h, r) = \frac{r^h}{\Gamma(h)} \cdot x^{h-1} e^{-r \cdot x}. \quad (\text{C-117})$$

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 an 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

$$\min_{P_s} \text{tr}\{(BP_s - \hat{B})' \hat{A}' \hat{A} (BP_s - \hat{B})\}, \quad (\text{D-118})$$

where $\hat{\theta}$ denotes the Maximum Likelihood (ML) estimator of θ .⁴ By multiplying matrices in (D-118) and using the fact that $\hat{B} = \hat{A}^{-1}$, one can show that (D-118) is equivalent to any of the following maximization problems:

$$\max_{P_s} \text{tr}\{P_s \hat{A} B\} = \max_{P_s} \text{tr}\{\hat{A} B P_s\} = \max_{P_s} \text{tr}\{\hat{A} A^{-1} P_s\} = \max_{P_s} \text{tr}\{P_s' A'^{-1} \hat{A}'\}. \quad (\text{D-119})$$

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 (D-118) 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 solution to the last formula in (D-119) 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 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 $-1 \cdot e_i' A^{-1} \hat{a}_i = e_i' I_k^* A^{-1} \hat{a}_i = e_i' (A I_k^*)^{-1} \hat{a}_i$, 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 (D-119). To this end, let us denote $G = \hat{A} A^{-1}$

⁴Note 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 (D-118) remains the same if instead of the original data y_t we use $H y_t$, where H is any nonsingular matrix - see Proposition 5 in WZ.

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 see 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:

$$\begin{aligned} p_1 g_{\pi(1),1} + p_2 g_{\pi(2),2} + \dots + p_k g_{\pi(k),k} &= |p_1 g_{\pi(1),1}| + |p_2 g_{\pi(2),2}| + \dots + |p_k g_{\pi(k),k}|, \\ &= |p_1| |g_{\pi(1),1}| + |p_2| |g_{\pi(2),2}| + \dots + |p_k| |g_{\pi(k),k}|, \\ &= |g_{\pi(1),1}| + |g_{\pi(2),2}| + \dots + |g_{\pi(k),k}|. \end{aligned} \quad (\text{D-120})$$

This suggests the modified problem:

$$\max_P \text{tr}\{P \cdot |G|\} = \max_P \text{tr}\{P \cdot |\hat{A}A^{-1}|\}, \quad (\text{D-121})$$

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 (D-119) is attained by the modified (i.e. constrained) optimization problem (D-121). When working with (D-121), 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).

All in all, instead of evaluating the objective function (D-118) (or D-119) for every P_s , which is computationally infeasible even for medium-scale models, we only need to solve

$$\max_P \text{tr}\{P \cdot |G|\} = \min_P \text{tr}\{P \cdot -|G|\}. \quad (\text{D-122})$$

However, the formulation in (D-122) shows that we encounter the 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 (D-122), and repeat. The following summarizes, on average over 100,000 repetitions, how long it took to solve (D-122): 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 (D-122), 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 (D-118). 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:

Given a target matrix \hat{A} apply the following steps at each iteration of the sampler, after drawing (A, B) :

1. solve $\min_P tr\{P \cdot -|\hat{A}B|\}$ 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 (D-118);
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

$$\min_{P_s} (\text{vec}(A'P_s) - \text{vec}(\hat{A}'))' \hat{V}^{-1} (\text{vec}(A'P_s) - \text{vec}(\hat{A}')), \quad (\text{D-123})$$

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 (D-118) as

$$\min_{P_s} [(\text{vec}(BP_s) - \text{vec}(\hat{B}))'(I_k \otimes \hat{\Omega})^{-1}(\text{vec}(BP_s) - \text{vec}(\hat{B}))], \quad (\text{D-124})$$

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 (D-123).

E A Comparison with the Keweloh and Wang (2026) Normalization

Let \mathcal{P}_s denote the set of signed permutation matrices. Fix an “anchor” (target) matrix \bar{B} (e.g., a preliminary maximum likelihood estimate). In this comparison, we will use the same target matrix for both approaches.

A normalization selects, for each draw B , a representative BP_s from the equivalence class $\{BP_s : P_s \in \mathcal{P}_s\}$. With the *same anchor* \bar{B} as in our normalization, Keweloh and Wang (2026) define $C = \bar{B}^{-1}B$ and pick the signed permutation that corresponds to the specific choice of C with the following properties:

$$|C_{jj}| > |C_{j\ell}| \quad \forall \ell > j, \quad C_{jj} > 0 \quad \text{for } j = 1, \dots, k. \quad (\text{E-125})$$

Operationally, (E-125) behaves like a *sequential/lexicographic* rule: row 1 determines which column is labeled as shock 1; row 2 chooses among remaining columns for shock 2; etc.

The dominance rule (E-125) does *not* solve the global minimization (D-118). Therefore, it can select a different signed permutation than the LP objective, even when both procedures use the *same anchor* \bar{B} .

Example (3 shocks). Consider the following example for $C = \bar{B}^{-1}B$:

$$|C| = \begin{pmatrix} 0.86 & 0.97 & 0.10 \\ 0.10 & 0.88 & 0.65 \\ 0.24 & 0.87 & 0.79 \end{pmatrix}. \quad (\text{E-126})$$

The [Keweloh and Wang \(2026\)](#) normalization assigns shock 1 to column 2 (largest entry in row 1), then shock 2 to column 3 (largest among remaining in row 2), leaving column 1 for shock 3. This corresponds to the permutation (2, 3, 1) and yields the diagonal “match” sum

$$0.97 + 0.65 + 0.24 = 1.86.$$

Our approach instead uses the permutation (1, 2, 3), with diagonal sum

$$0.86 + 0.88 + 0.79 = 2.53,$$

yielding a strictly smaller objective function (D-118).

Because P_s^* *minimizes* Equation (D-118), any alternative normalization that selects a different P_s produces a representative that is *weakly farther* from the anchor in the likelihood-preserving metric (strictly farther if the minimizer is unique), increasing the scope for label instability/mode switching in a posterior sampler.

F 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 $\hat{B} = \hat{A}^{-1}$ using e.g. the maximum likelihood estimator.⁵ Then, at each iteration of the sampler:
 1. draw ϕ from the Normal conditional posterior from equation (C-56) - in large dimensional settings, one can use a dummy observation approach as in Bańbura et al. (2010);
 2. draw L from the Normal conditional posterior from equation (C-92);
 3. draw U from the Normal conditional posterior from equation (C-101) (possibly under the additional identifying restrictions on A);
 4. draw x_i 's from the Gamma conditional posteriors from equation (C-112);
 5. compute Λ by setting each (i, i) entry either as $\lambda_i = -\sqrt{x_i}$ or $\lambda_i = \sqrt{x_i}$ (with equal probability);
 6. compute \bar{A} associated with (Λ, L, U) using equation (C-77) and set $\bar{B} = \bar{A}^{-1}$;
 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 (C-66);
 9. draw \mathbf{v} from the discretized conditional posterior from equation (C-74);

⁵To 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.

10. repeat from step 1.

To initialize the sampler one can set $\phi = \text{vec}(\Pi)$ to the OLS estimate, (\mathbf{l}, \mathbf{u}) to zero, and λ_i 's equal to the standard deviation of the residual in an univariate autoregressive process estimated on a training sample, in the spirit of the Minnesota prior. \mathbf{v} can, instead, be set at the ML estimates. If no MLE is to be computed before the sampler, one can run the burn-in part of the Gibbs sampler without applying any normalization, and then set \hat{B} equal to the value associated with the highest evaluation of the posterior distribution in the burn-in part of the sampler, see footnote 17 in the main text. Then, an alternative value for the initialization of \mathbf{v} is to be selected.

We end this section with two final remarks. First, note that placing the normalization step 7 before drawing (D, \mathbf{v}) ensures that the ordering of the structural shocks is consistent with the ordering of the degrees of freedom. Second, note that the stored values of B are associated with a matrix A that may not admit a decomposition via equation (C-77). This fact is without loss of generality, as the decomposition is only employed as an operational procedure to develop a Gibbs sampler.

G Additional material on the simulation exercise

The parameter values of the data generating process are

$$B = \begin{pmatrix} 0.60 & 0.40 \\ 0.70 & -0.70 \end{pmatrix}, \quad (\text{G-127})$$

$$\mathbf{v} = \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \quad (\text{G-128})$$

$$\Pi_1 = \begin{pmatrix} 1.0612 & -0.0759 \\ -0.2502 & 1.1404 \end{pmatrix}, \quad (\text{G-129})$$

$$\Pi_2 = \begin{pmatrix} -0.0660 & 0.0093 \\ -0.0253 & -0.0905 \end{pmatrix}, \quad (\text{G-130})$$

$$\Pi_3 = \begin{pmatrix} -0.0641 & 0.0109 \\ 0.0286 & -0.0655 \end{pmatrix}, \quad (\text{G-131})$$

$$\Pi_4 = \begin{pmatrix} -0.0530 & 0.0119 \\ 0.0639 & -0.0434 \end{pmatrix}, \quad (\text{G-132})$$

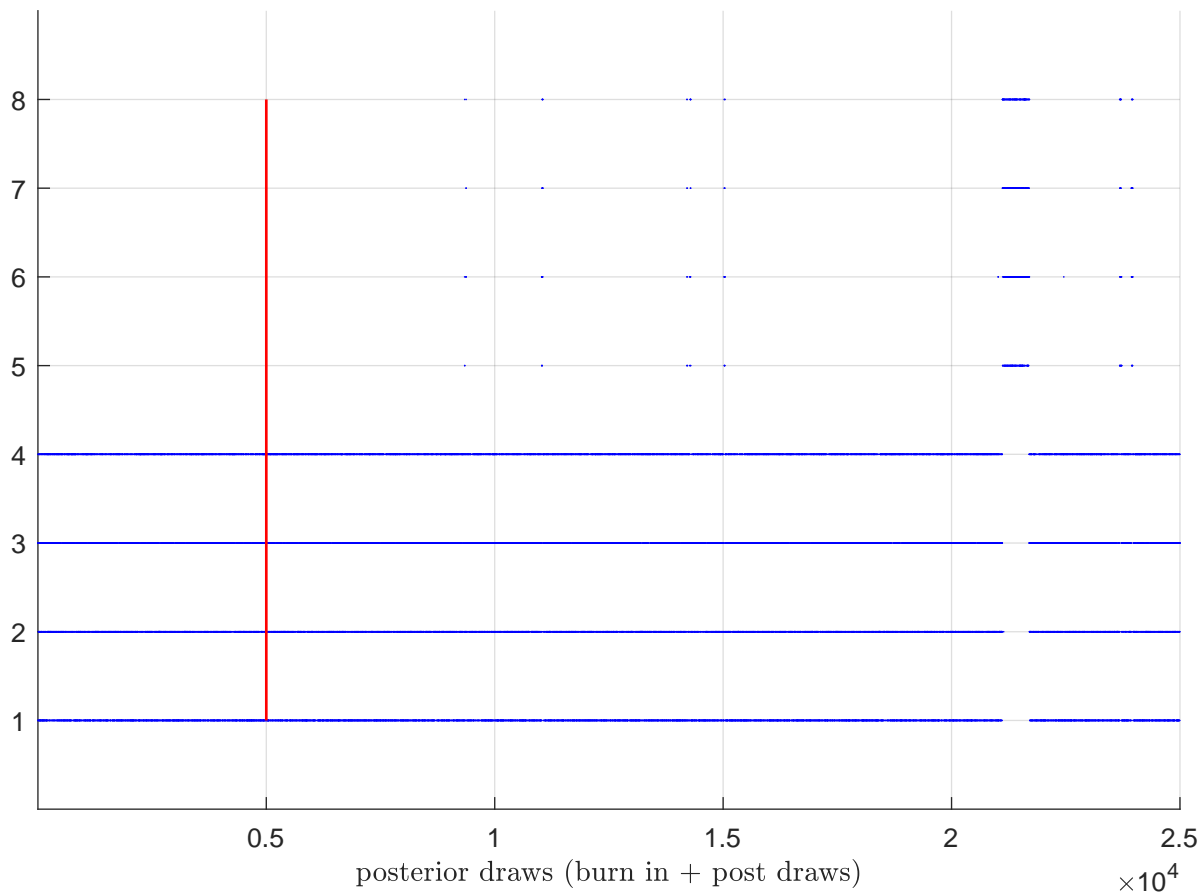
$$\Pi_5 = \begin{pmatrix} -0.0355 & 0.0113 \\ 0.0660 & -0.0304 \end{pmatrix}, \quad (\text{G-133})$$

$$\Pi_6 = \begin{pmatrix} -0.0165 & 0.0084 \\ 0.0425 & -0.0230 \end{pmatrix}. \quad (\text{G-134})$$

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

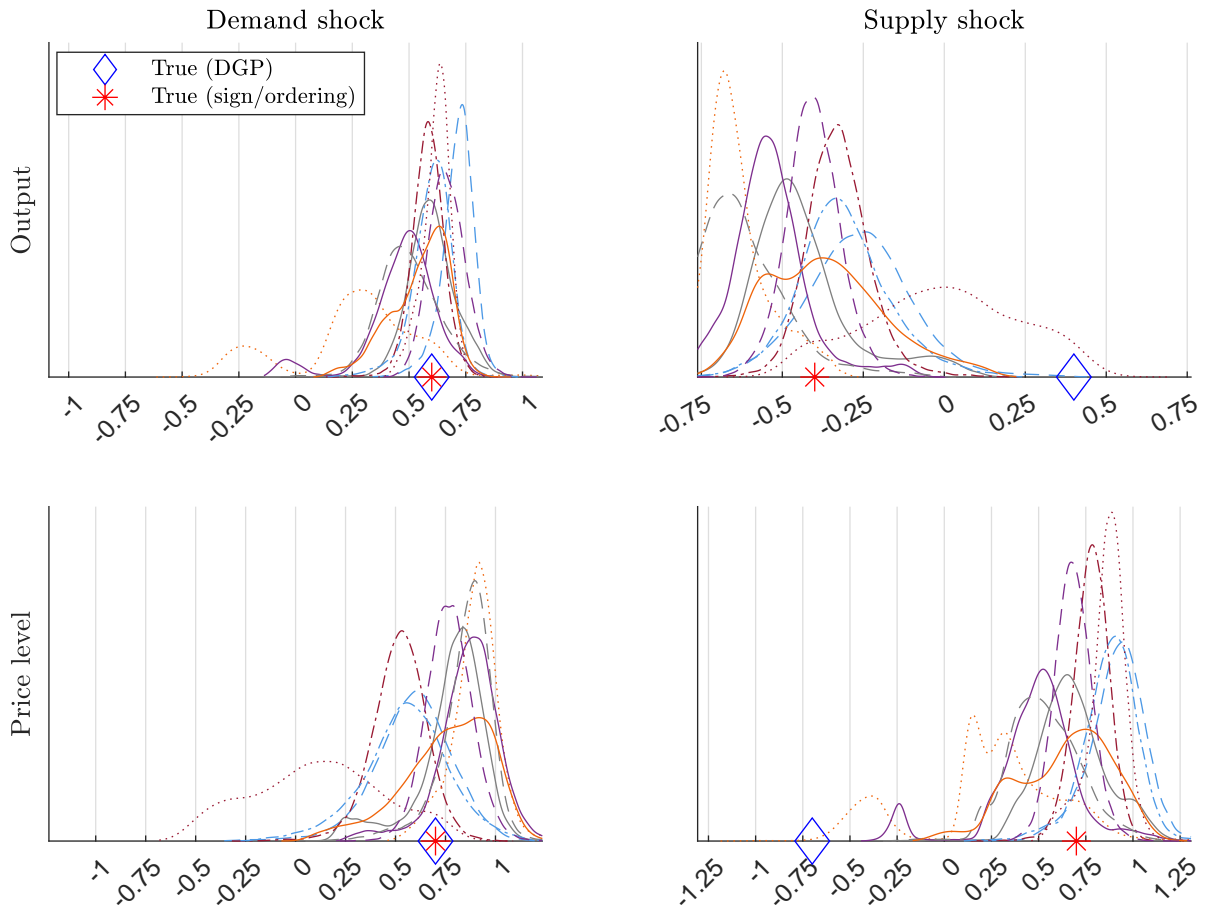
$$B_{target} = \begin{pmatrix} 0.7015 & -0.3279 \\ 0.7271 & 0.7236 \end{pmatrix}. \quad (\text{G-135})$$

Figure G-2: Generalized LP normalization



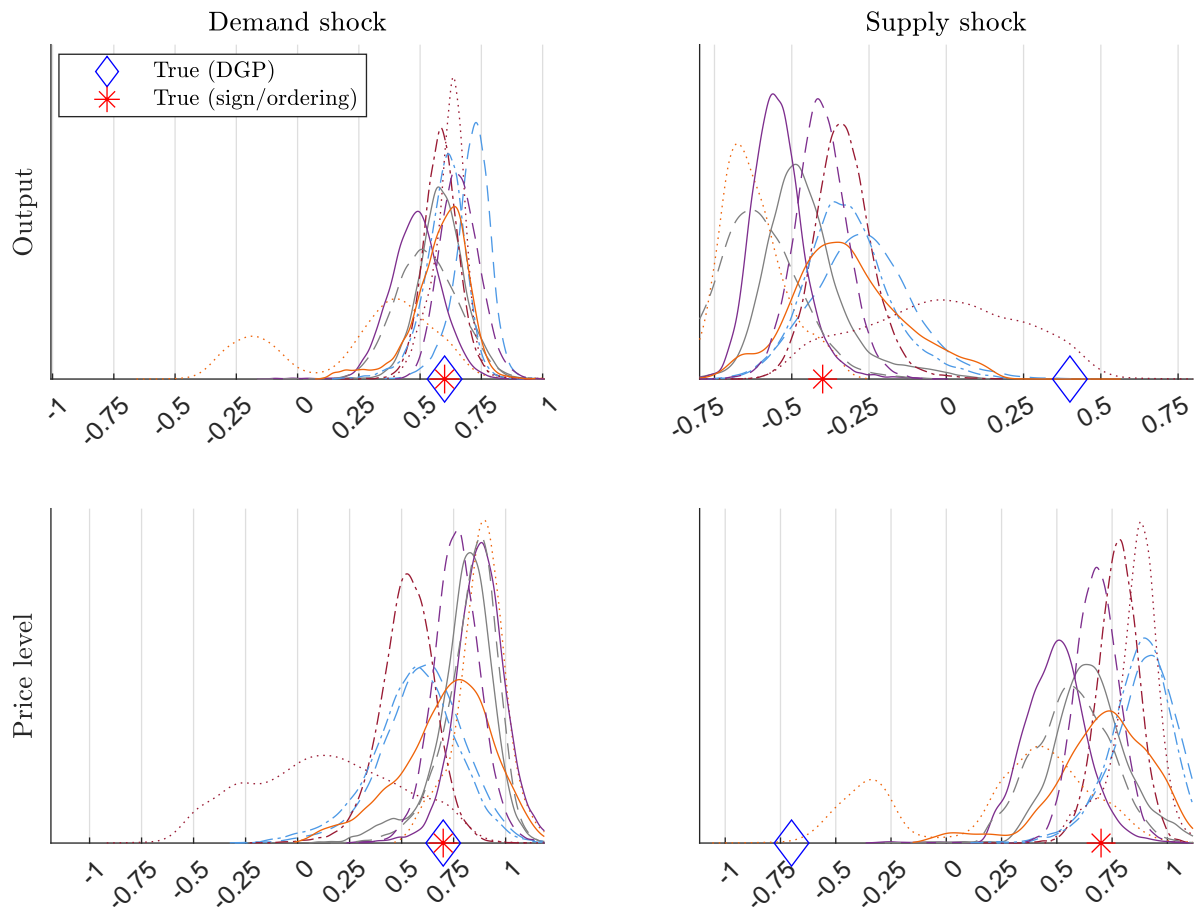
Note: The vertical red line shows when the burn-in draws end. For each of the 25,000 posterior draws, the figure indicates if the generalized LP normalization permutes the ordering of the columns of B (vertical values 5-8) or not (vertical values 1-4). It also indicates if the sign of the columns of 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 G-3: Comparing normalizations: our approach



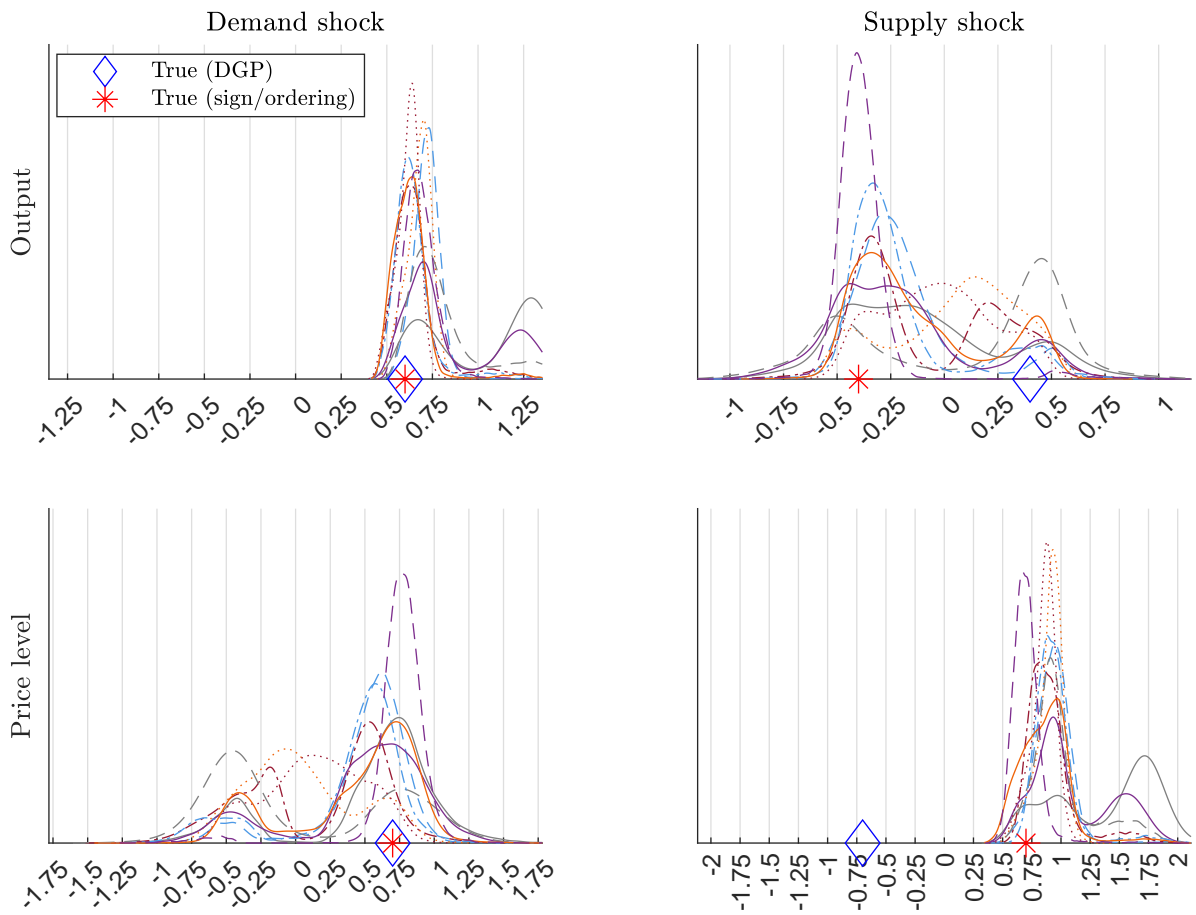
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

Figure G-4: Comparing normalizations: [Jarociński \(2024\)](#)



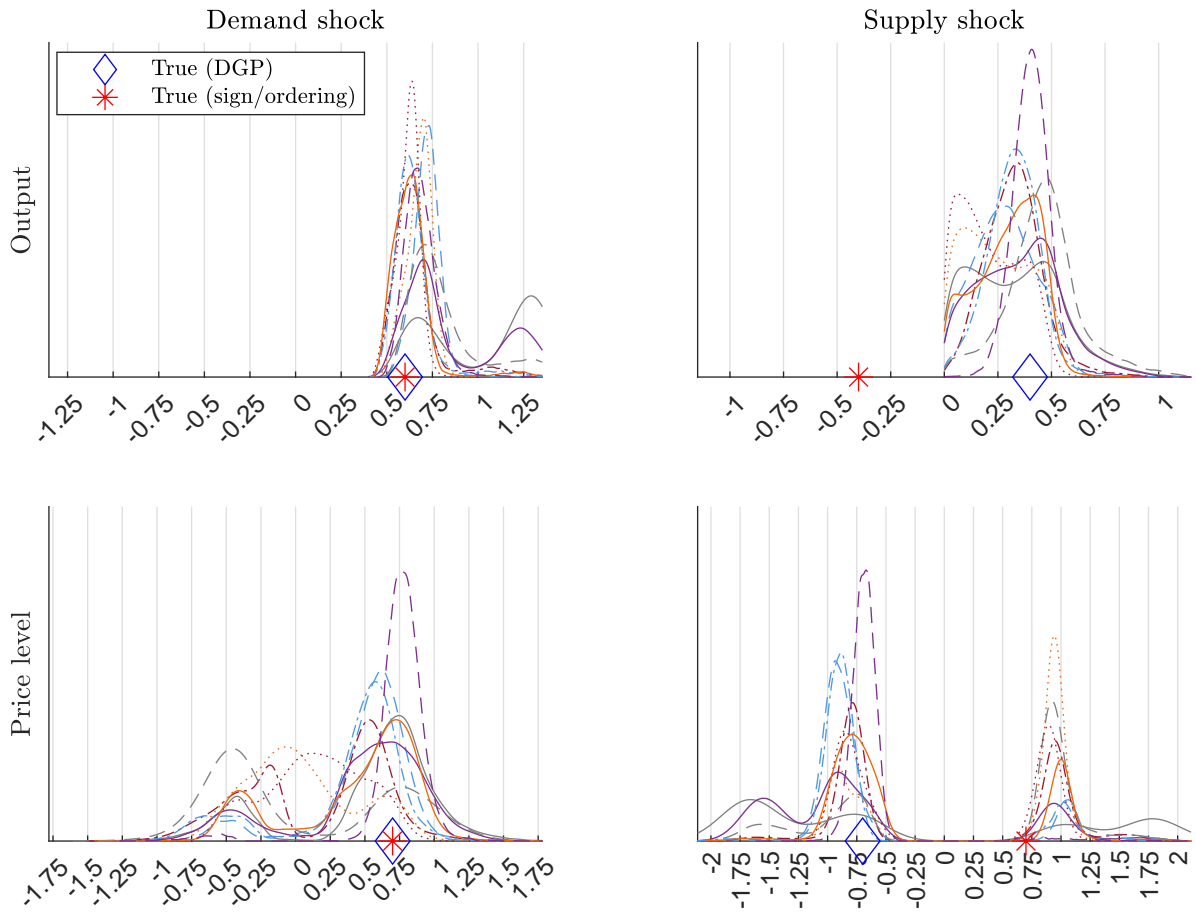
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

Figure G-5: Comparing normalizations: [Lanne et al. \(2017\)](#)



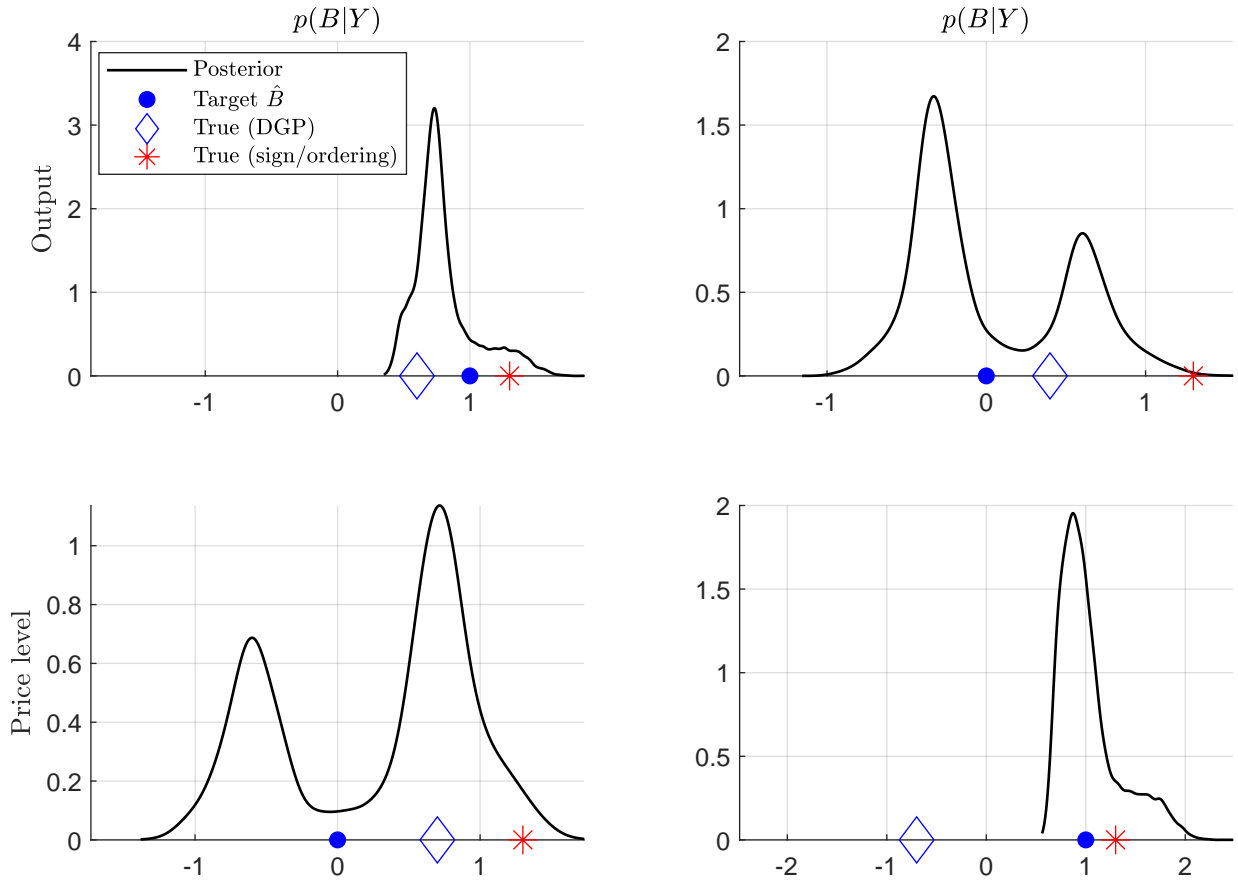
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

Figure G-6: Comparing normalizations: [Gouriéroux et al. \(2020\)](#)



Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

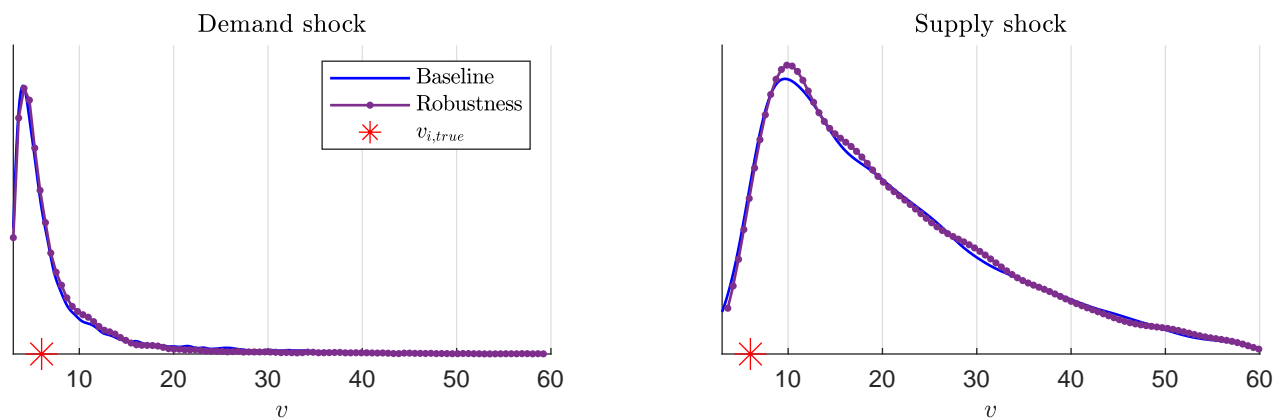
Figure G-7: Impact effect of the shocks (B) when the normalization targets an arbitrary matrix: $\hat{B} = I_2$



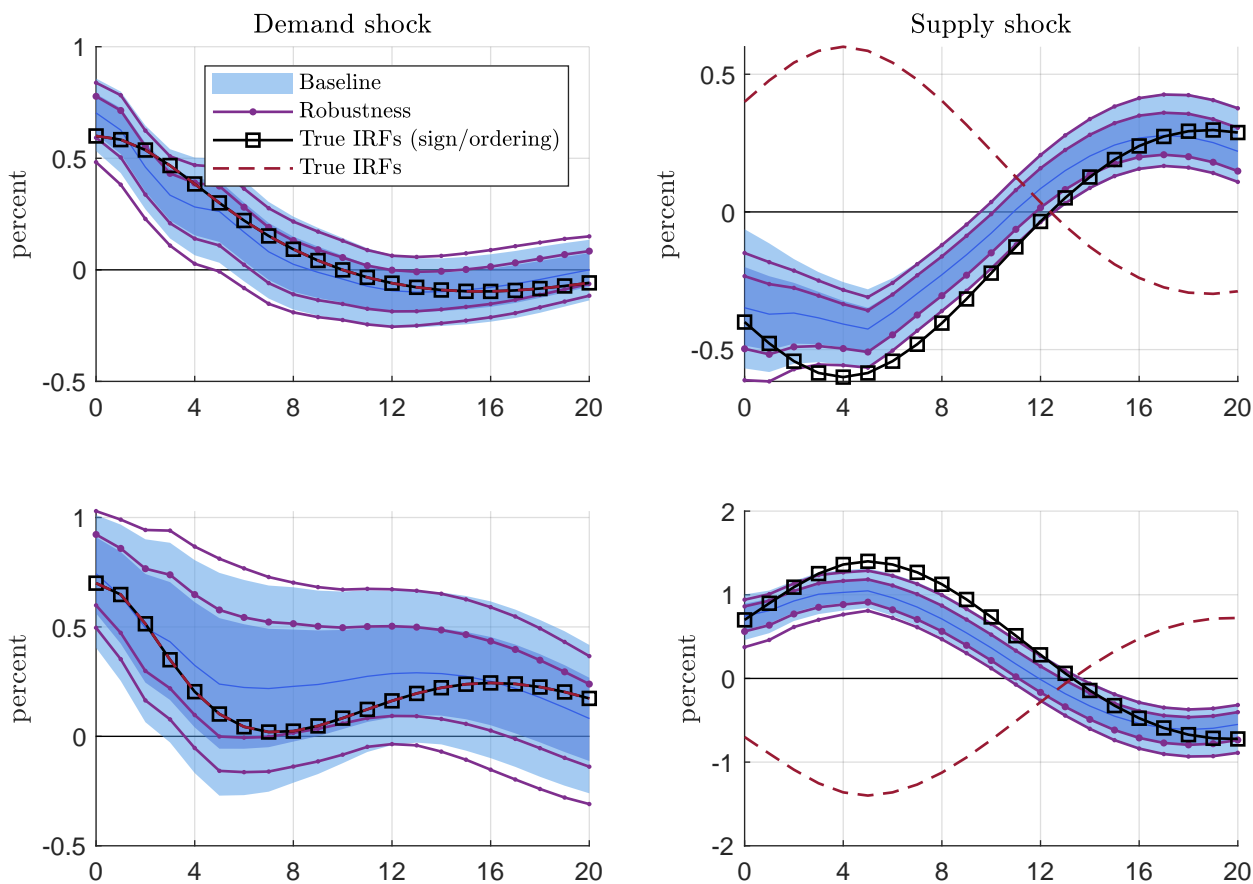
Note: The blue diamond indicates B_{true} . The blue dot indicates the target matrix \hat{B} used for the normalization. The red star indicates the sign/permutation of B_{true} that is the closest to \hat{B} . The continuous line shows the marginal posterior of the entries of B from 20,000 posterior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

Figure G-8: Robustness when computing the target matrix in the sampler

A) Degrees of freedom



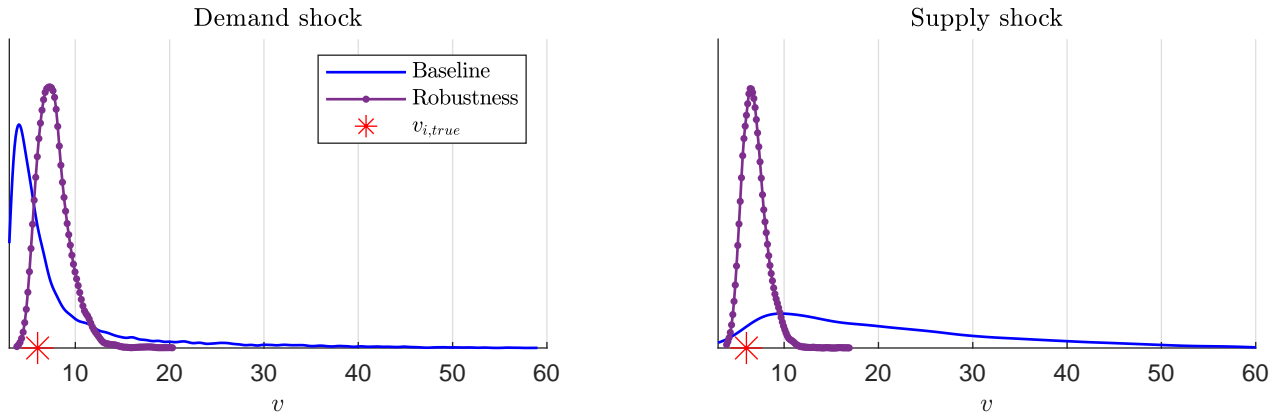
B) Impulse responses



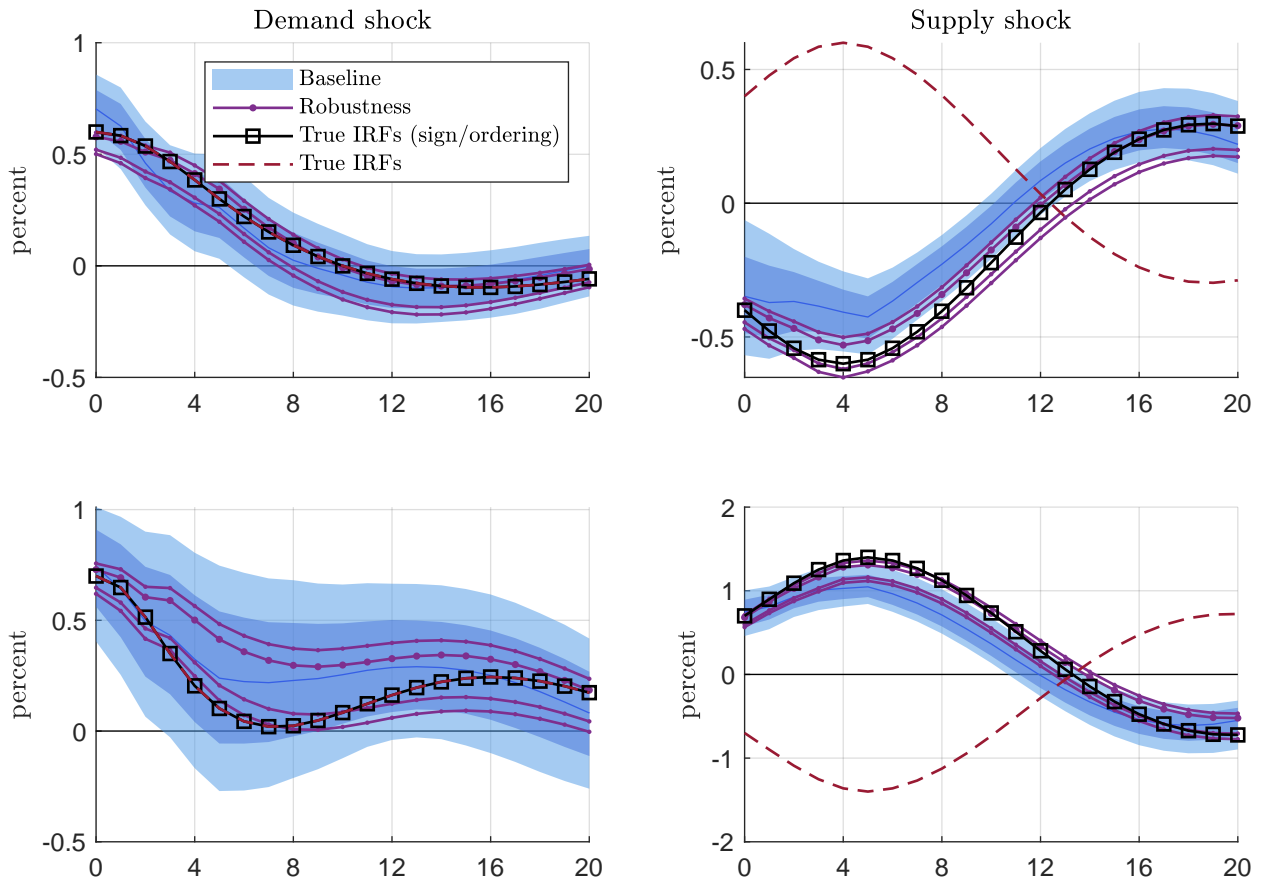
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure G-9: Robustness for $T = 1,000$

A) Degrees of freedom



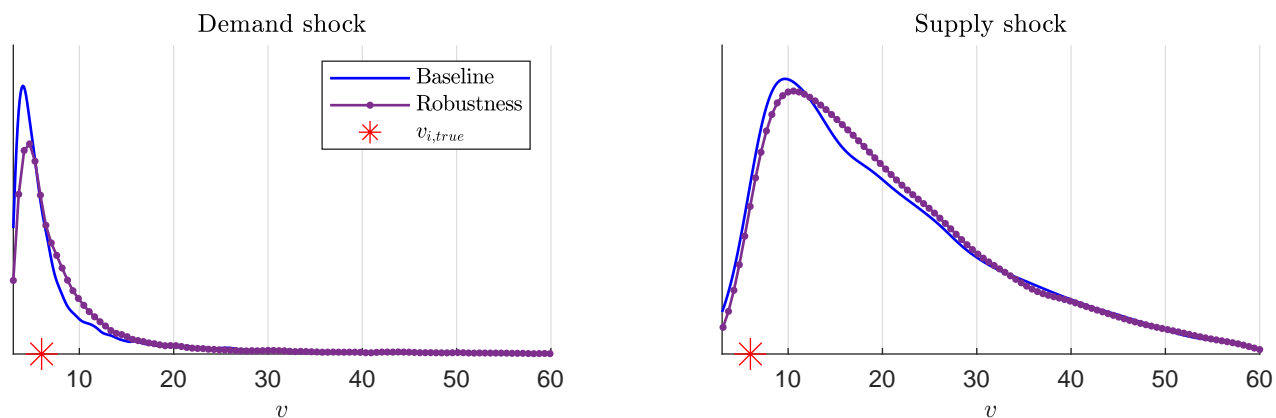
B) Impulse responses



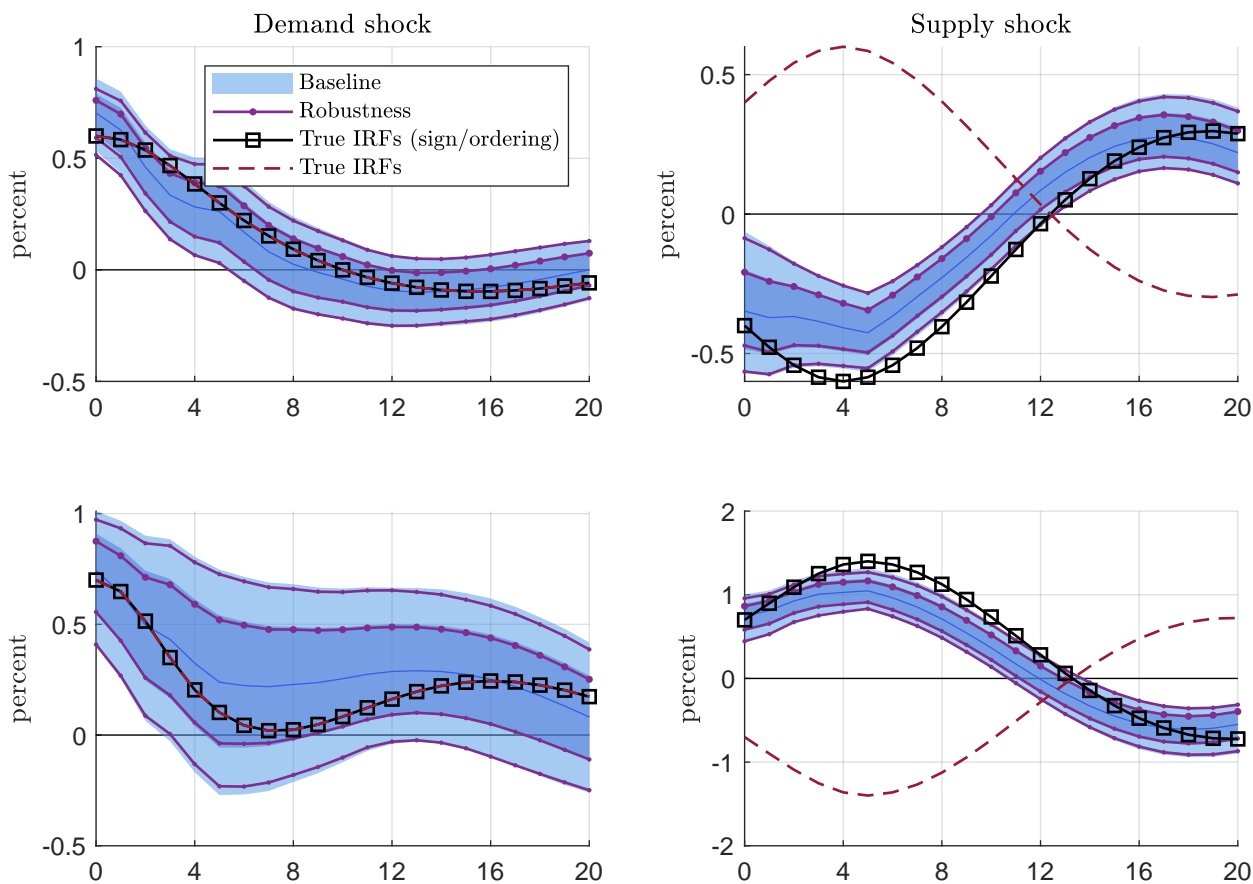
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure G-10: Robustness for flat prior on A

A) Degrees of freedom



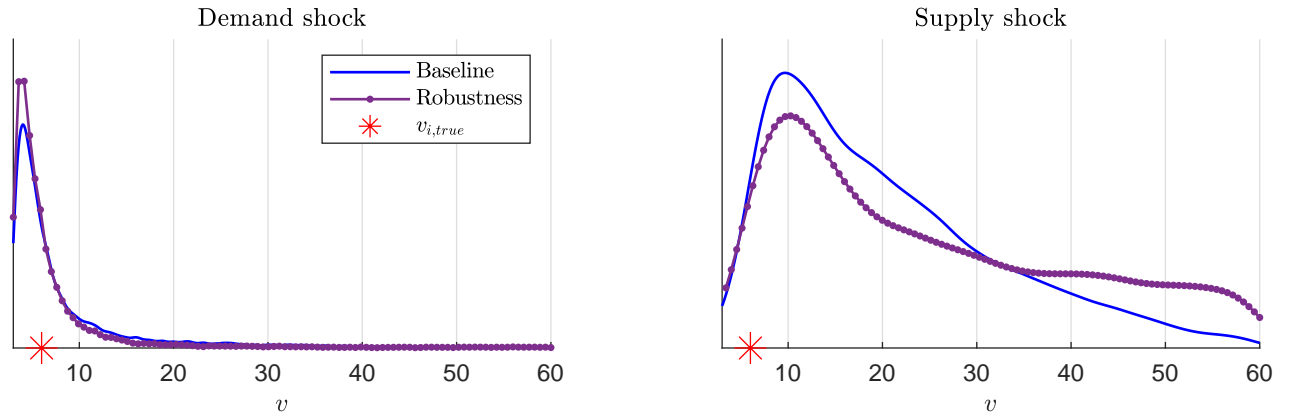
B) Impulse responses



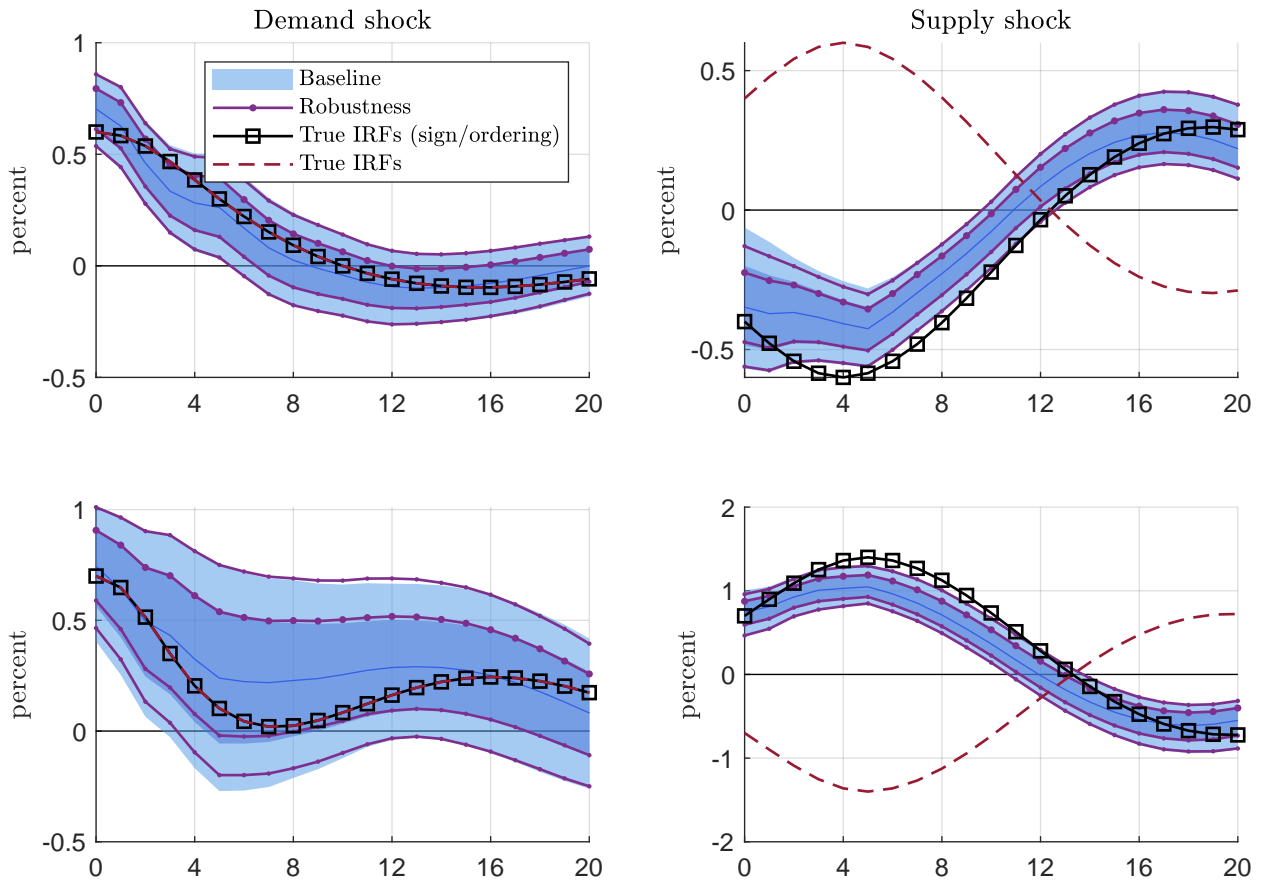
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure G-11: Robustness for flat prior on v

A) Degrees of freedom



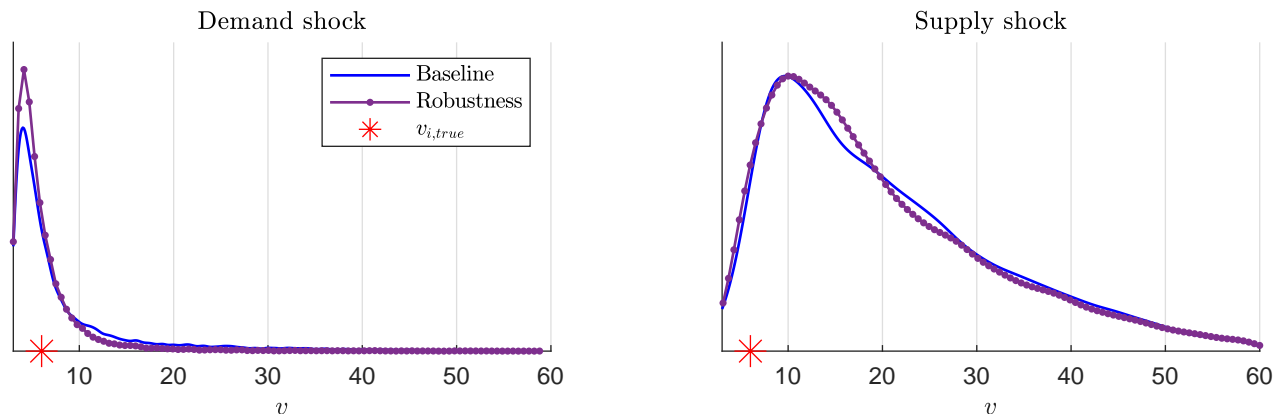
B) Impulse responses



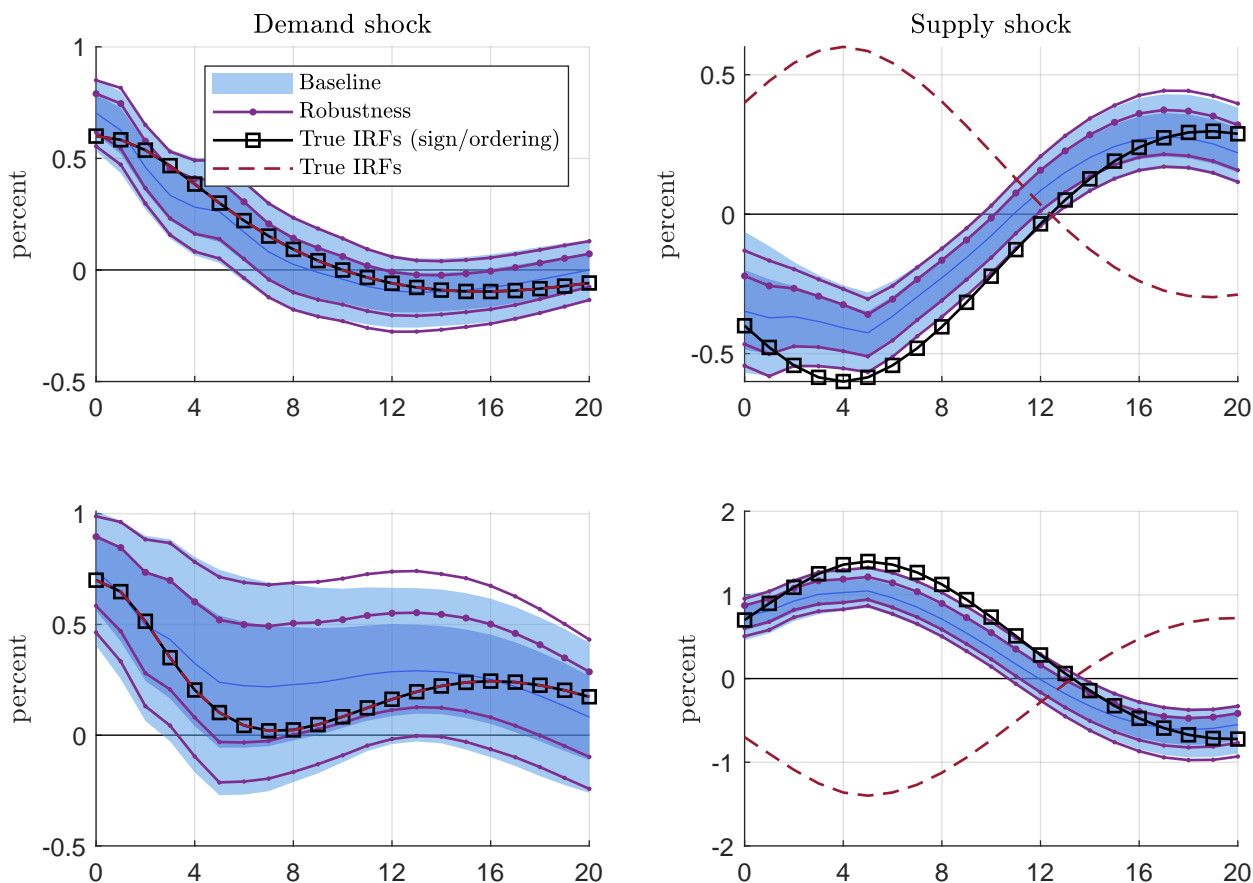
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure G-12: Robustness for looser prior on ϕ ($\lambda_1 = 0.2$ rather than 0.1 using the notation in Canova, 2007)

A) Degrees of freedom



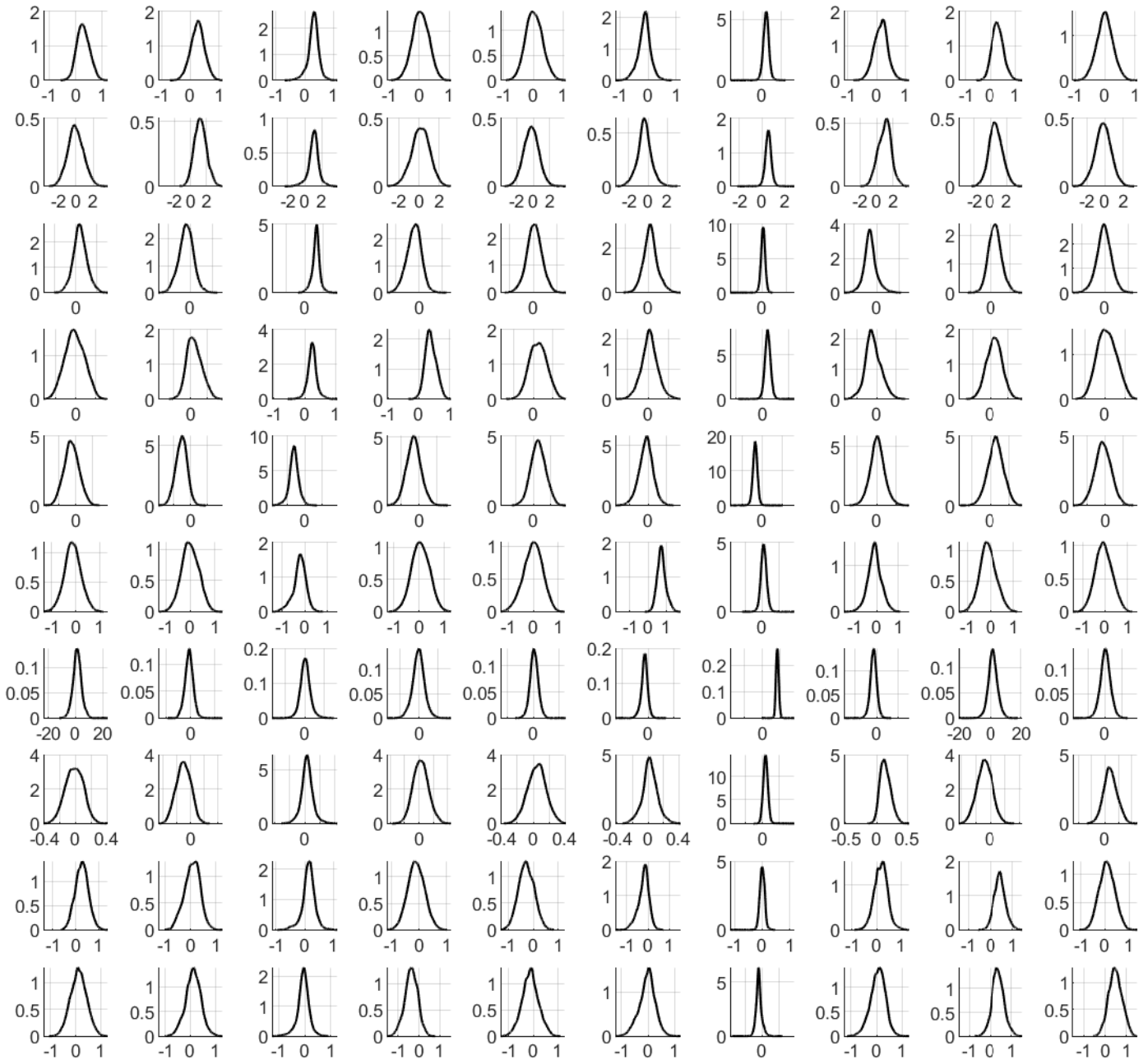
B) Impulse responses



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 solid lines show the pointwise credible sets in the alternative specification.

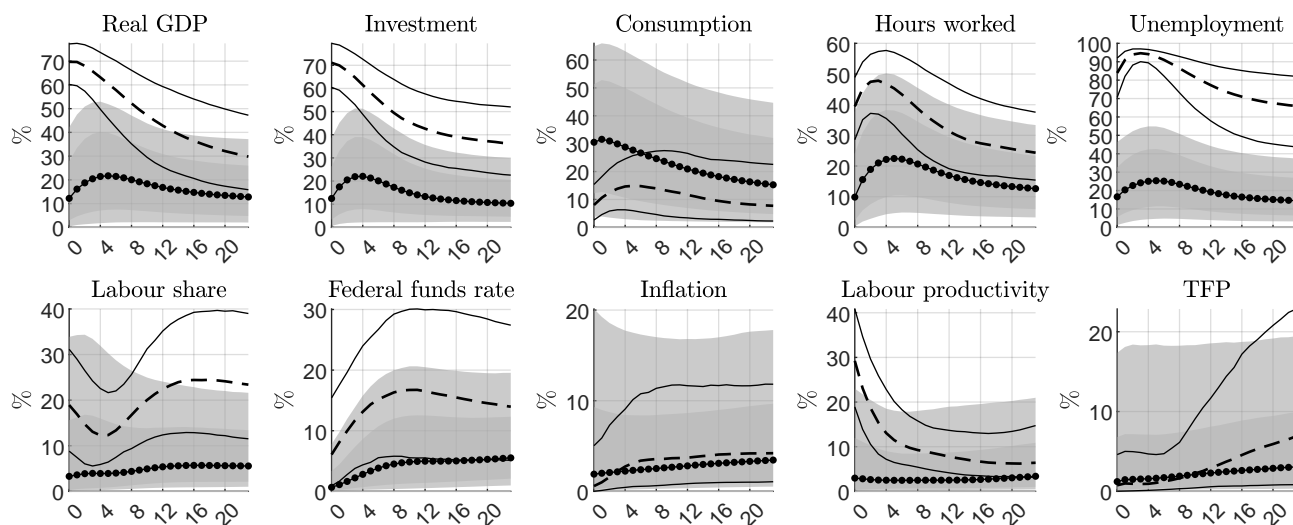
H Additional material on the application to the US GDP

Figure H-13: Marginal posterior distributions, $p(B|Y)$

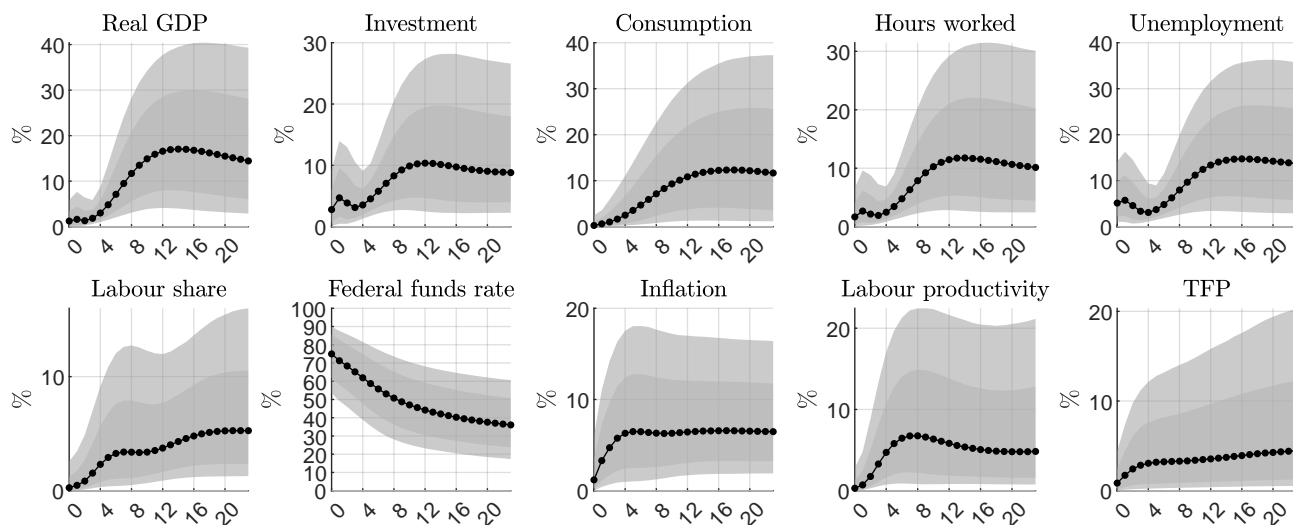


Note: the rows in the figure indicate the variables in the model, the columns indicate the shocks.

Figure H-14: Forecast error variance decomposition
Shock 3: demand shock



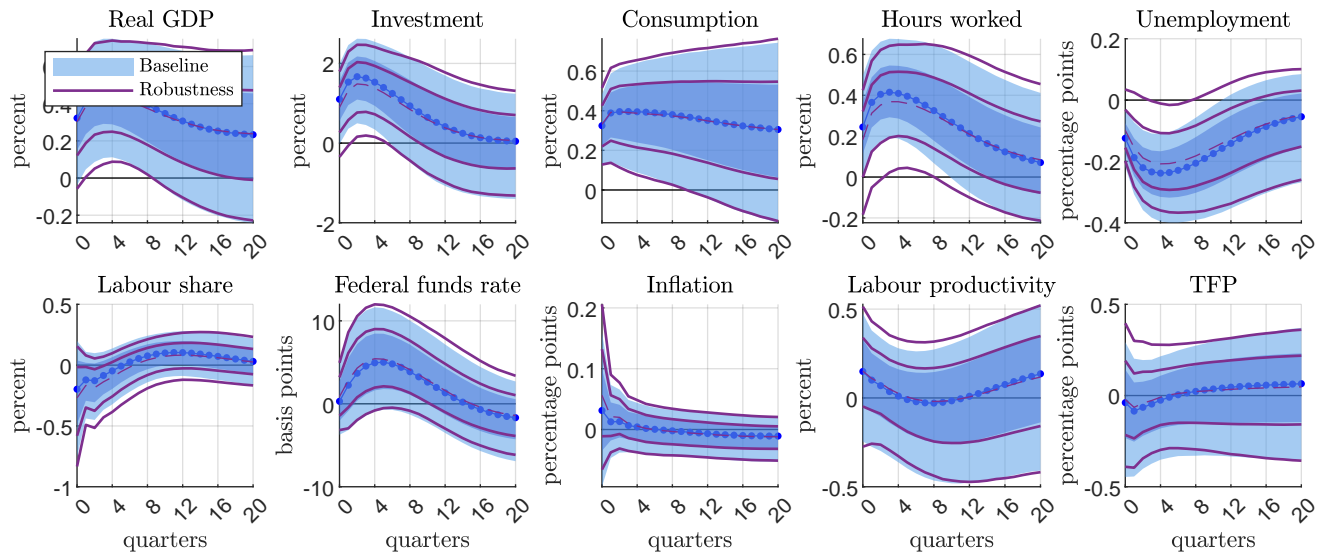
Shock 7: supply shock



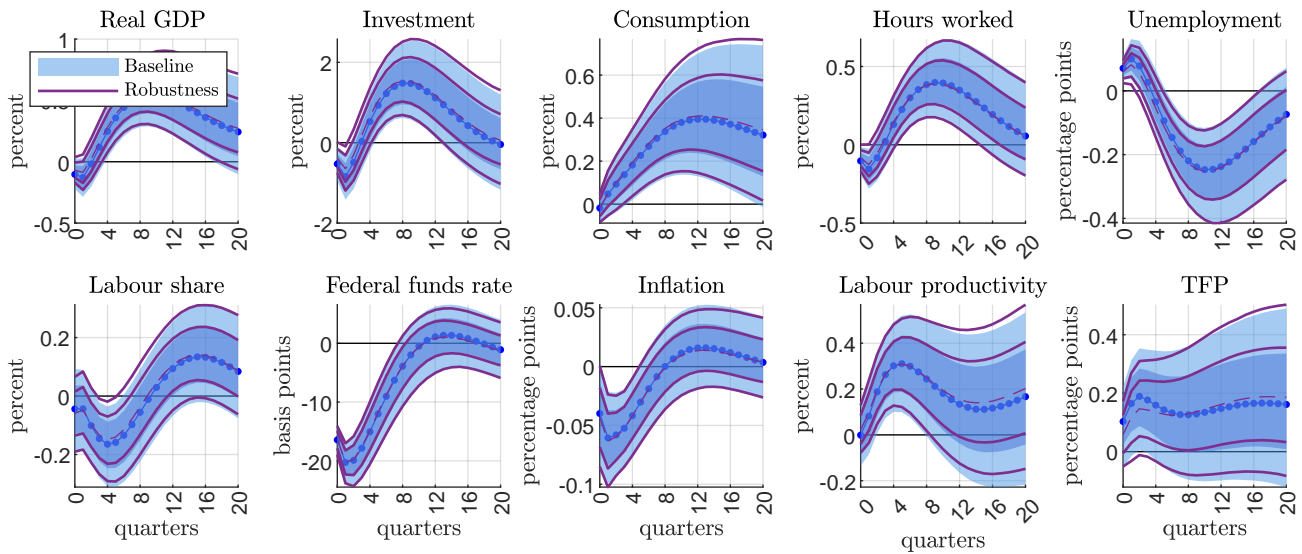
Note: The black dotted lines show the pointwise posterior median from our estimation, with corresponding 68% and 90% pointwise posterior credible sets shown with shaded areas. The black dashed and solid lines in the top figure show the pointwise median and 90% credible sets estimated by [Angeletos et al. \(2020\)](#) for the Main Business Cycle shock.

Figure H-15: Robustness computing target matrix \hat{B} in the sampler: impulse responses

A) Demand shock



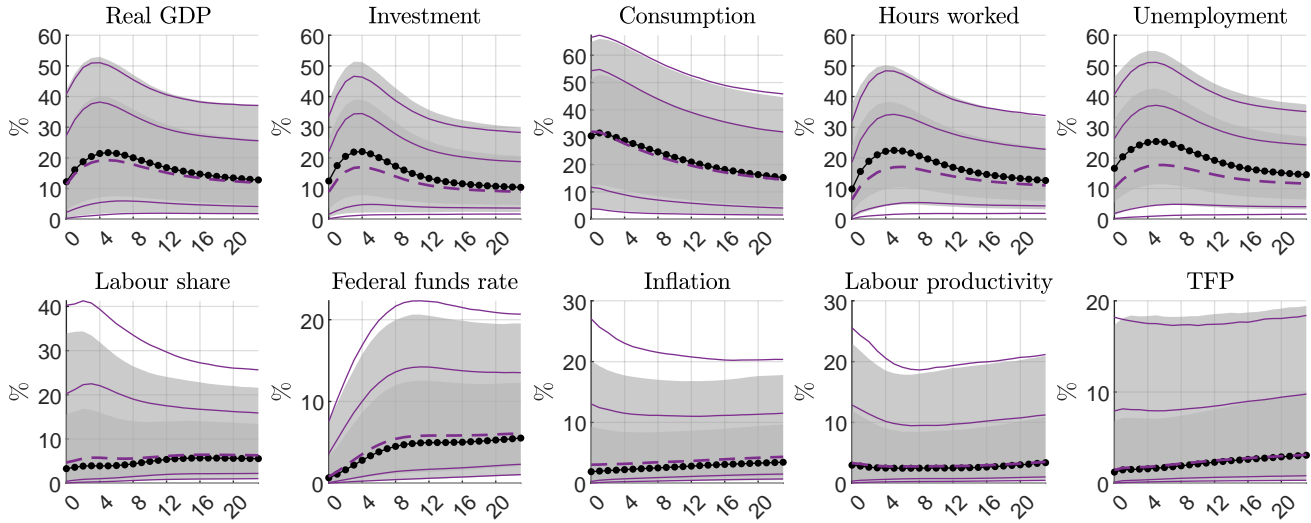
B) supply shock



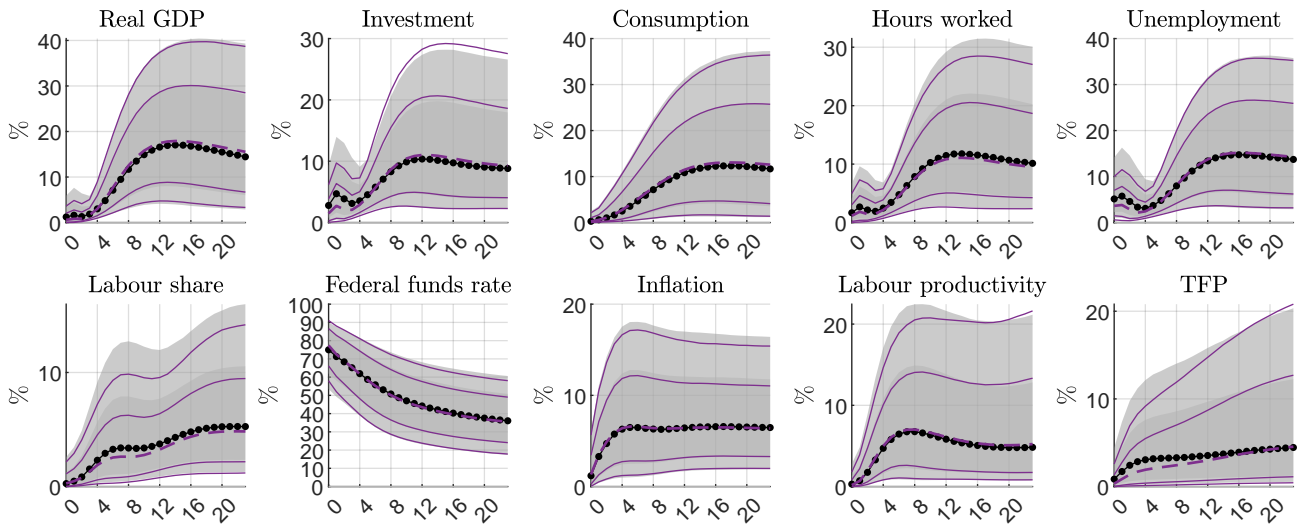
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure H-16: Robustness computing target matrix \hat{B} in the sampler: forecast error variance decomposition

A) Demand shock



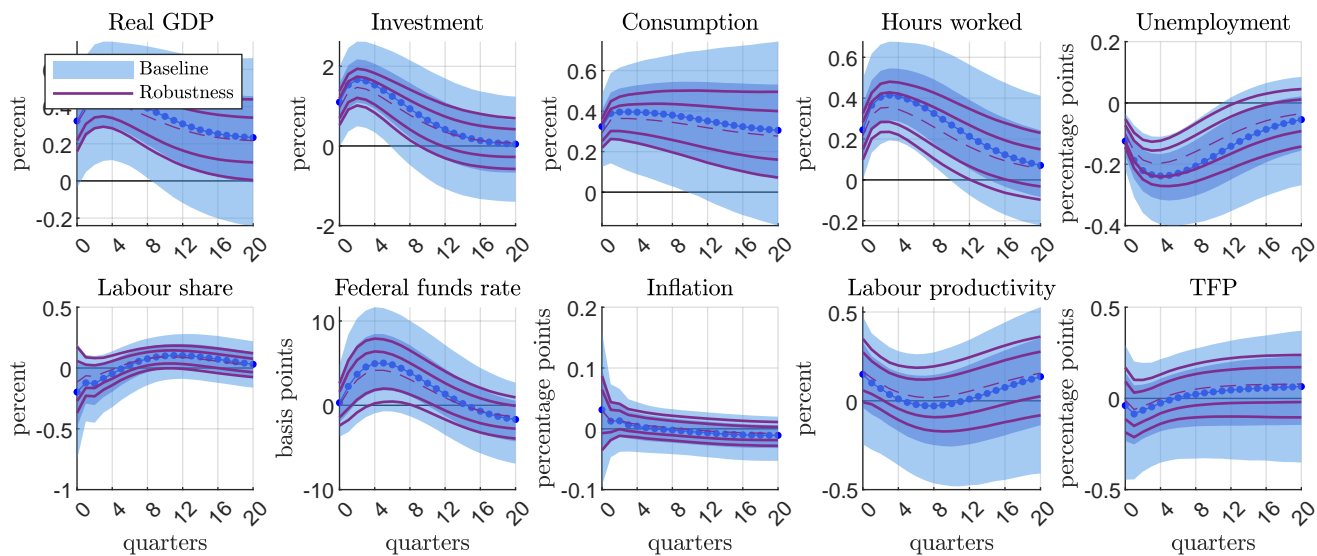
B) supply shock



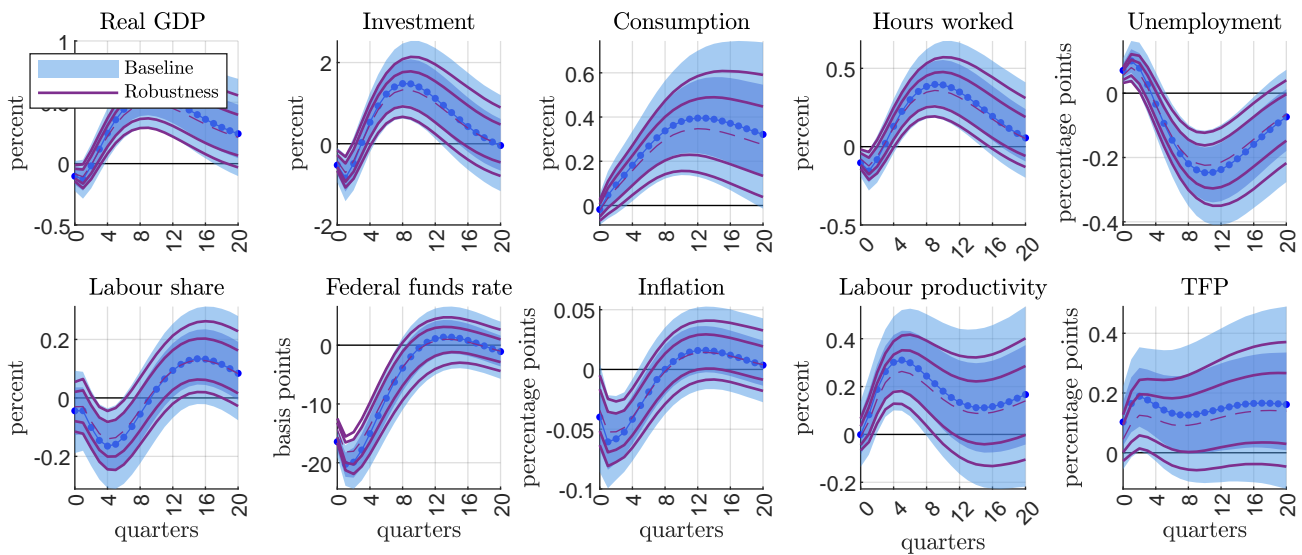
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure H-17: Robustness for flat prior on A: impulse responses

A) Demand shock



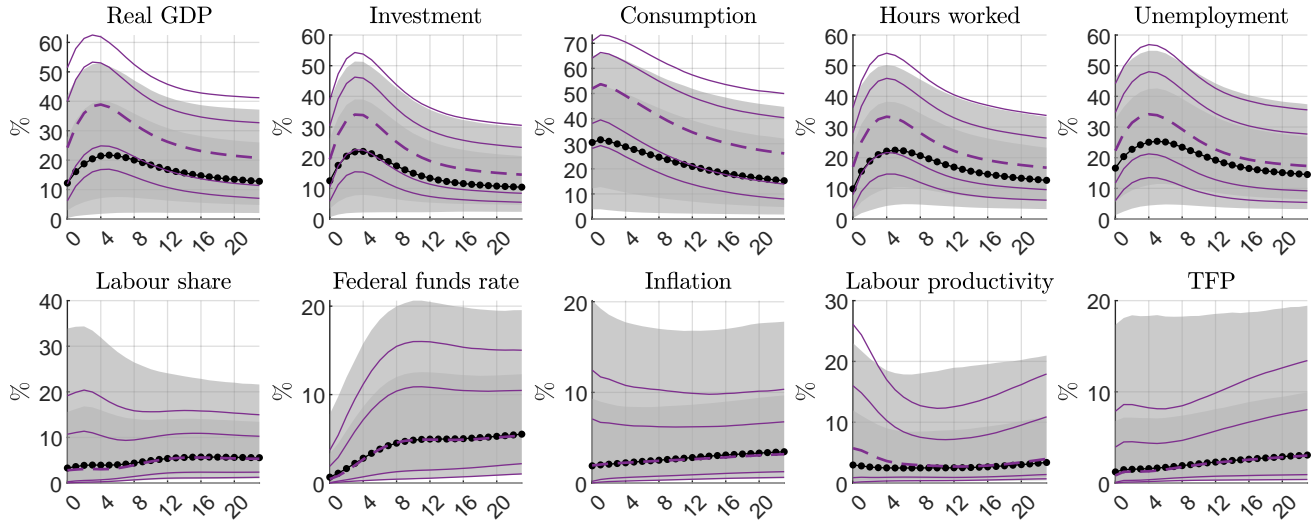
B) supply shock



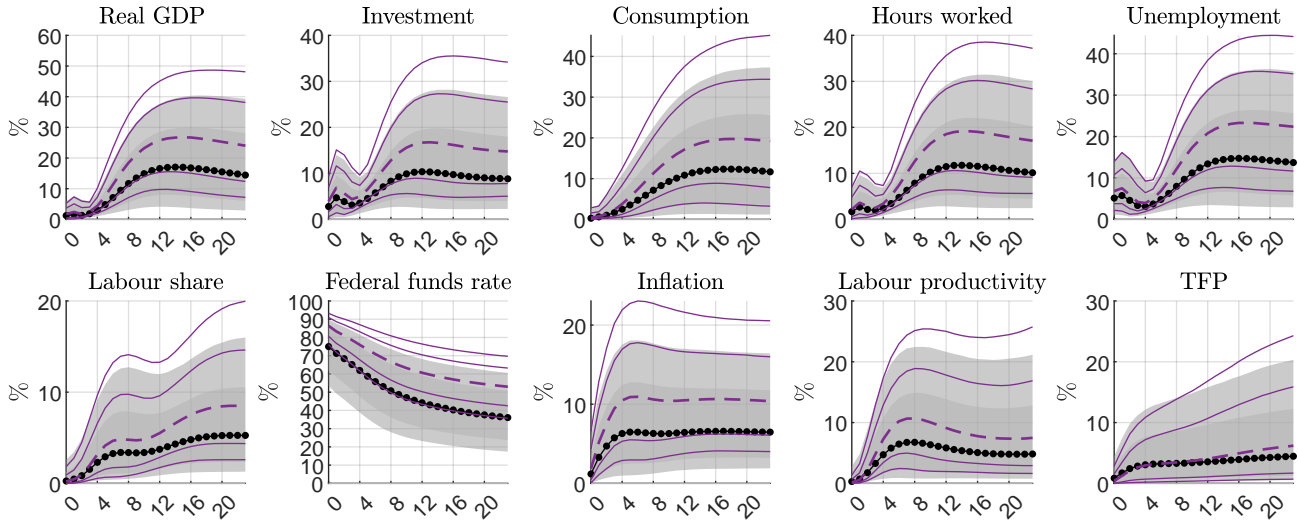
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure H-18: Robustness for flat prior on A : forecast error variance decomposition

A) Demand shock



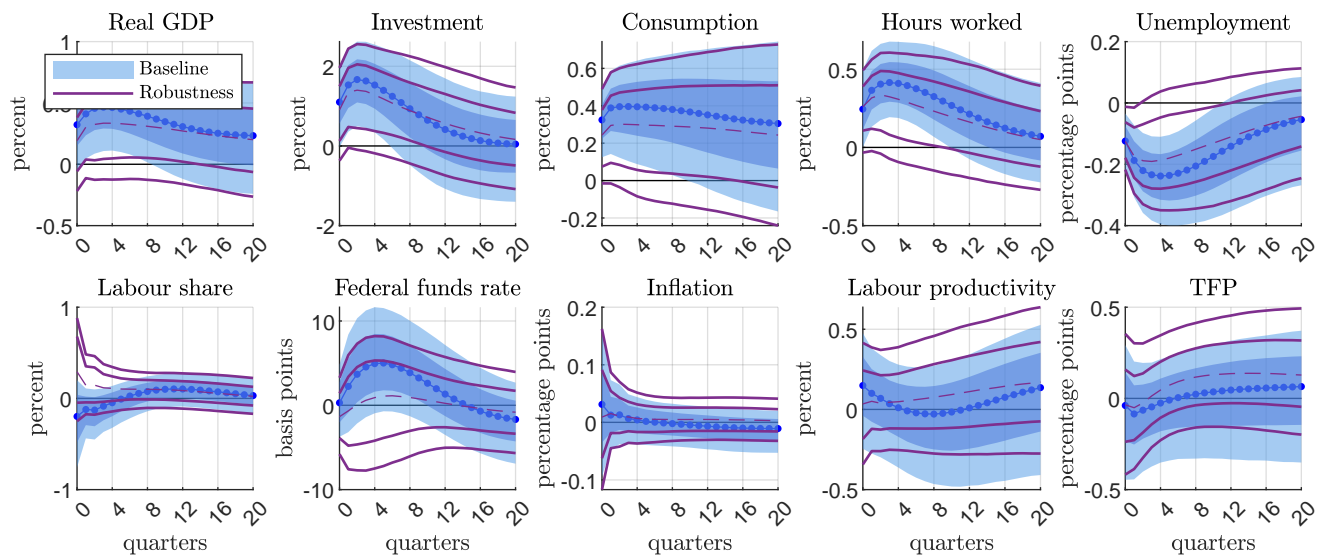
B) supply shock



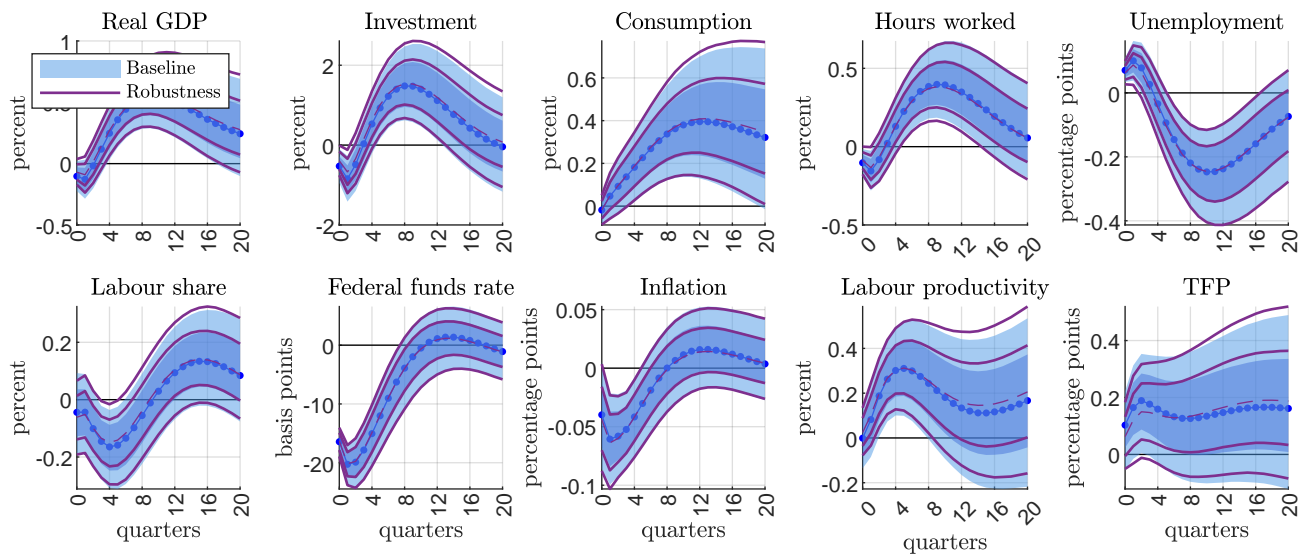
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure H-19: Robustness for flat prior on ν : impulse responses

A) Demand shock



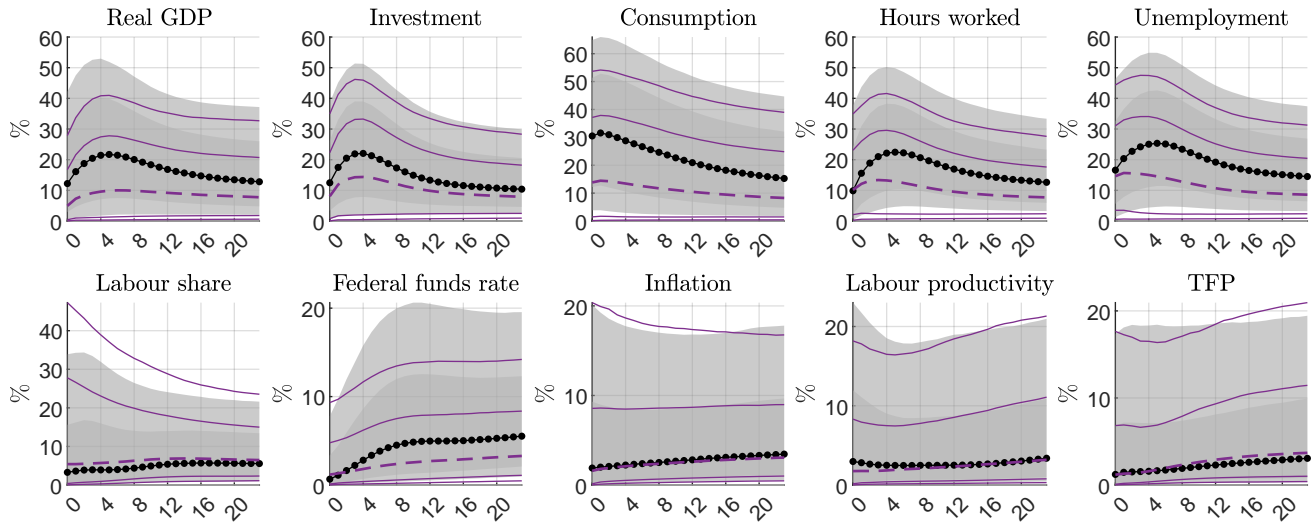
B) supply shock



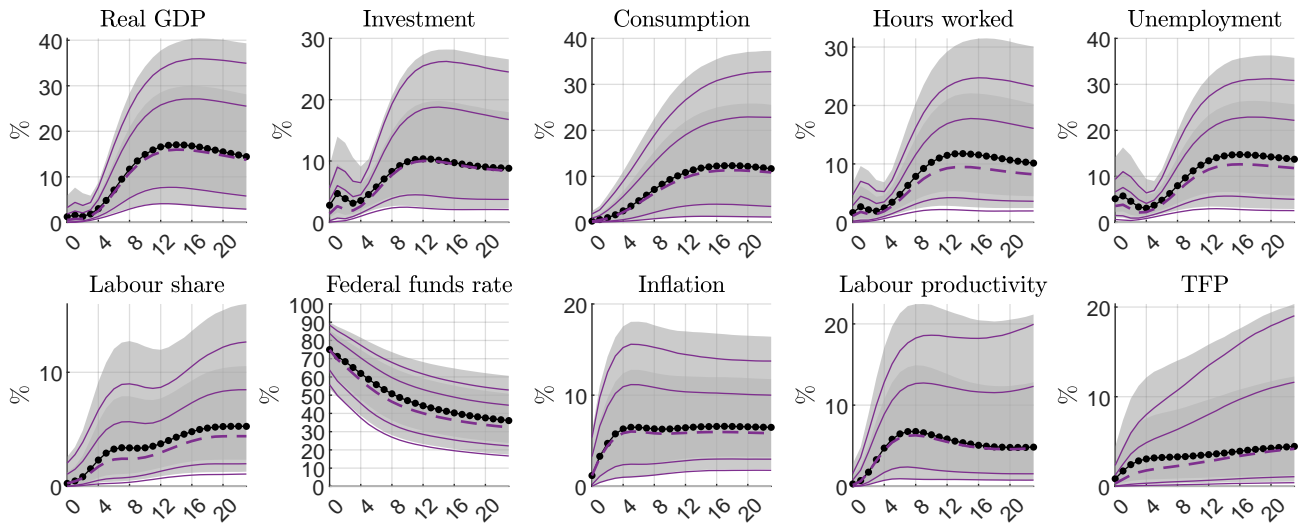
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure H-20: Robustness for flat prior on v : forecast error variance decomposition

A) Demand shock



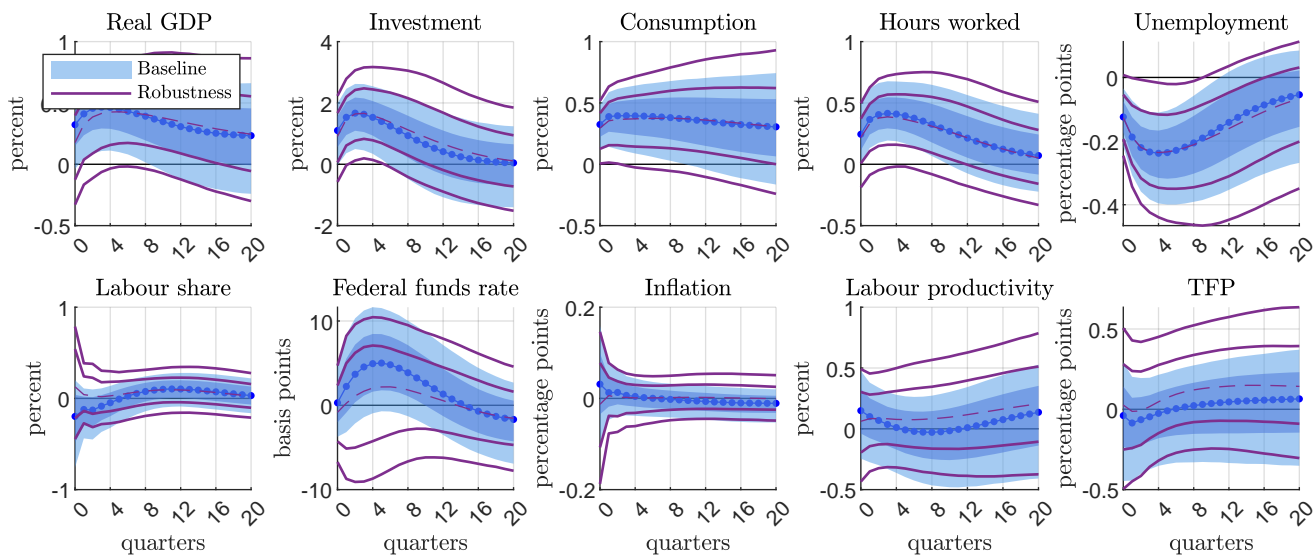
B) supply shock



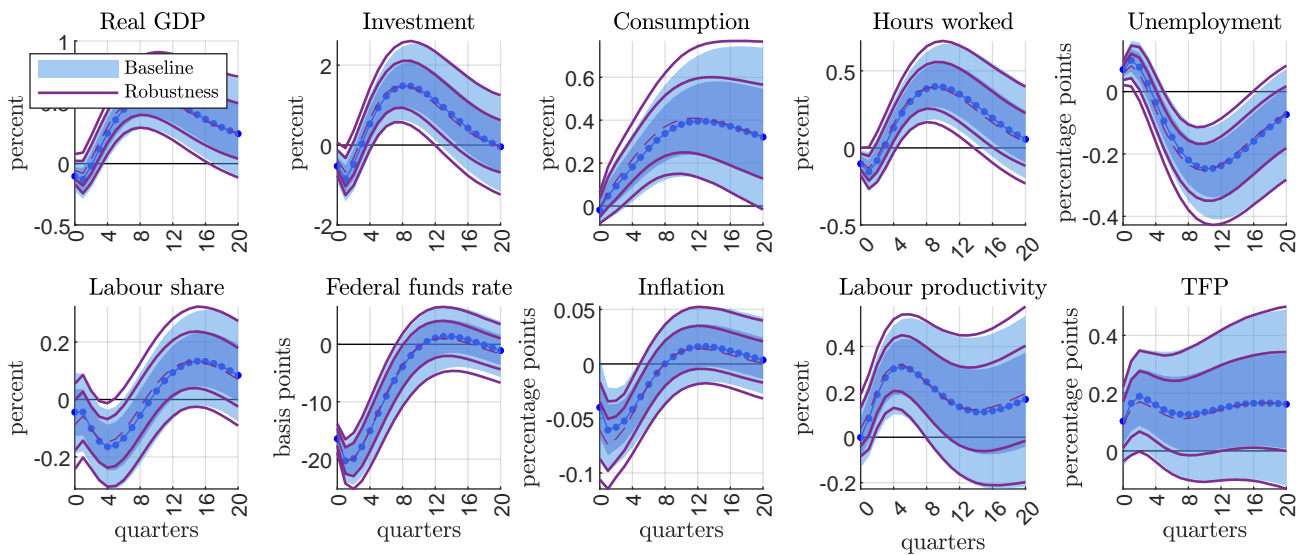
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure H-21: Robustness for looser prior on ϕ : impulse responses

A) Demand shock



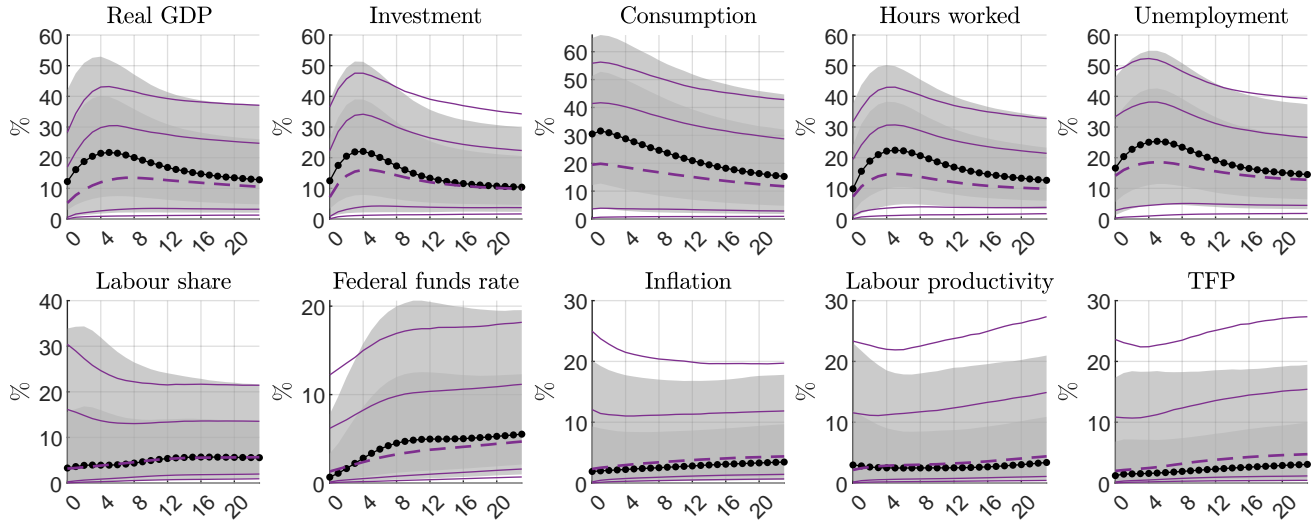
B) supply shock



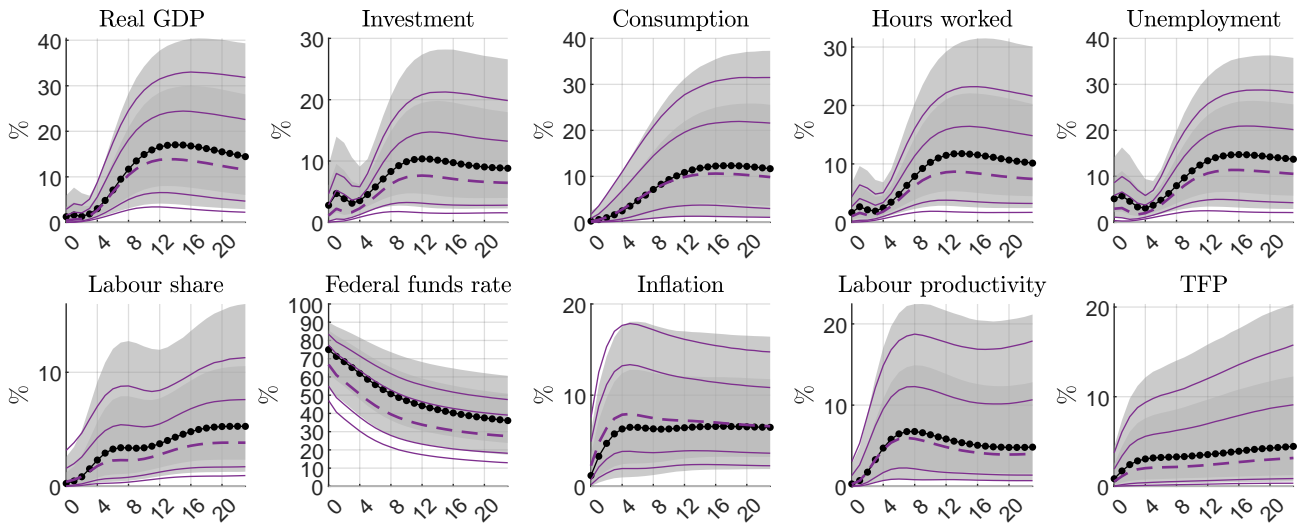
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure H-22: Robustness for looser prior on ϕ : forecast error variance decomposition

A) Demand shock



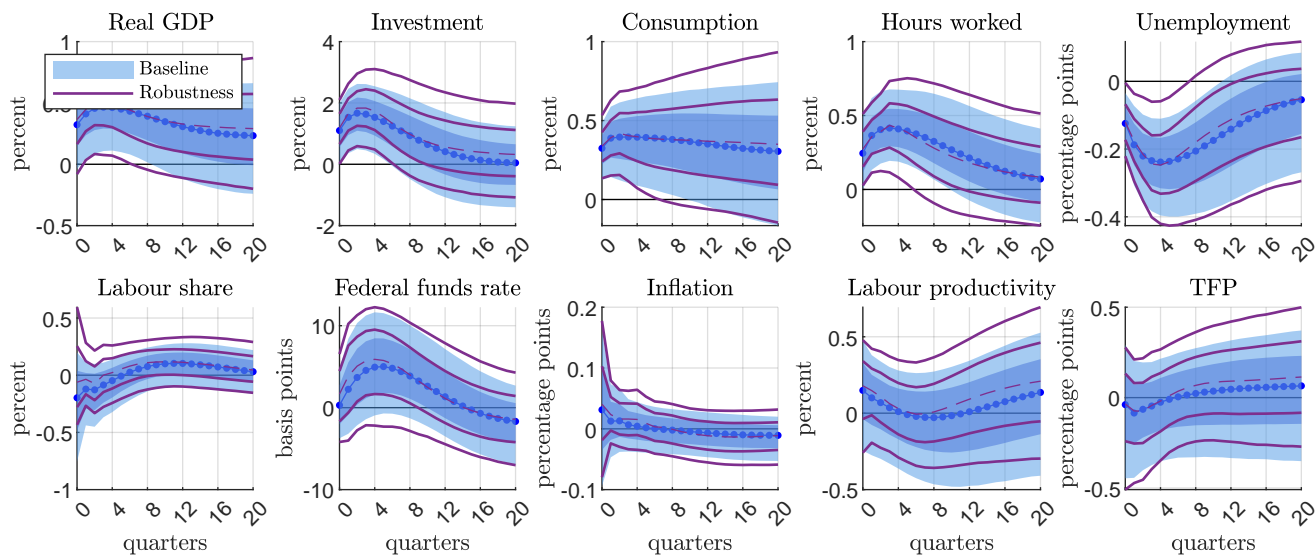
B) supply shock



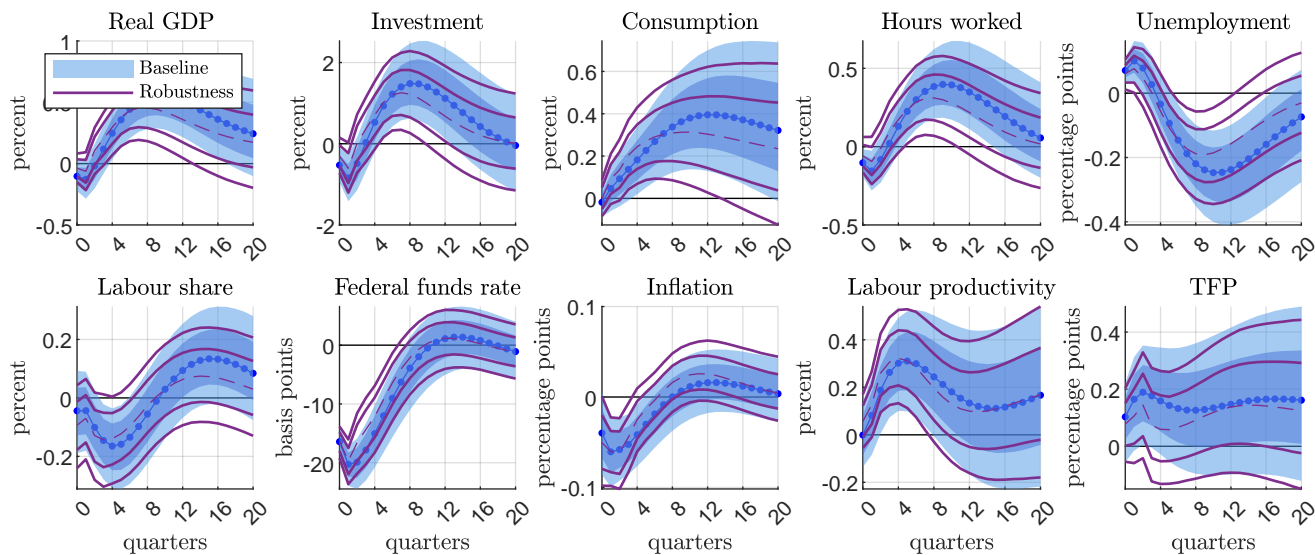
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure H-23: Robustness for $p = 4$: impulse responses

A) Demand shock



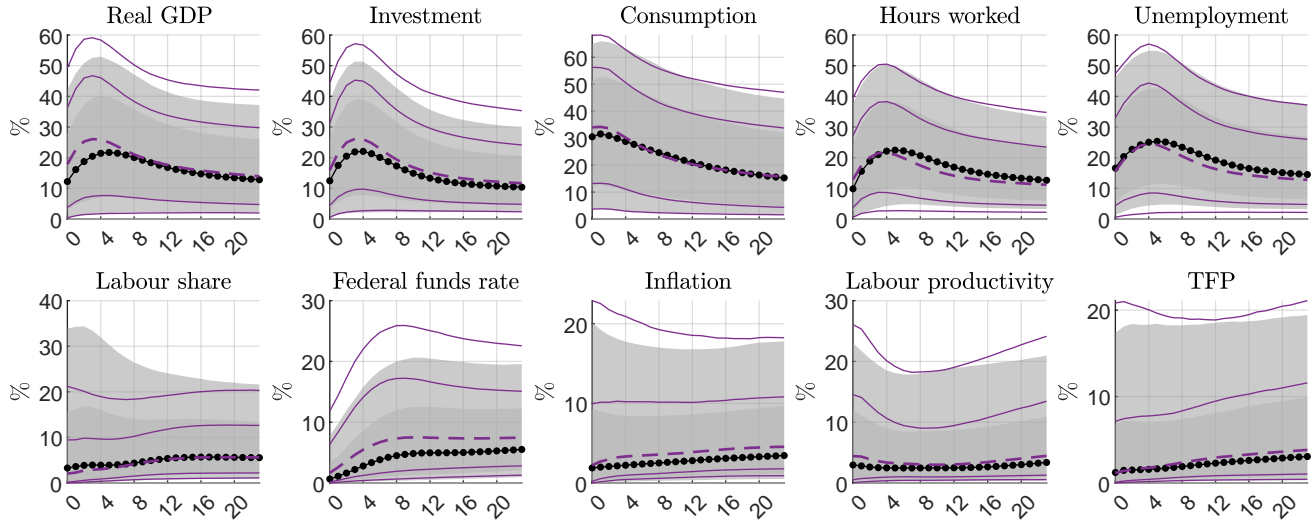
B) supply shock



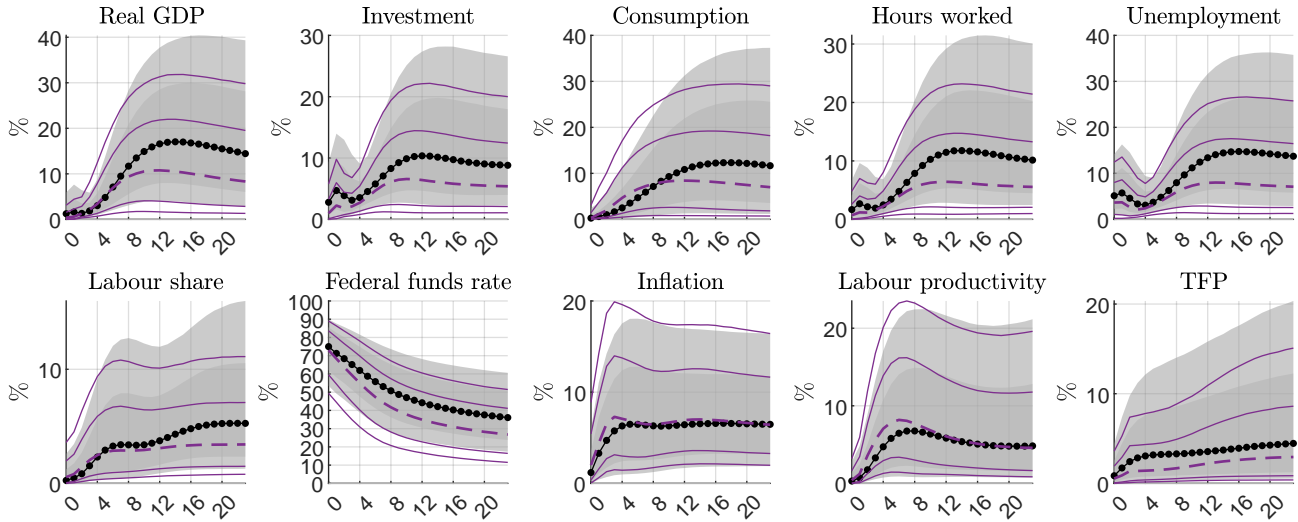
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure H-24: Robustness for $p = 4$: forecast error variance decomposition

A) Demand shock



B) supply shock



Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.