

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 $p(\phi Y, B, D, \mathbf{v})$	C-11
C.2 Conditional posterior $p(D Y, \phi, B, \mathbf{v})$	C-12
C.3 Conditional posterior $p(\mathbf{v} Y, \phi, B, D)$	C-13
C.4 Conditional posterior $p(B Y, \phi, D, \mathbf{v})$	C-14
D Identification up to sign and ordering	D-18
E Summary of the algorithm to sample from the posterior of a SVAR with t -distributed structural shocks	E-22
F Additional material on the simulation exercise	F-24
G Additional material on the application to the US GDP	G-36

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 $\nu = (\nu_1, \nu_2)'$. We normalize the variance of these shocks to 1 (which implies that we assume the variance to be finite and thus $\nu_1, \nu_2 > 2$). 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 $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. Our goal is to identify (B, ν) , or equivalently, (B_c, Q, ν) .

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.

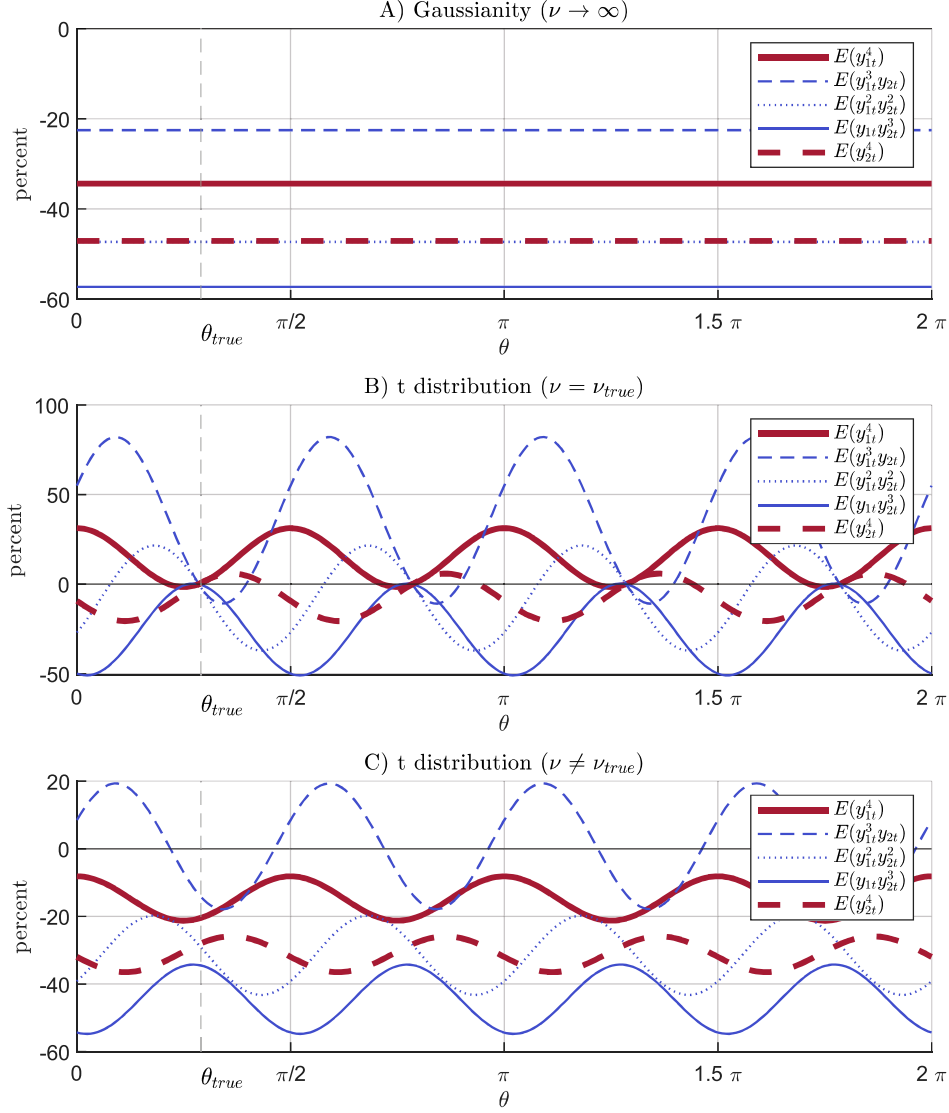
Things change dramatically when evaluating (A-4) using $\kappa_i = 6$ for $i = 1, 2$, which coincides with the true value. The middle panel of Figure A-1 shows that the fourth moments in (A-4) are no longer constant in θ , hence fourth moments provide additional information for identification. In addition, 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}, v_{true}) . The top panel evaluates equations (A-4) for $\kappa_i = 3$, $i = 1, 2$, which is associated with a Gaussian model ($v_1 = v_2 = \infty$). The middle and lower panel set κ_i implied by $v = v_{true} = 6$ or $v = 9$.

(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

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

and $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$ are independently t -distributed structural shocks with degrees of freedom $\nu = (v_1, v_2)'$ 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 ϵ_t are *i.i.d.* with zero mean. Individual components of ϵ_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 $\mathbf{v} = (v_1, \dots, v_k)'$, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)'$. 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:

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

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

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

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

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

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

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

The k stochastic terms in the new specification, \mathbf{e}_t , are Gaussian with the identity covariance matrix. The latent variables d_{it} are treated as unknown parameter with inverse Gamma prior that is independent across i and t , parametrized according to

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\mathbf{z}_t \sim N(\mathbf{0}, A^{-1} D_t A^{-1})$, 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

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 $-1 \cdot e_i'A^{-1}\hat{a}_i = e_i'I_k^*A^{-1}\hat{a}_i = e_i'(AI_k^*)^{-1}\hat{a}_i$, where I_k^* is the $k \times k$ identity matrix except its i -th diagonal element is set to -1 . This is exactly the LP normalization for Gaussian SVARs, rewritten in WZ's notation.

In our general setup, we need to compute the signed permutation matrix P_s that solves one of the equivalent problems in (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 Hy_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 $\text{tr}\{P_s G\} = p_1 g_{\pi(1),1} + p_2 g_{\pi(2),2} + \cdots + p_k g_{\pi(k),k}$, 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 $p_1 g_{\pi(1),1} + \cdots + p_i g_{\pi(i),i} + \cdots + p_k g_{\pi(k),k} < p_1 g_{\pi(1),1} + \cdots + p_i^* g_{\pi(i),i} + \cdots + p_k g_{\pi(k),k}$, i.e. contradiction. Hence at the maximum:

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

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 $\text{tr}\{P \cdot |G|\} = |g_{\pi(1),1}| + |g_{\pi(2),2}| + \cdots + |g_{\pi(k),k}|$, 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 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);
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);
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 16 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

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.

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.

F Additional material on the simulation exercise

The parameter values of the data generating process are

$$B = \begin{pmatrix} 0.60 & 0.40 \\ 0.70 & -0.70 \end{pmatrix}, \quad (\text{F-125})$$

$$\mathbf{v} = \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \quad (\text{F-126})$$

$$\Pi_1 = \begin{pmatrix} 1.0612 & -0.0759 \\ -0.2502 & 1.1404 \end{pmatrix}, \quad (\text{F-127})$$

$$\Pi_2 = \begin{pmatrix} -0.0660 & 0.0093 \\ -0.0253 & -0.0905 \end{pmatrix}, \quad (\text{F-128})$$

$$\Pi_3 = \begin{pmatrix} -0.0641 & 0.0109 \\ 0.0286 & -0.0655 \end{pmatrix}, \quad (\text{F-129})$$

$$\Pi_4 = \begin{pmatrix} -0.0530 & 0.0119 \\ 0.0639 & -0.0434 \end{pmatrix}, \quad (\text{F-130})$$

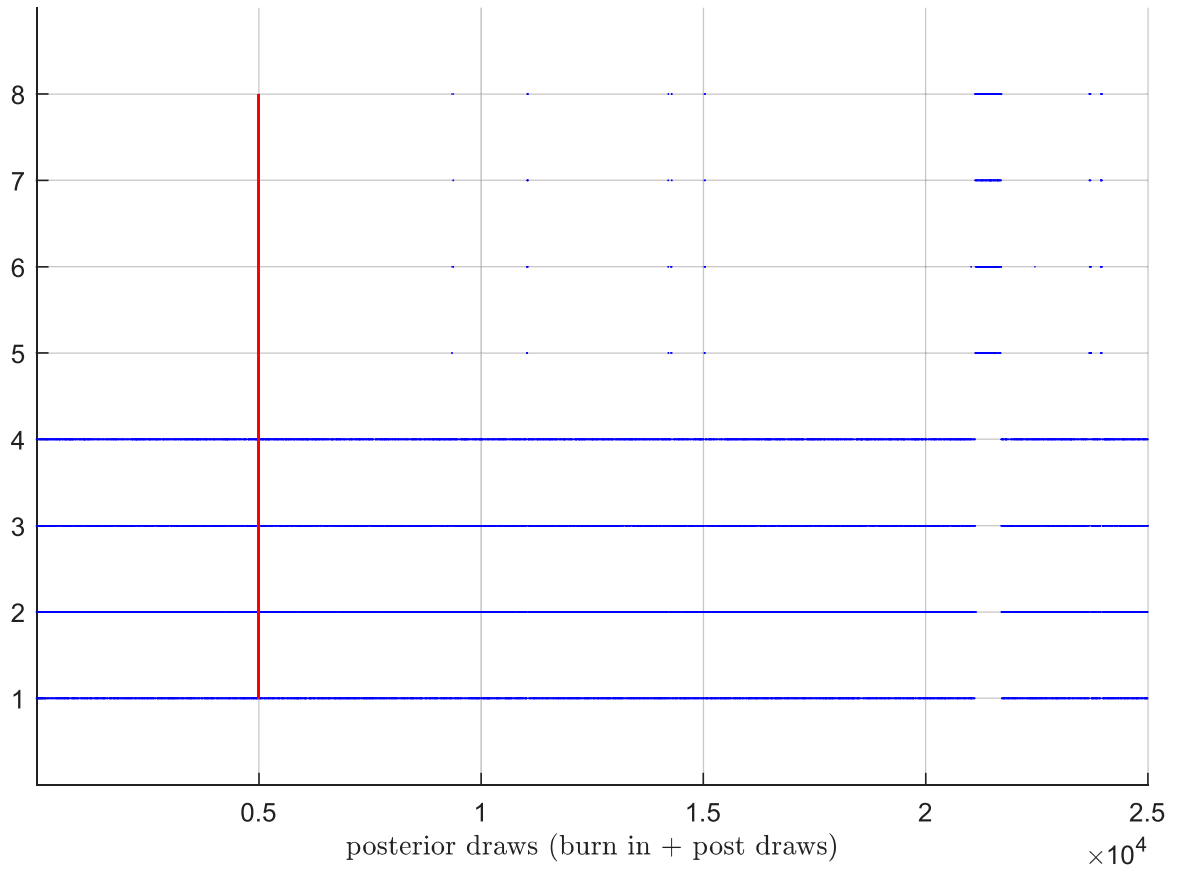
$$\Pi_5 = \begin{pmatrix} -0.0355 & 0.0113 \\ 0.0660 & -0.0304 \end{pmatrix}, \quad (\text{F-131})$$

$$\Pi_6 = \begin{pmatrix} -0.0165 & 0.0084 \\ 0.0425 & -0.0230 \end{pmatrix}. \quad (\text{F-132})$$

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

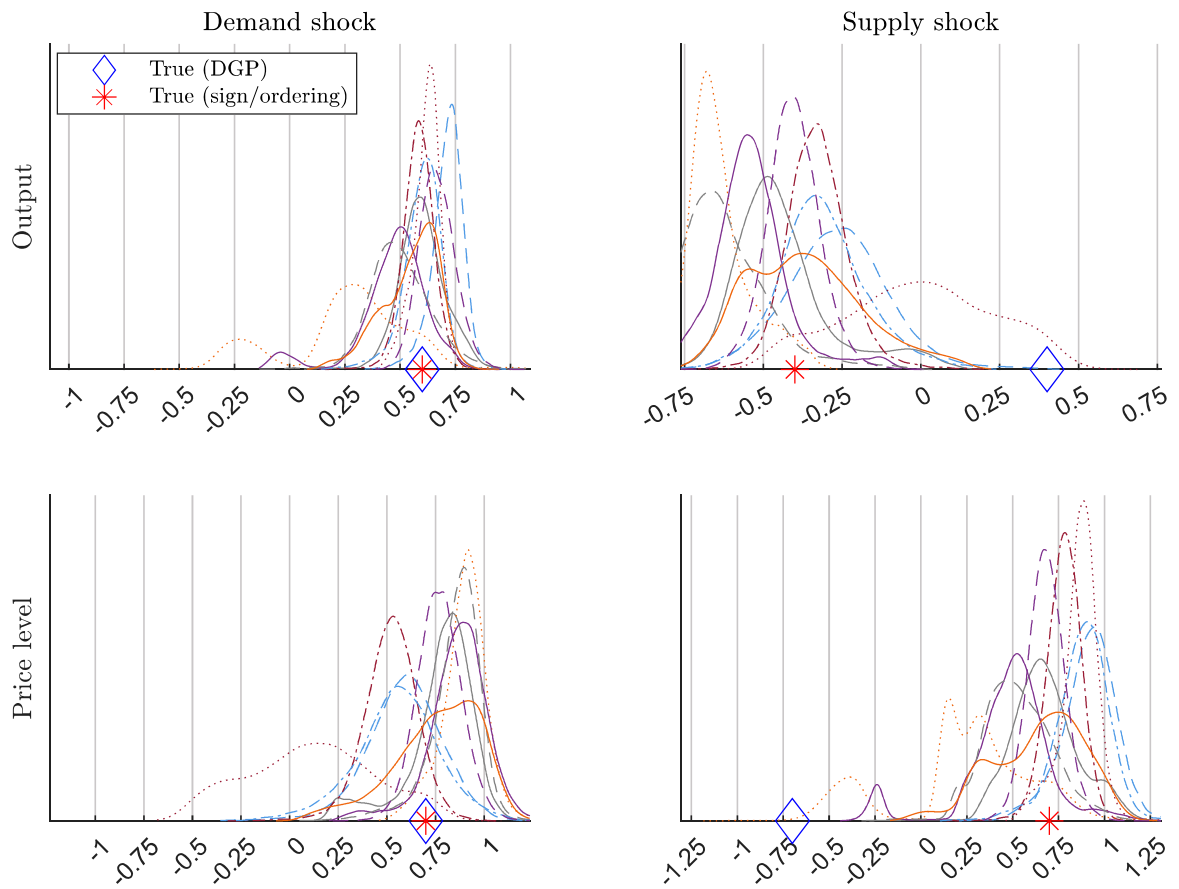
$$B_{target} = \begin{pmatrix} 0.7015 & -0.3279 \\ 0.7271 & 0.7236 \end{pmatrix}. \quad (\text{F-133})$$

Figure F-2: Generalized LP normalization



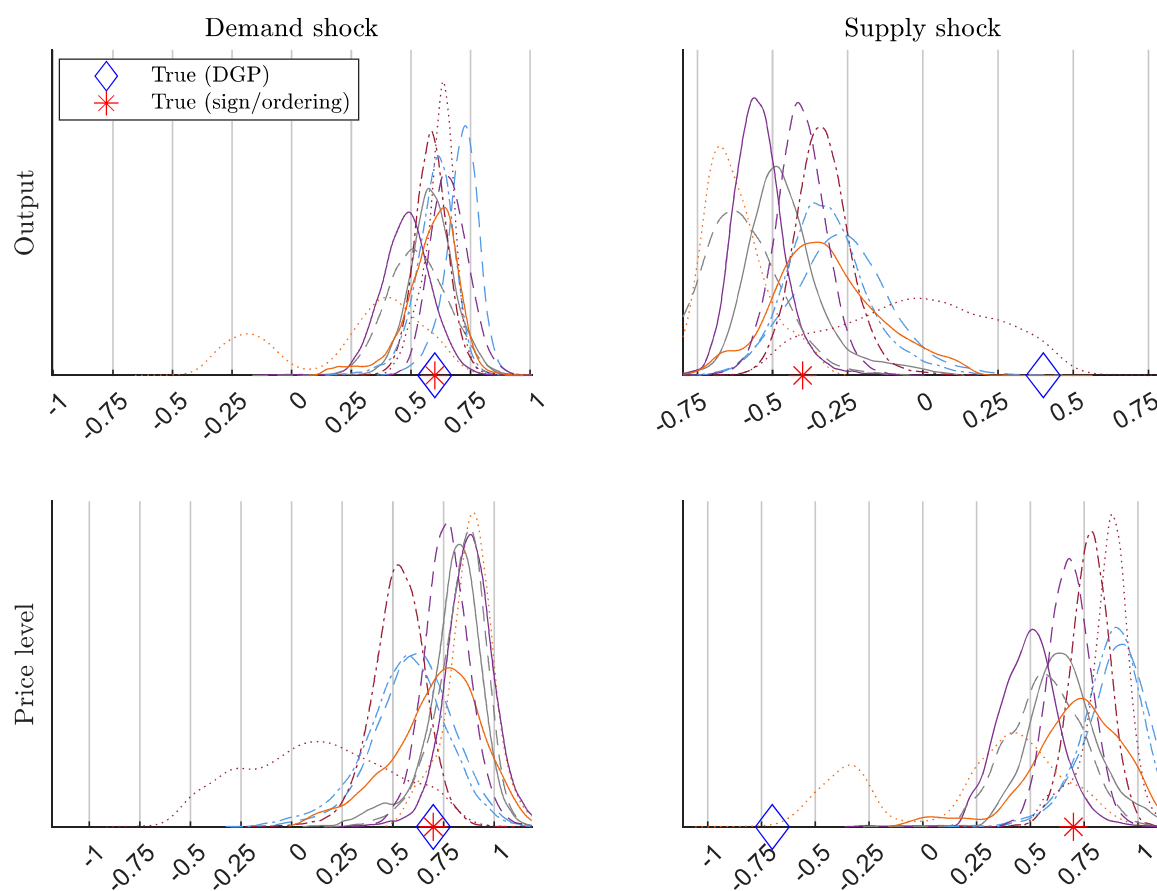
Note: The vertical red line shows when the burn-in draws end. For each of the 25,000 posterior draws, the figure indicates if the generalized LP normalization permutes the ordering of the columns of B (vertical values 5-8) or not (vertical values 1-4). It also indicates if the sign of the columns of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



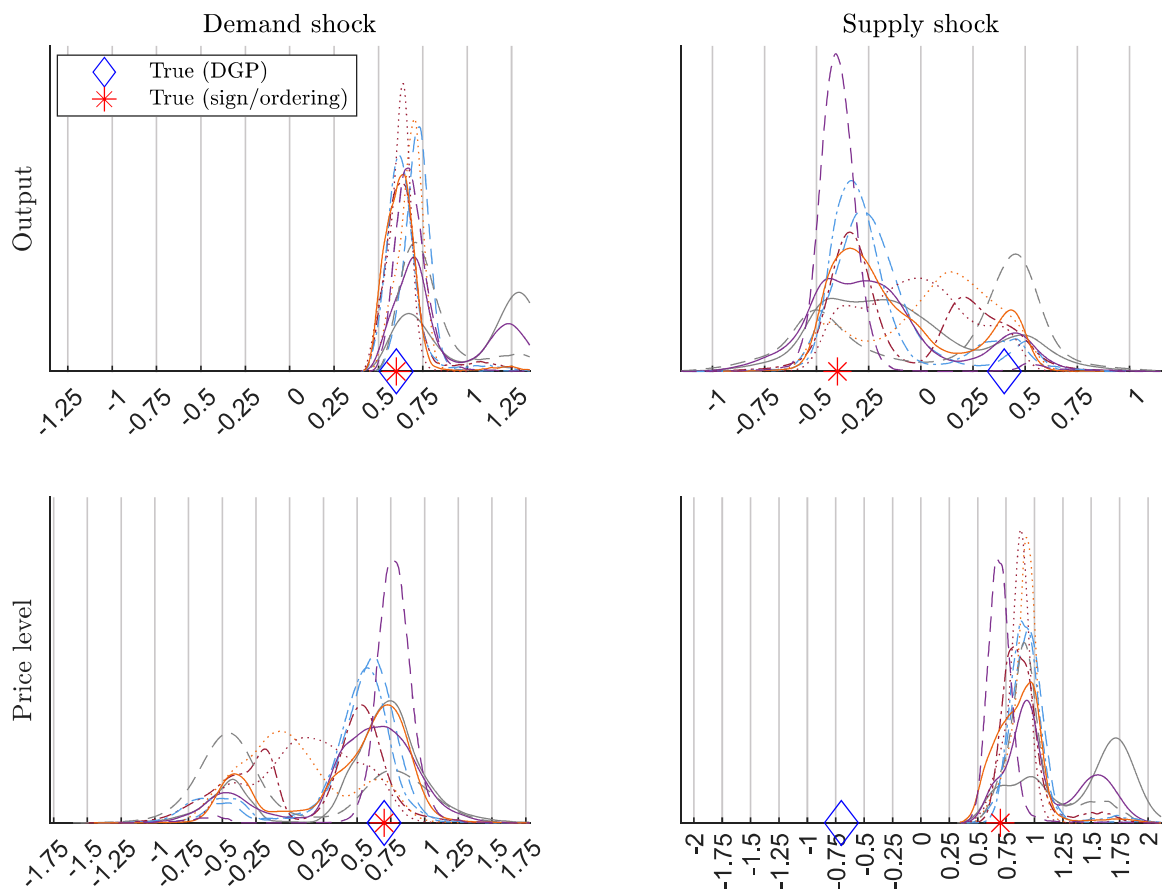
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

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



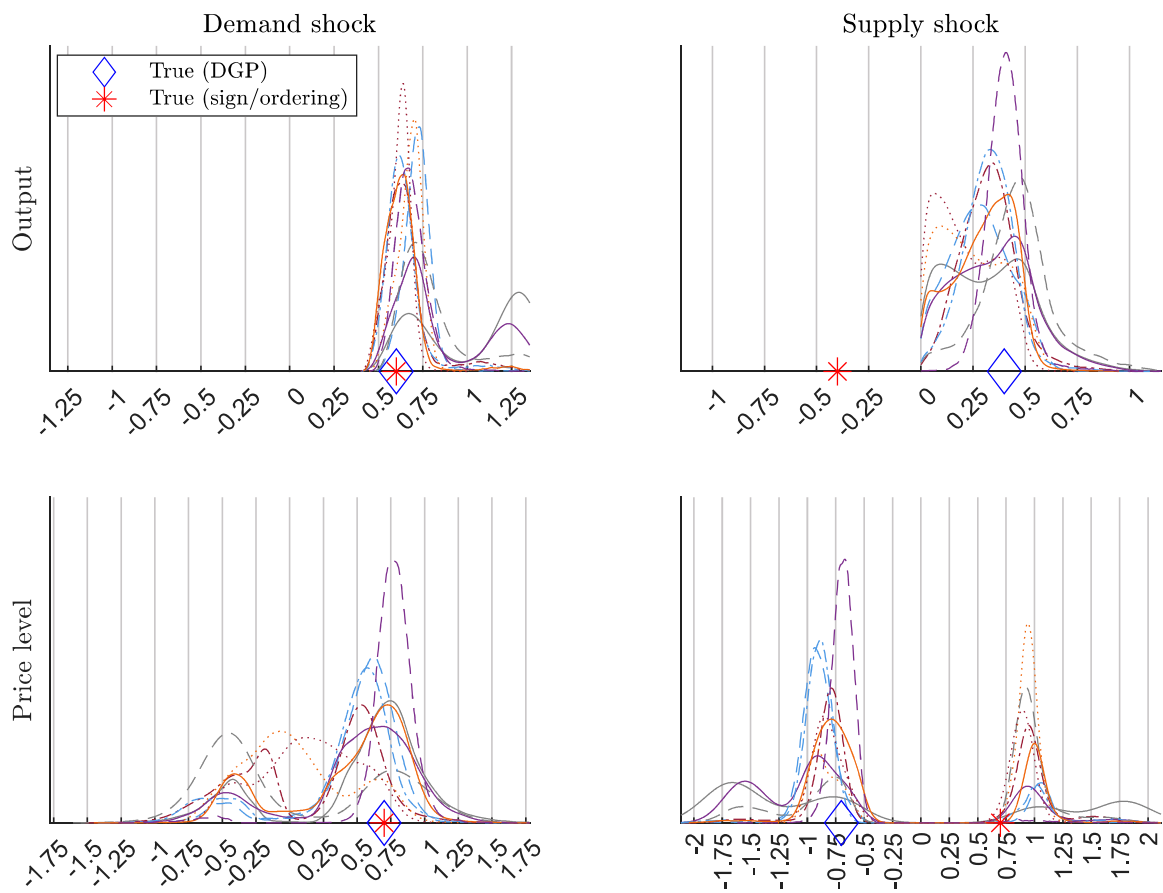
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

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



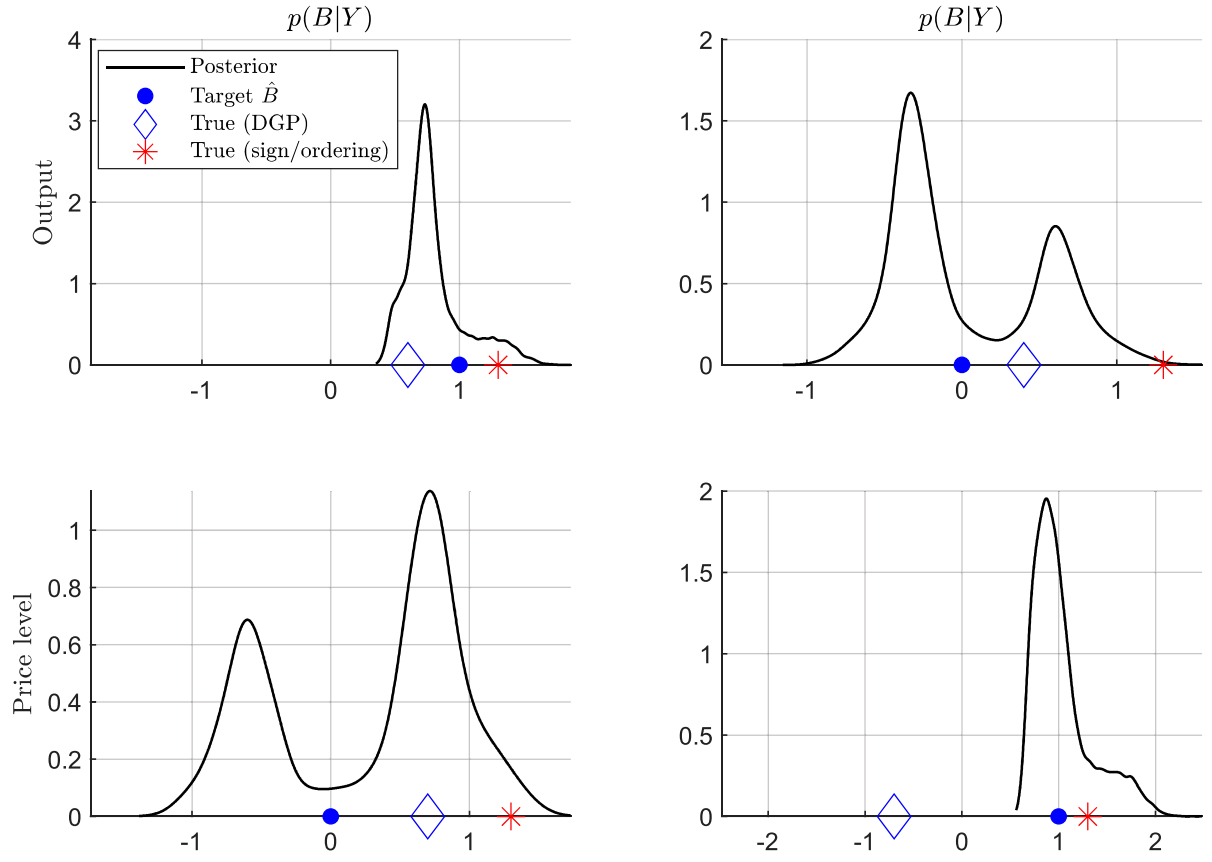
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

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



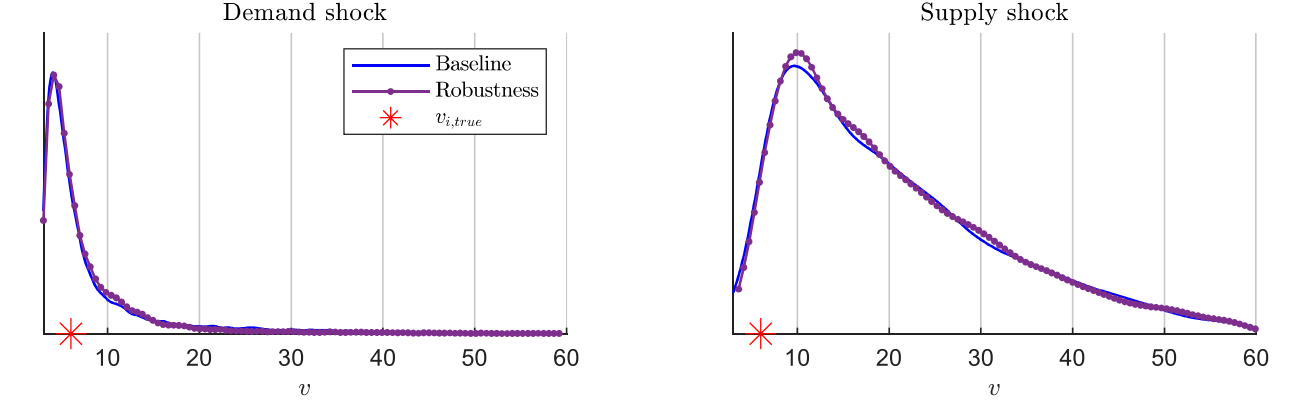
Note: Each line corresponds to the marginal posterior distribution estimated using one of the ten pseudo datasets generated in simulation.

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

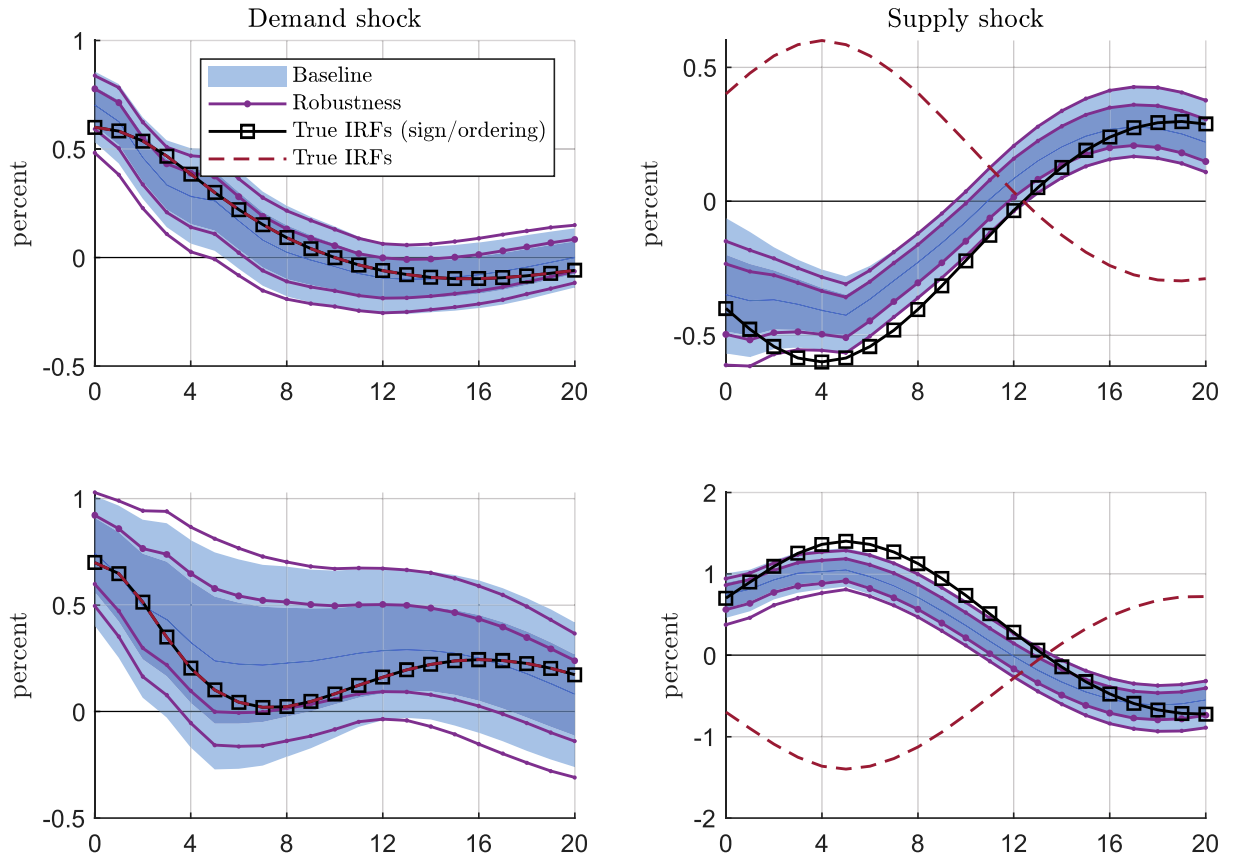


Note: The blue diamond indicates B_{true} . The blue dot indicates the target matrix \hat{B} used for the normalization. The red star indicates the sign/permutation of B_{true} that is the closest to \hat{B} . The continuous line shows the marginal posterior of the entries of B from 20,000 posterior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

Figure F-8: Robustness when computing the target matrix in the sampler
A) Degrees of freedom



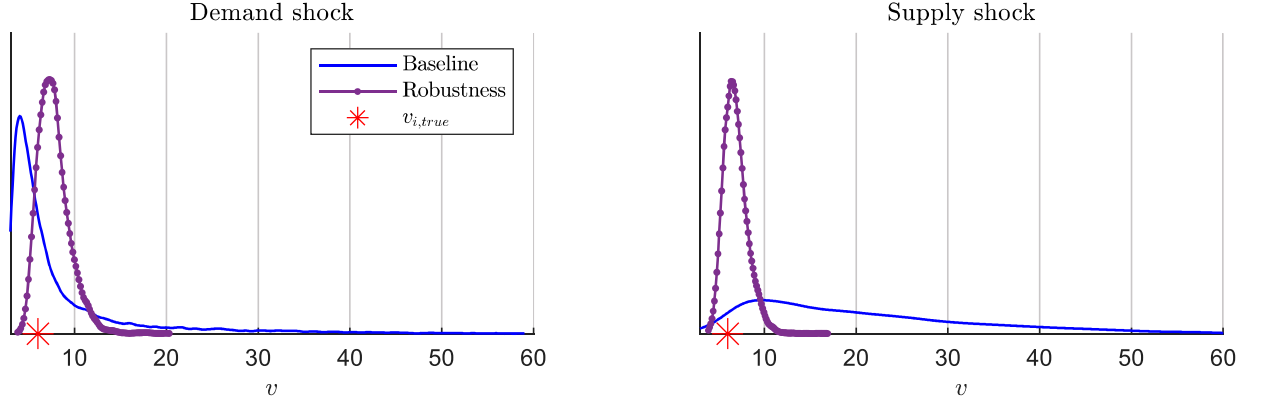
B) Impulse responses



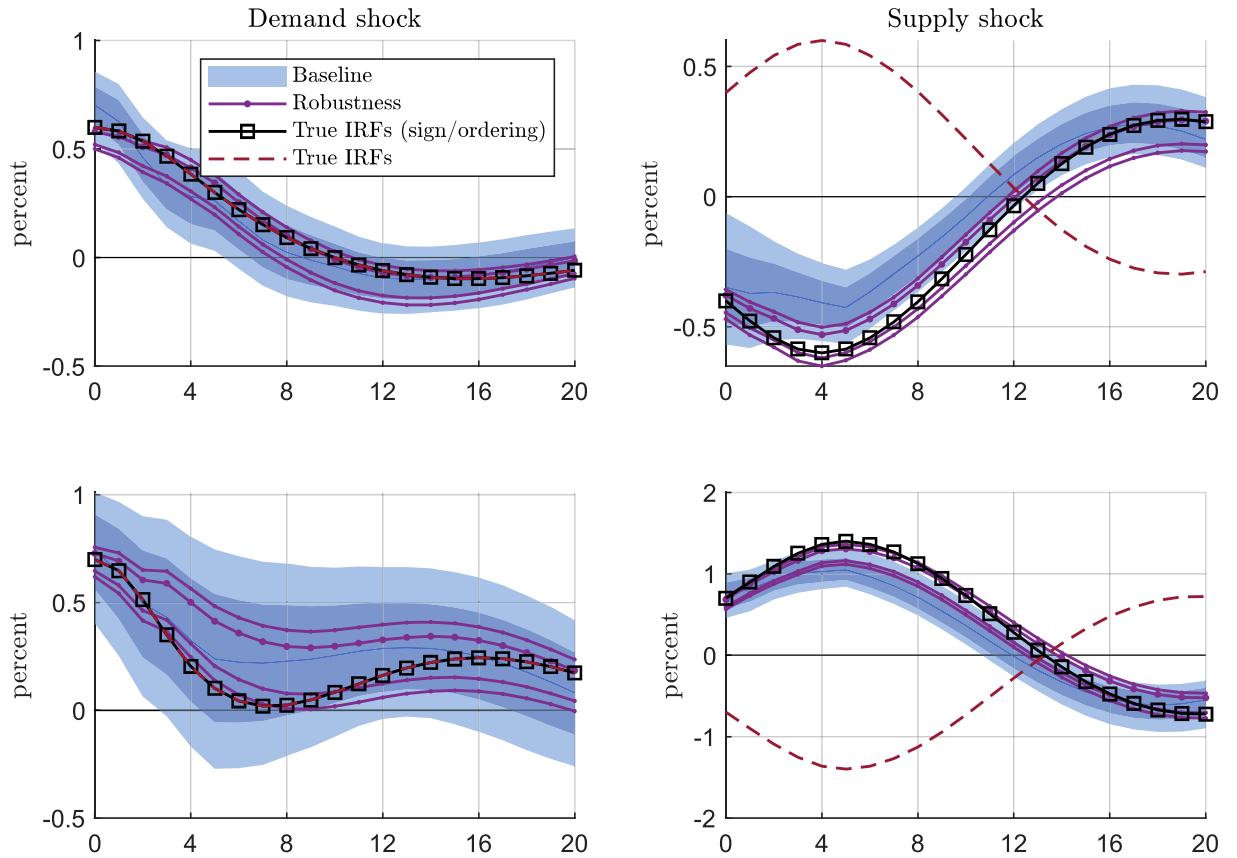
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

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

A) Degrees of freedom



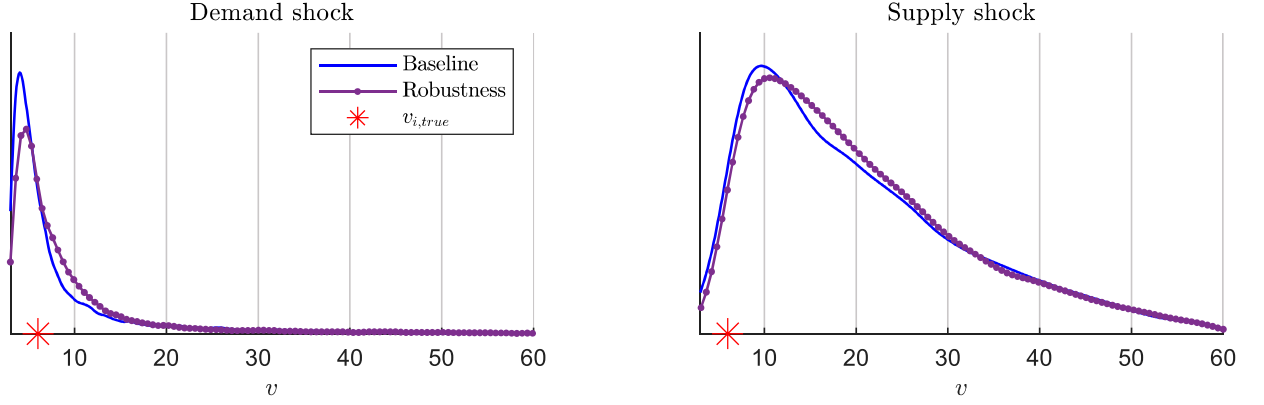
B) Impulse responses



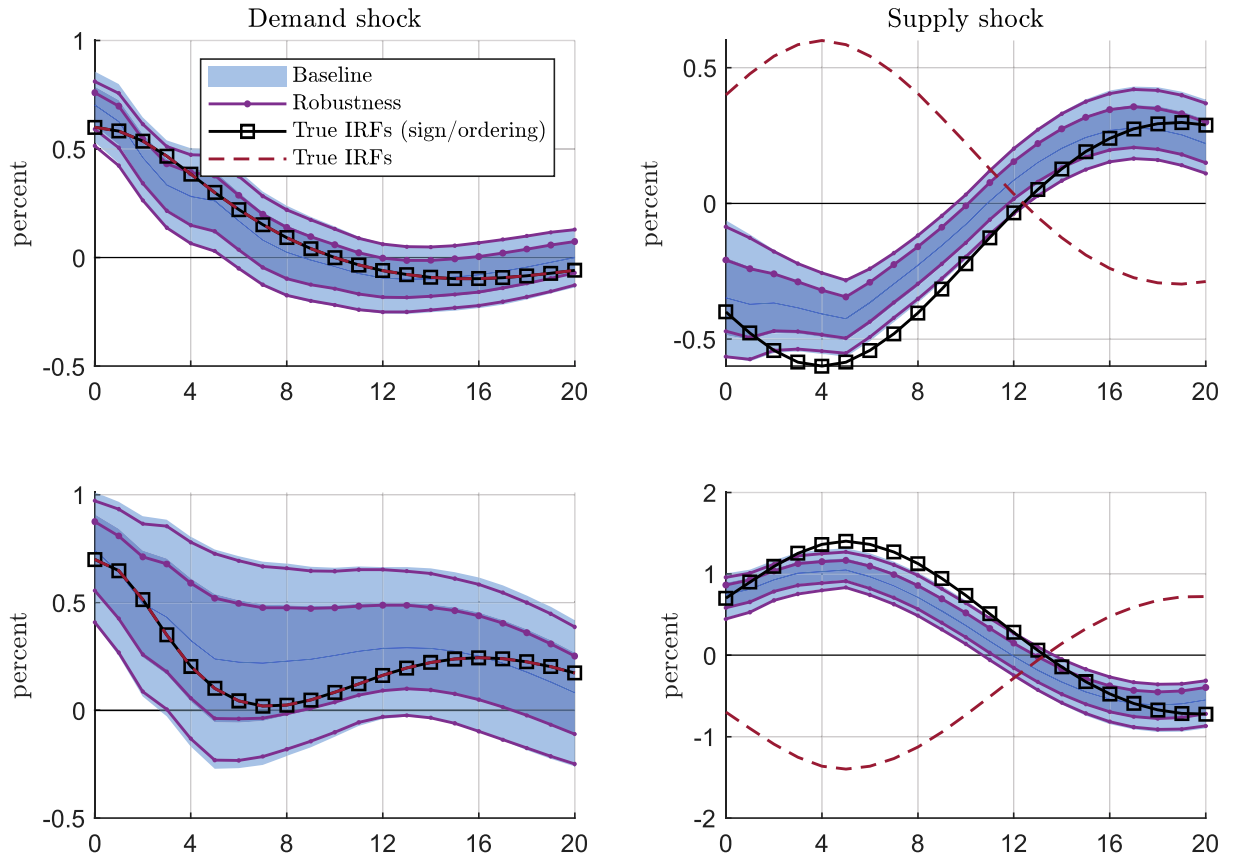
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure F-10: Robustness for flat prior on A

A) Degrees of freedom



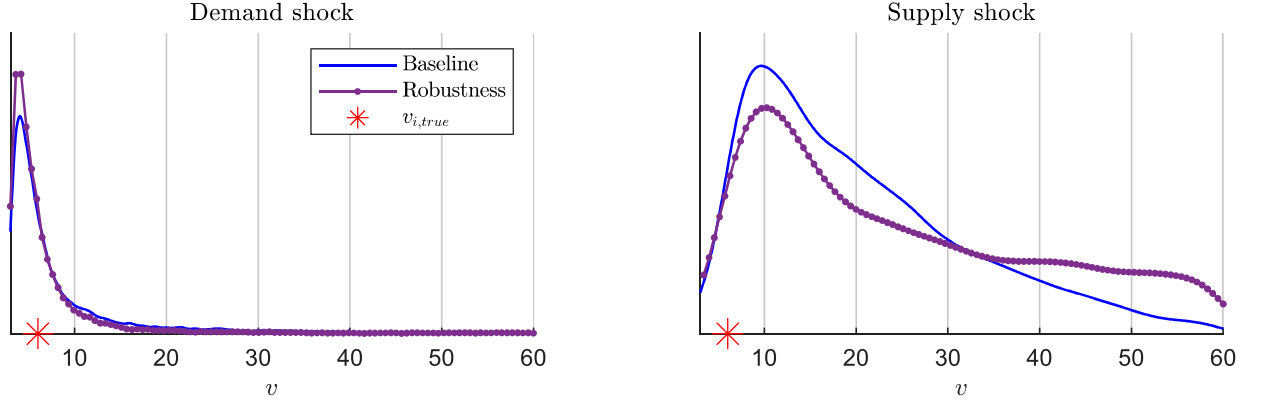
B) Impulse responses



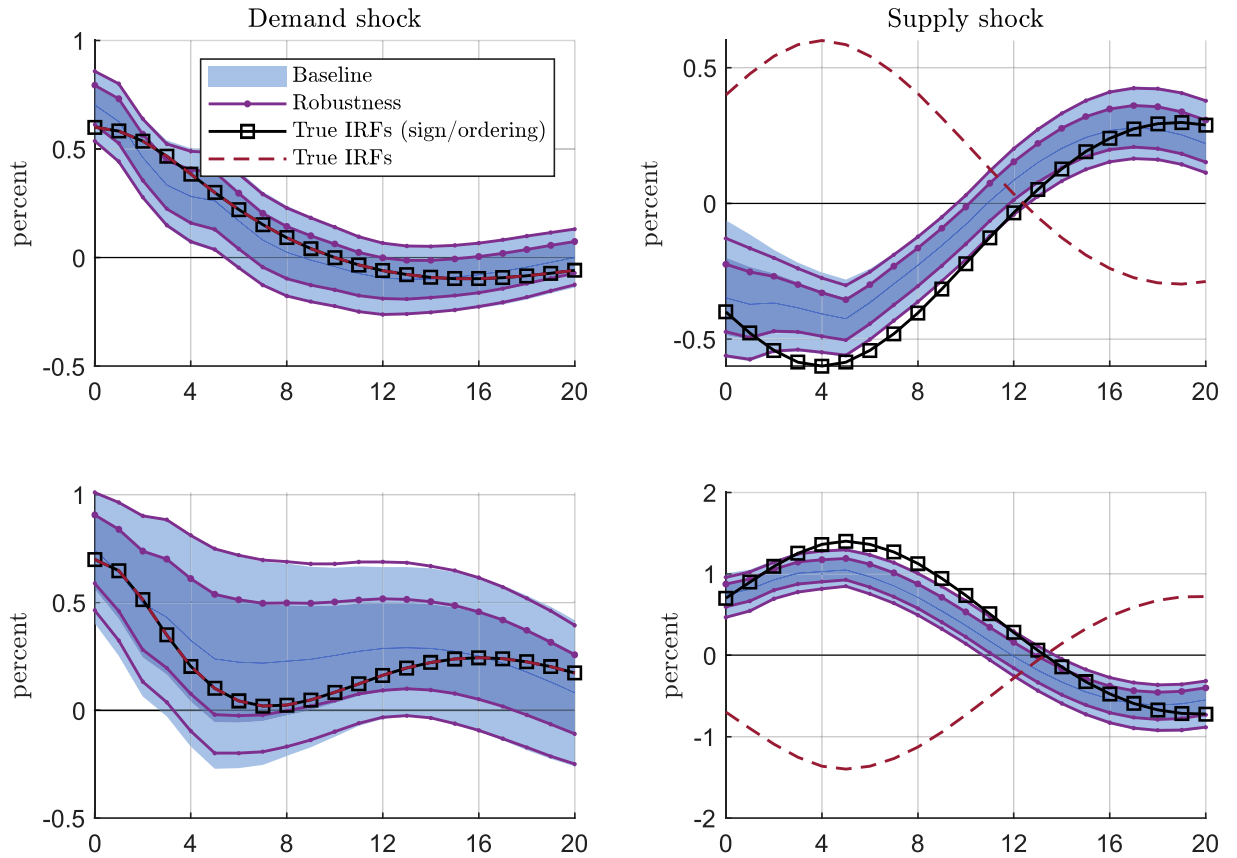
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure F-11: Robustness for flat prior on v

A) Degrees of freedom



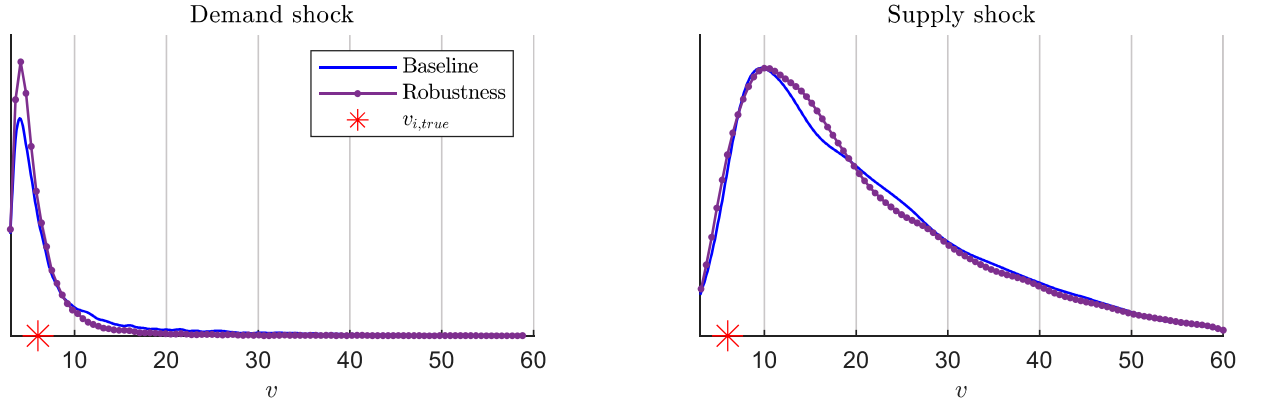
B) Impulse responses



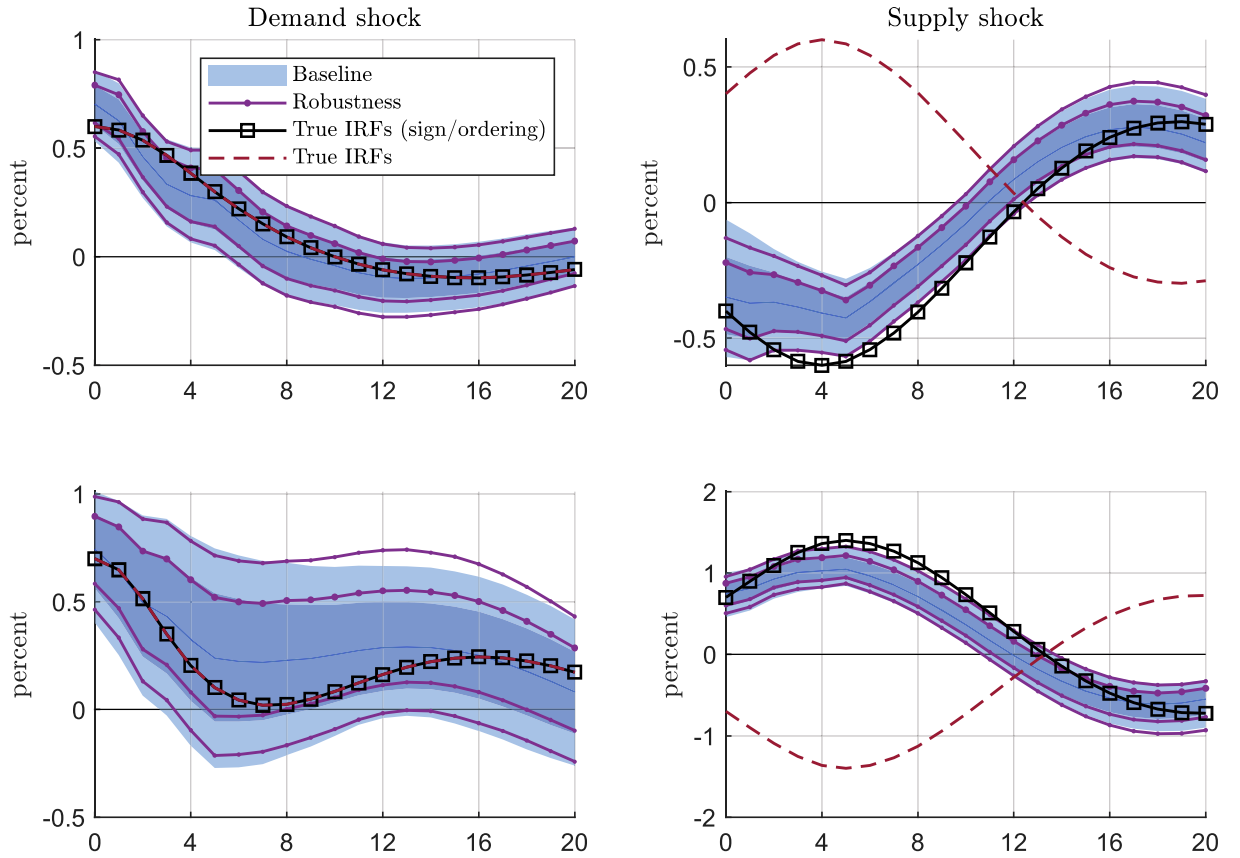
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

Figure F-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



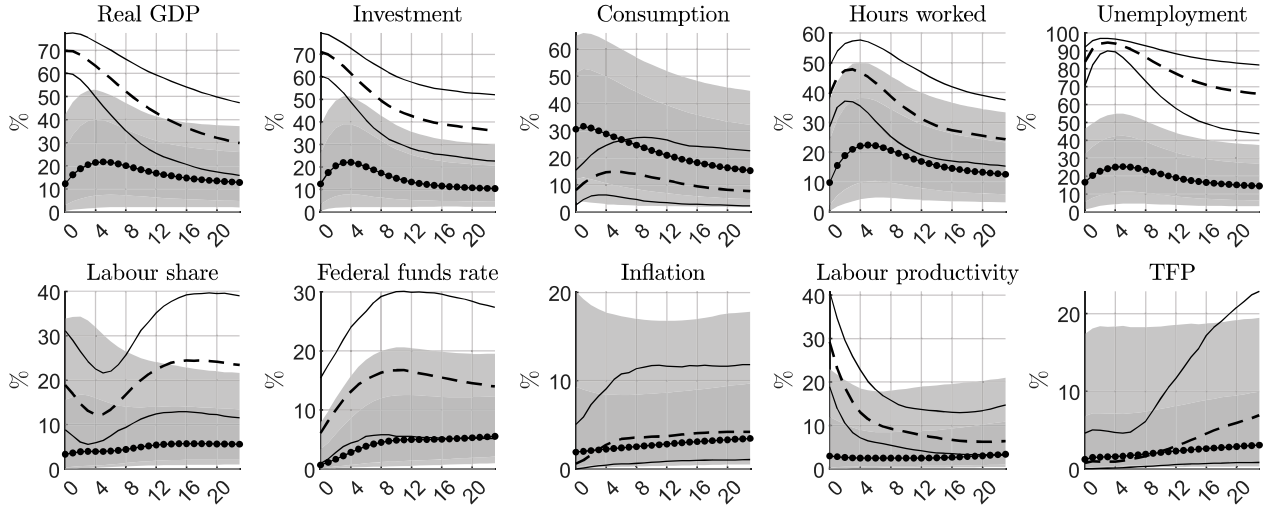
B) Impulse responses



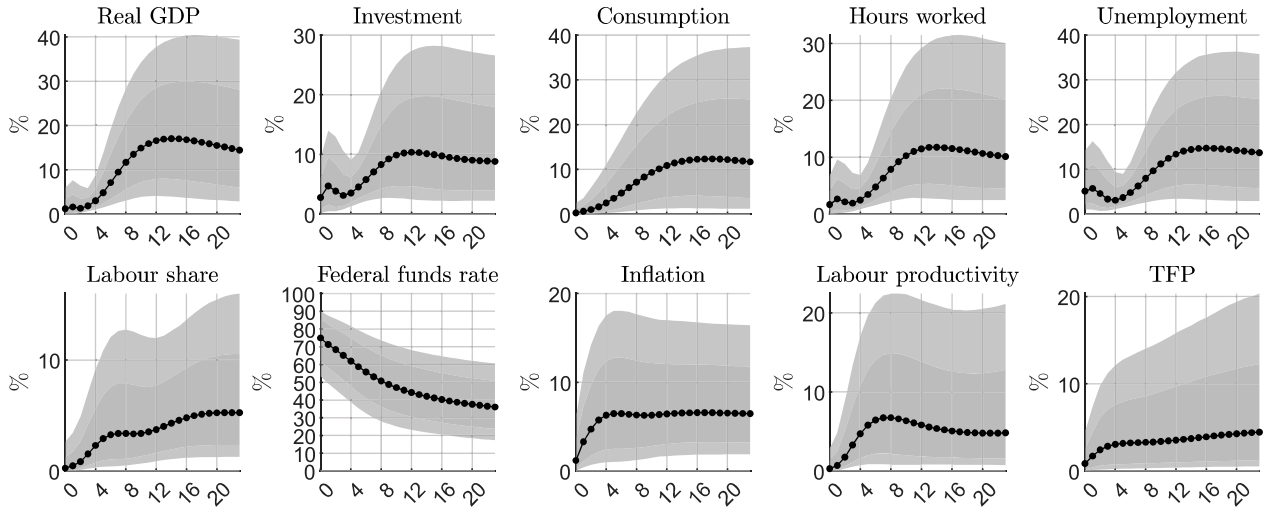
Note: In Panel A, the blue solid line shows the marginal posterior from the baseline specification while the thick purple line shows the marginal posterior under the alternative specification. In Panel B, the blue line and shaded areas show the pointwise median and credible sets corresponding to the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

G Additional material on the application to the US GDP

Figure G-13: Forecast error variance decomposition
Shock 3: demand shock



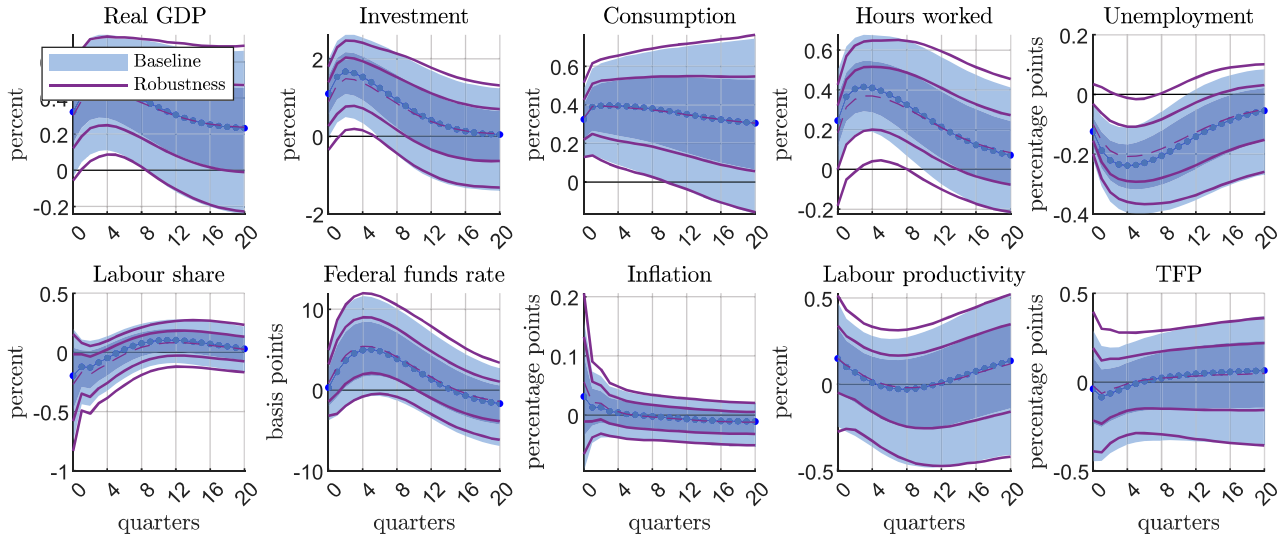
Shock 7: supply shock



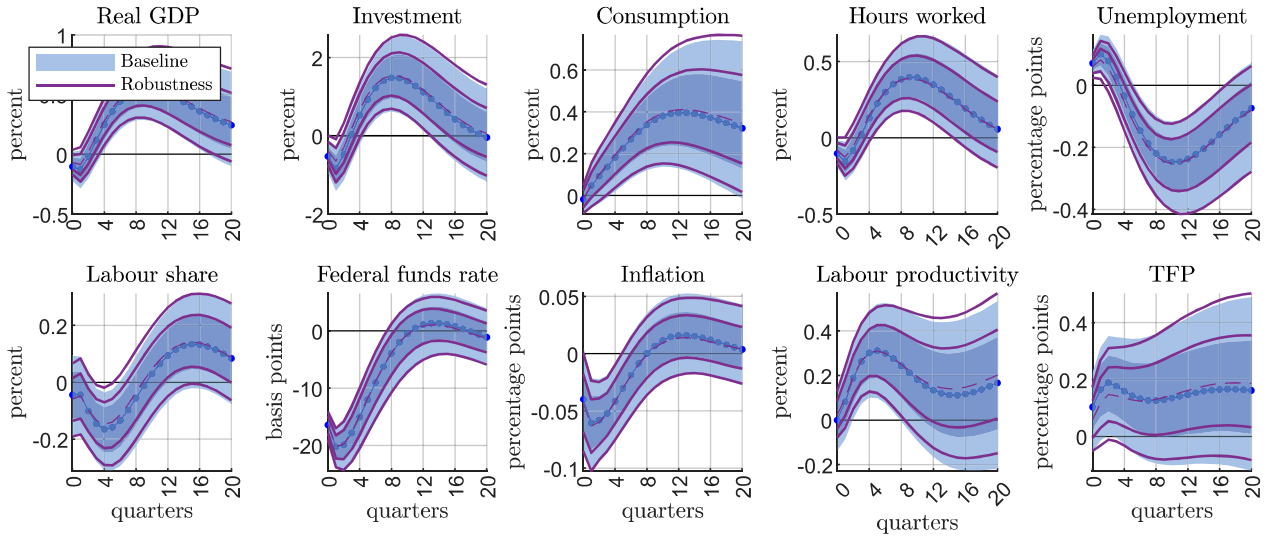
Note: The black dotted lines show the pointwise posterior median from our estimation, with corresponding 68% and 90% pointwise posterior credible sets shown with shaded areas. The black dashed and solid lines in the top figure show the pointwise median and 90% credible sets estimated by [Angeletos et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

Figure G-14: Robustness computing target matrix \hat{B} in the sampler: impulse responses

A) Demand shock



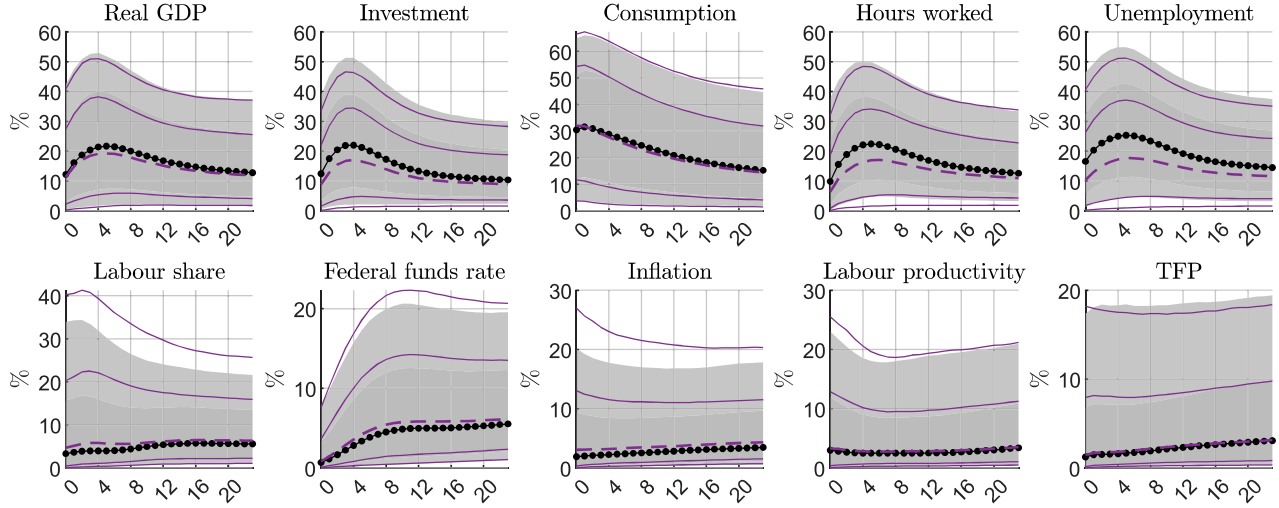
B) supply shock



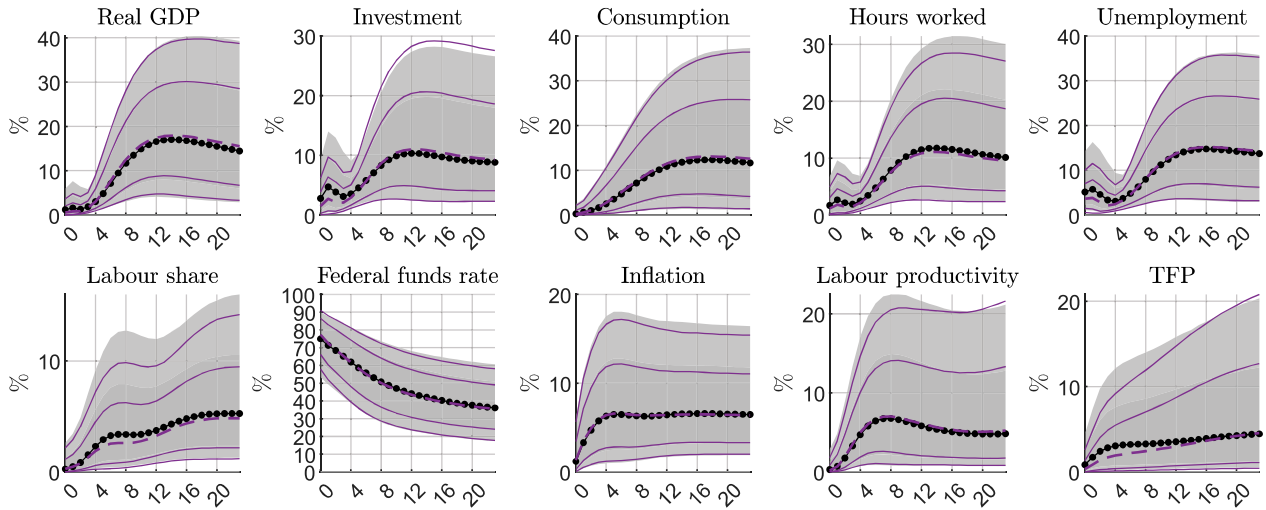
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure G-15: Robustness computing target matrix \hat{B} in the sampler: forecast error variance decomposition

A) Demand shock



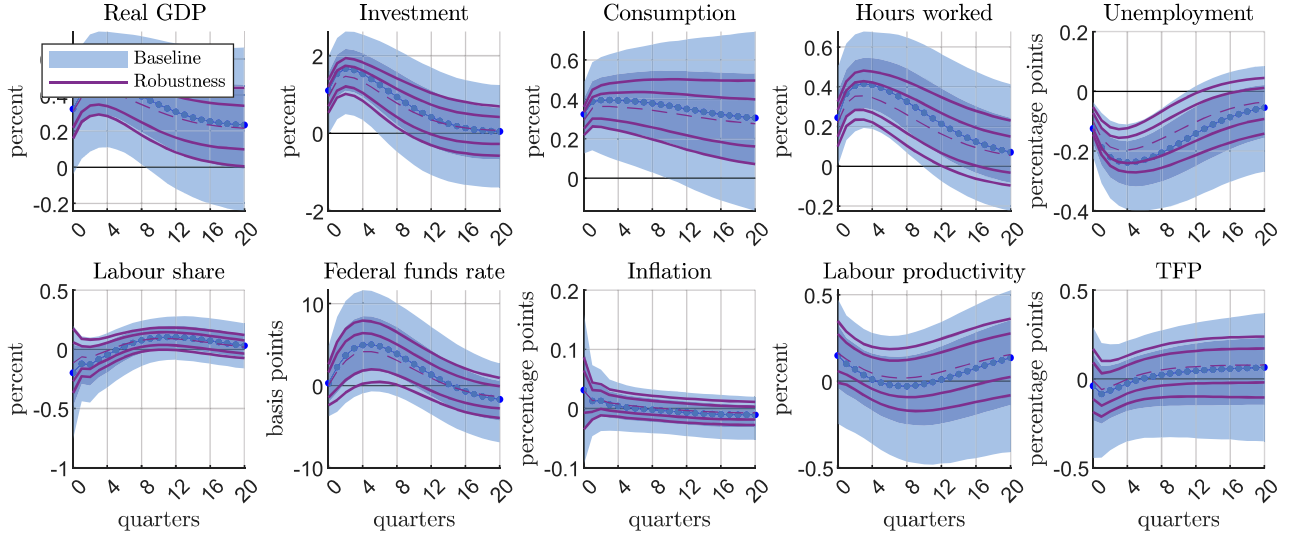
B) supply shock



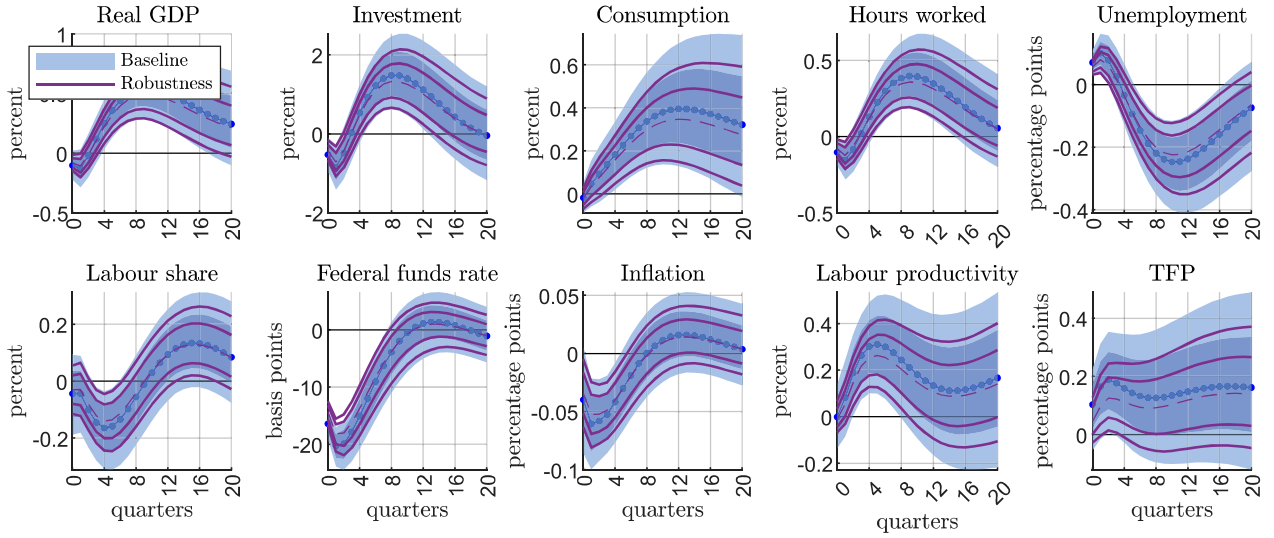
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure G-16: Robustness for flat prior on A : impulse responses

A) Demand shock



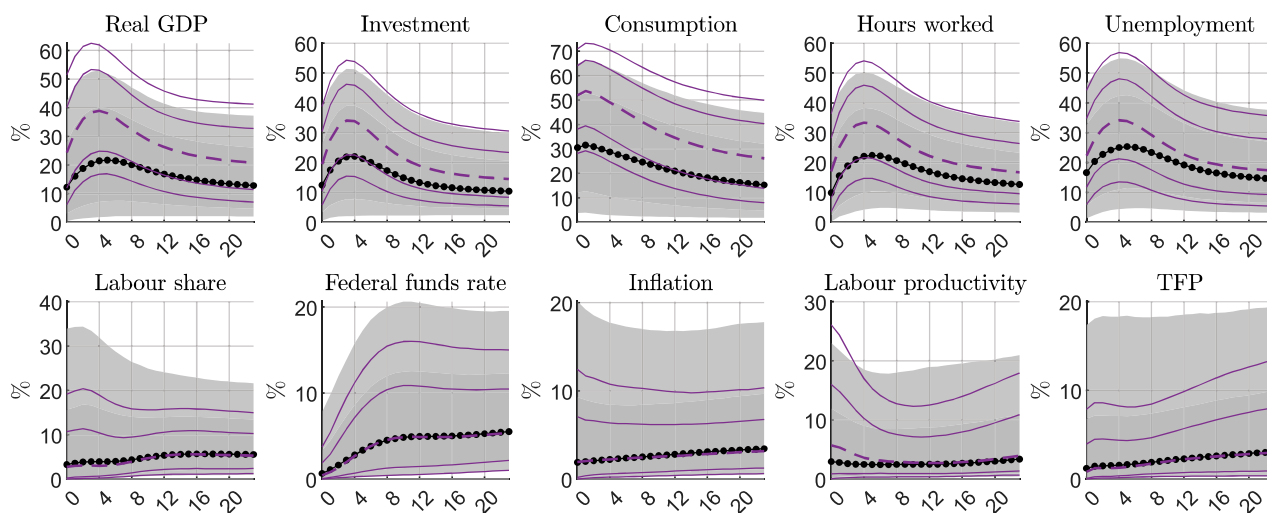
B) supply shock



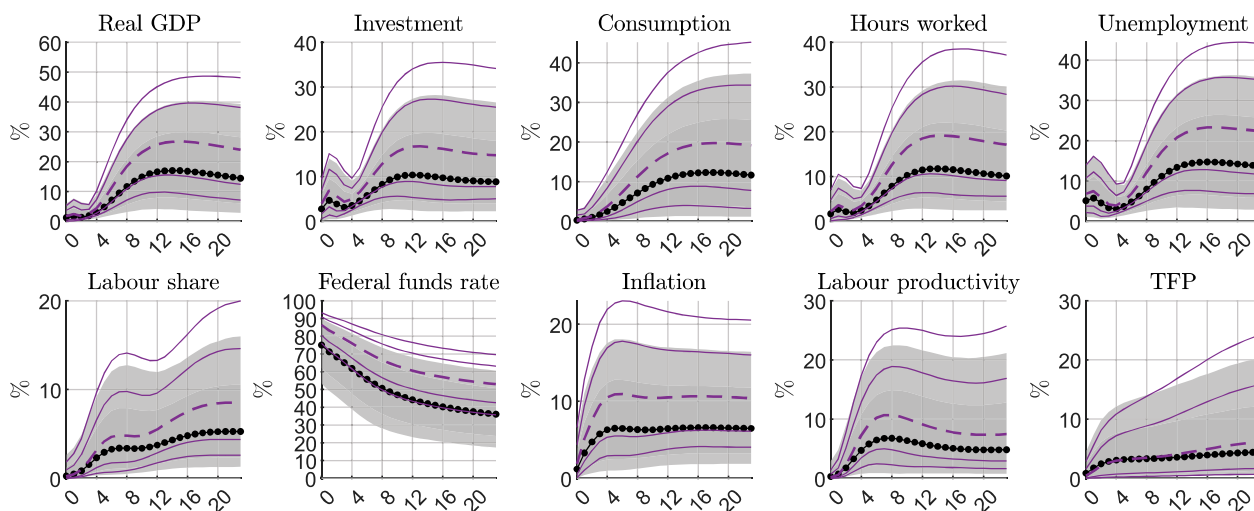
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure G-17: Robustness for flat prior on A : forecast error variance decomposition

A) Demand shock



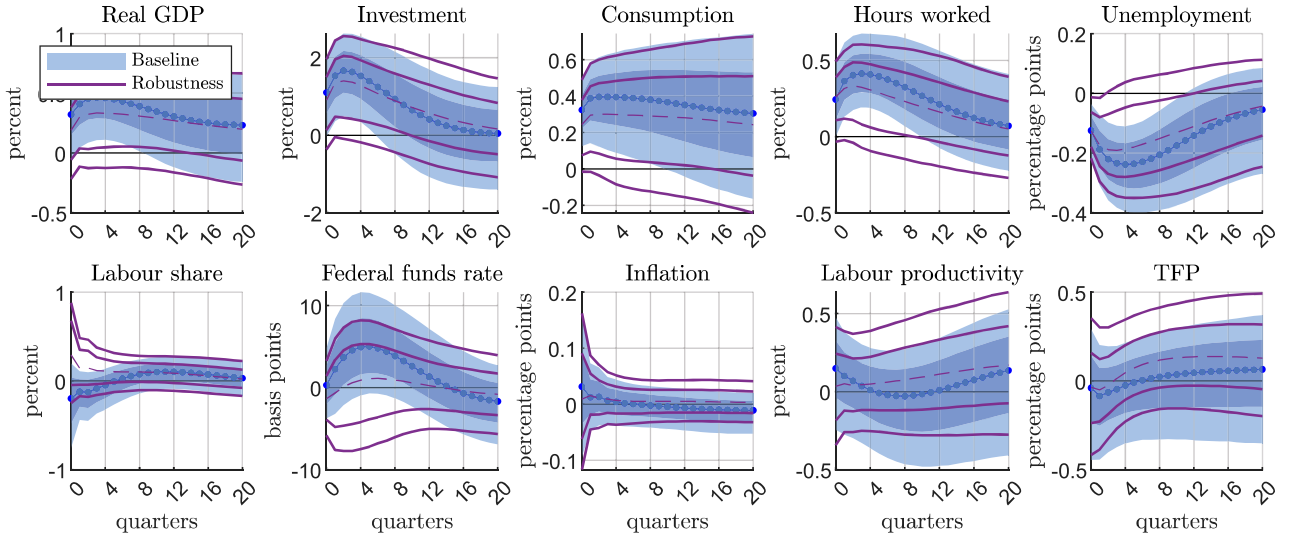
B) supply shock



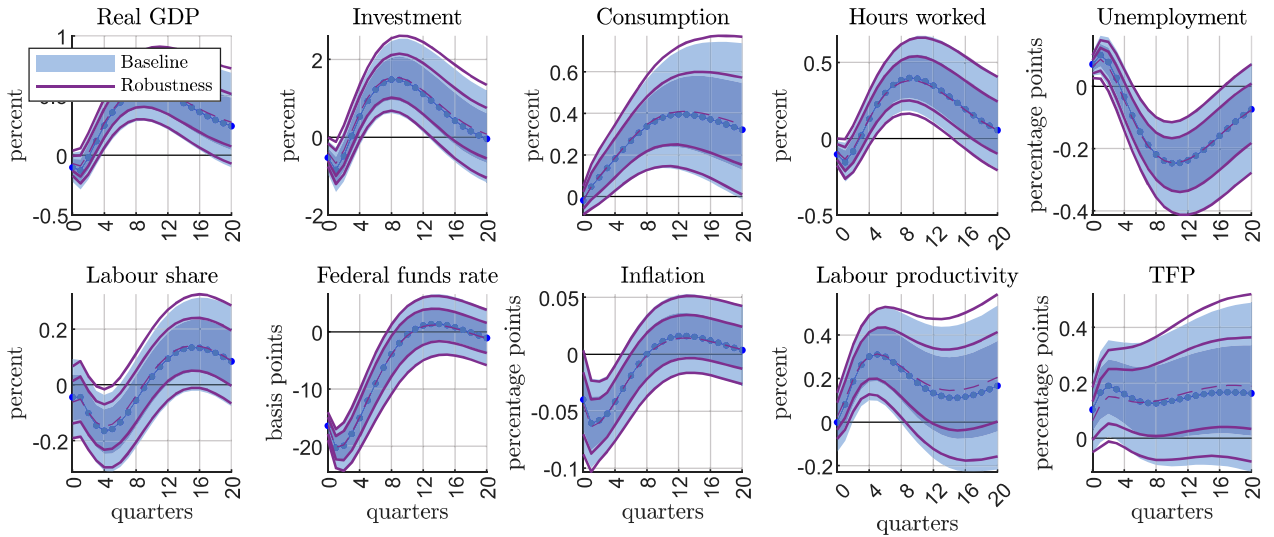
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure G-18: Robustness for flat prior on v : impulse responses

A) Demand shock



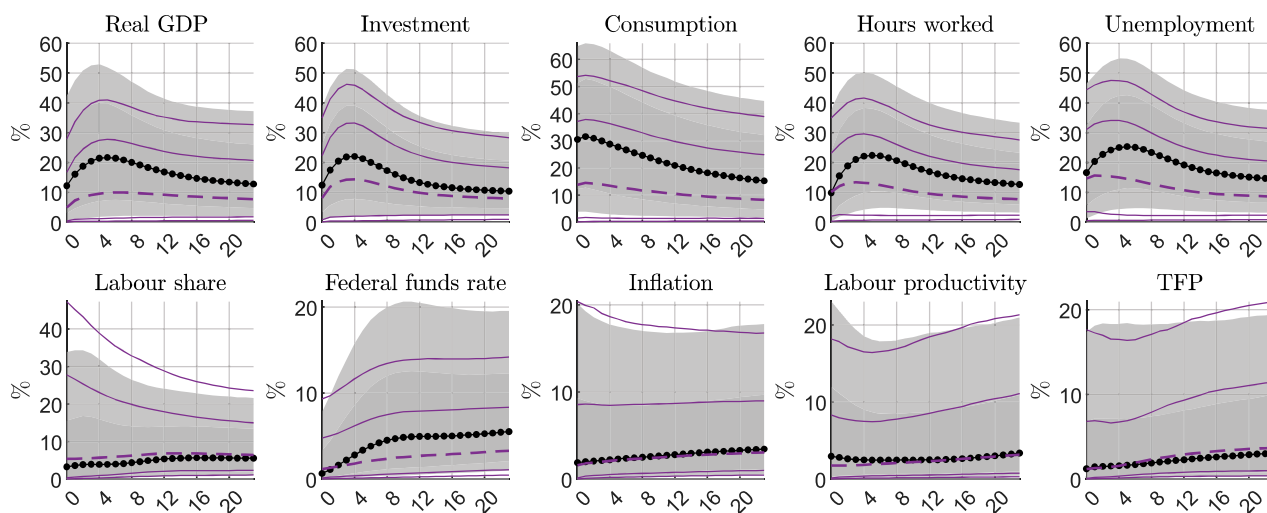
B) supply shock



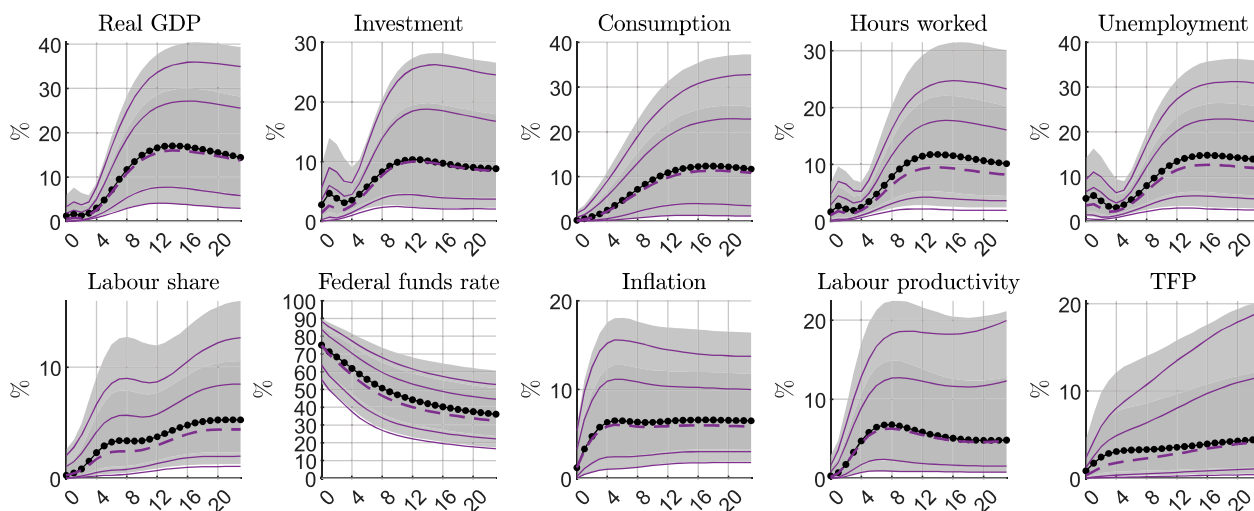
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure G-19: Robustness for flat prior on ν : forecast error variance decomposition

A) Demand shock



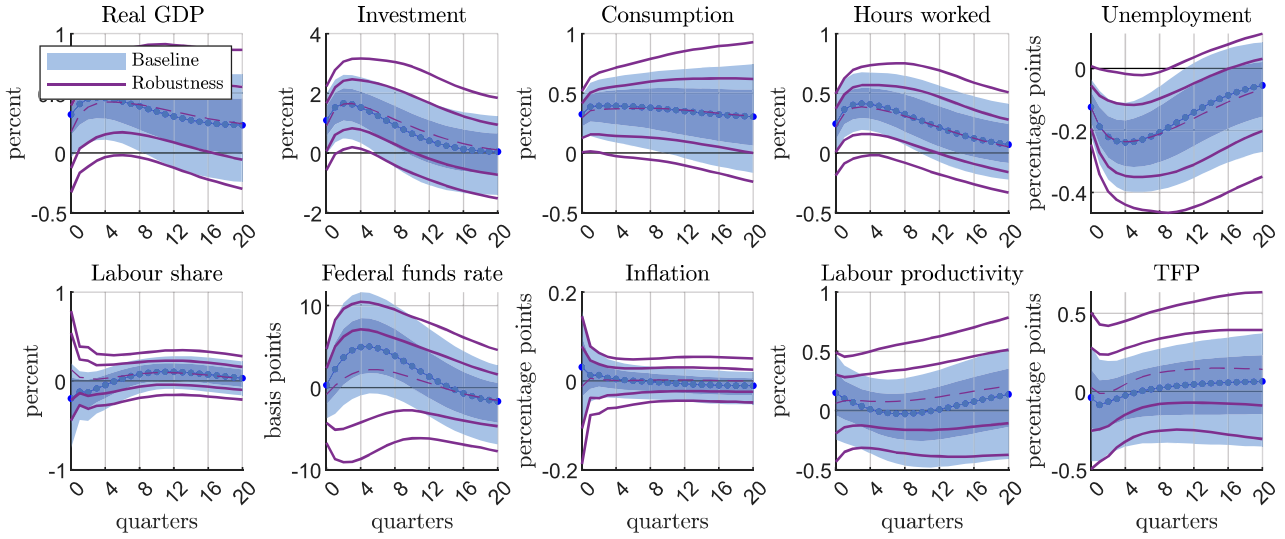
B) supply shock



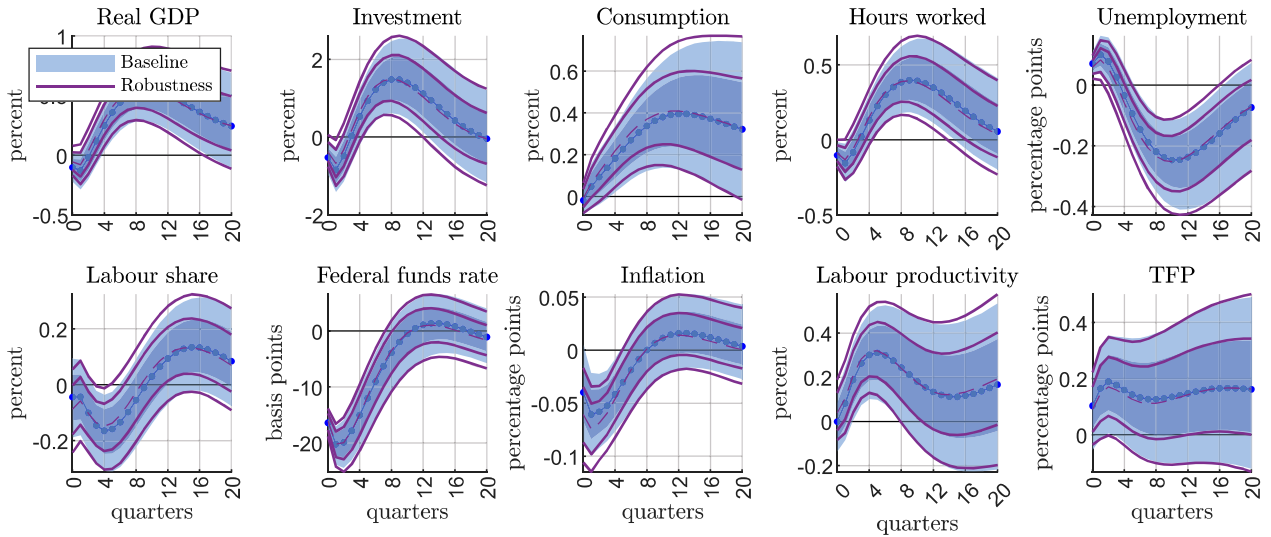
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure G-20: Robustness for looser prior on ϕ : impulse responses

A) Demand shock



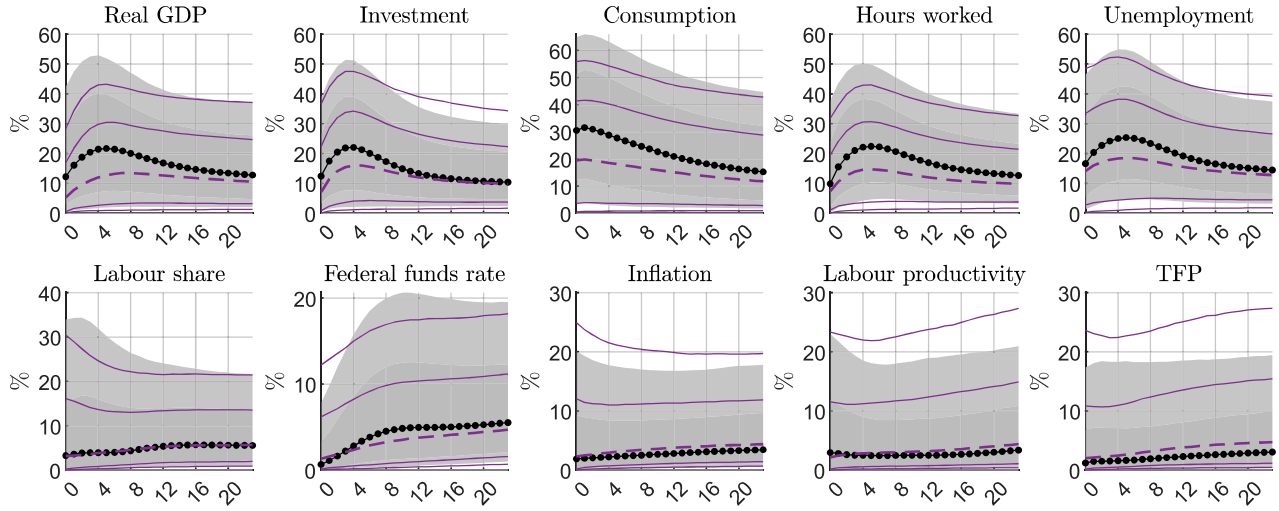
B) supply shock



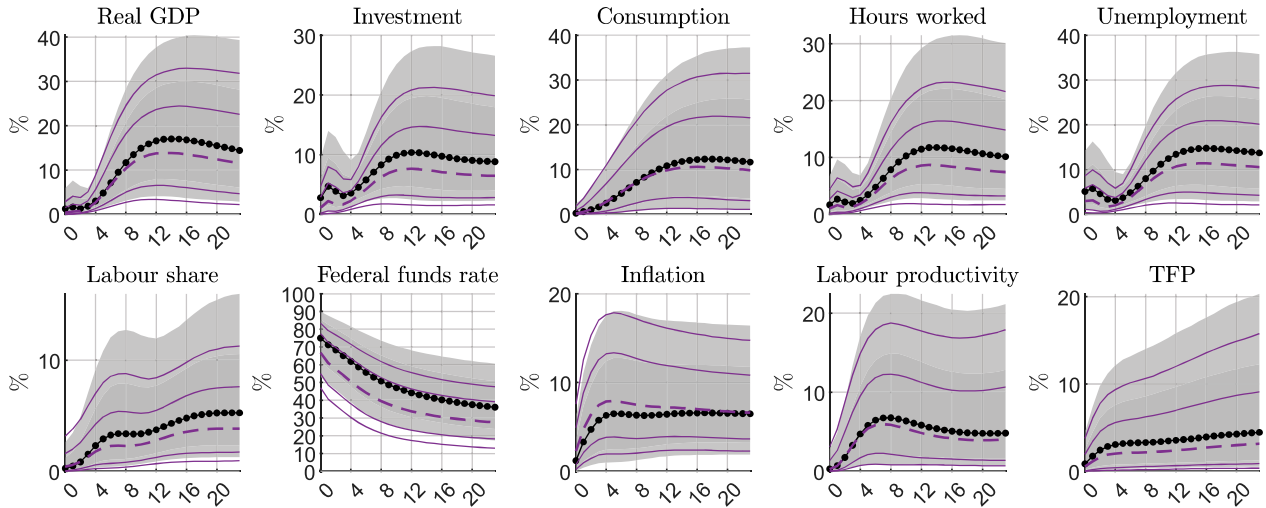
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure G-21: Robustness for looser prior on ϕ : forecast error variance decomposition

A) Demand shock



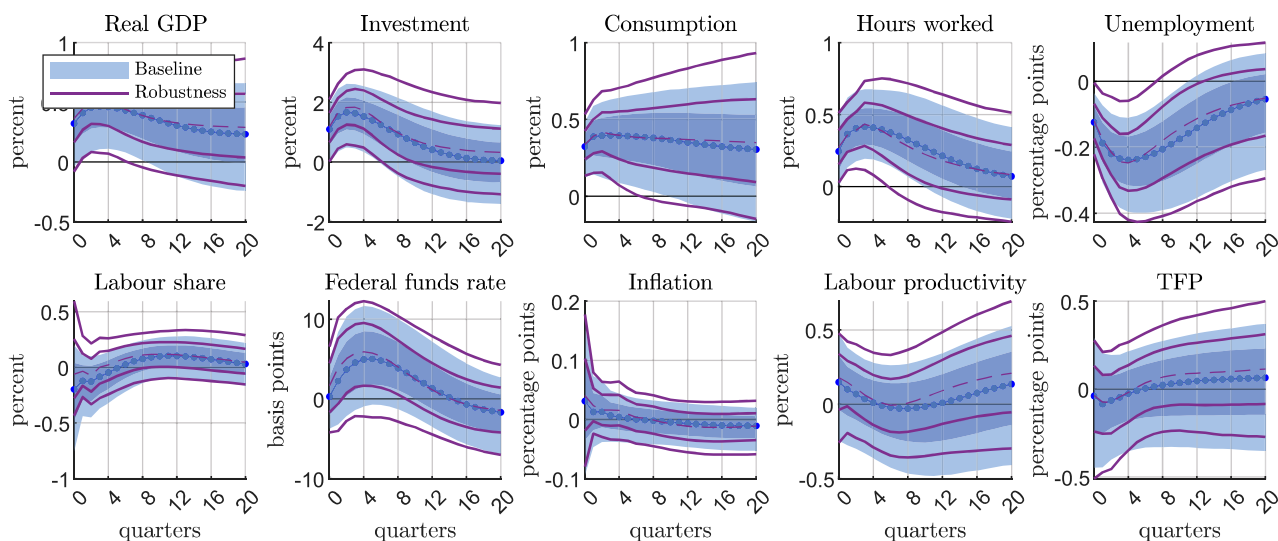
B) supply shock



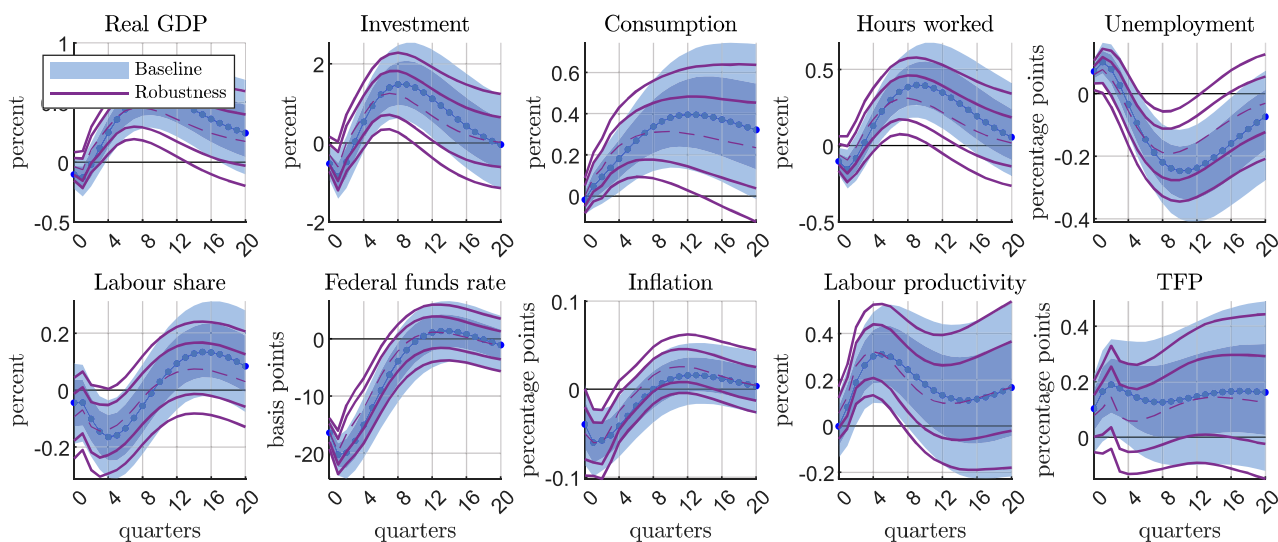
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The dashed and solid lines show the pointwise median and credible sets from the alternative specification.

Figure G-22: Robustness for $p = 4$: impulse responses

A) Demand shock



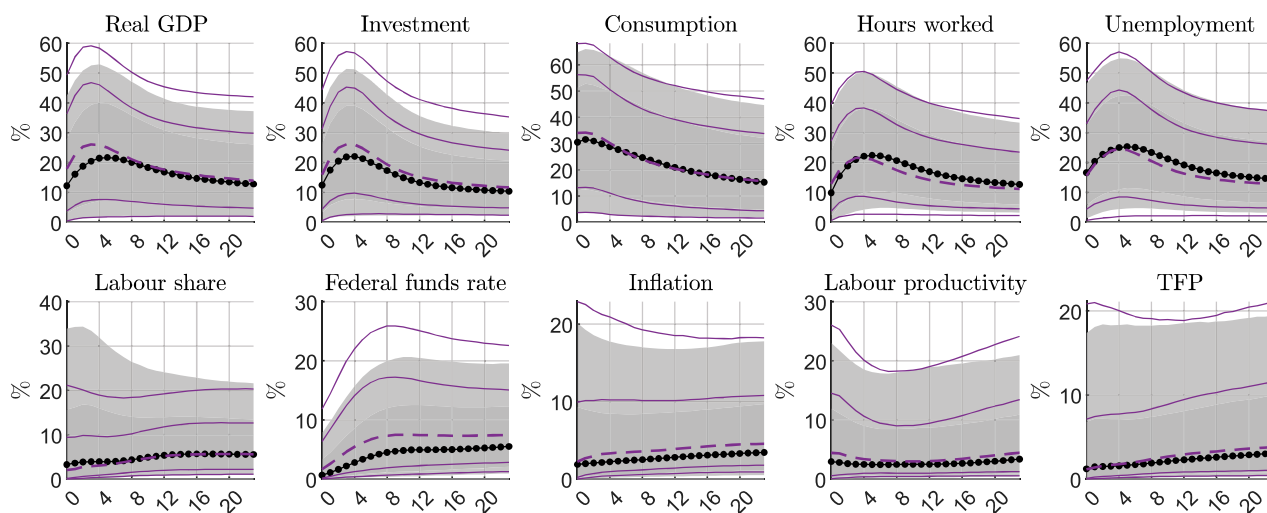
B) supply shock



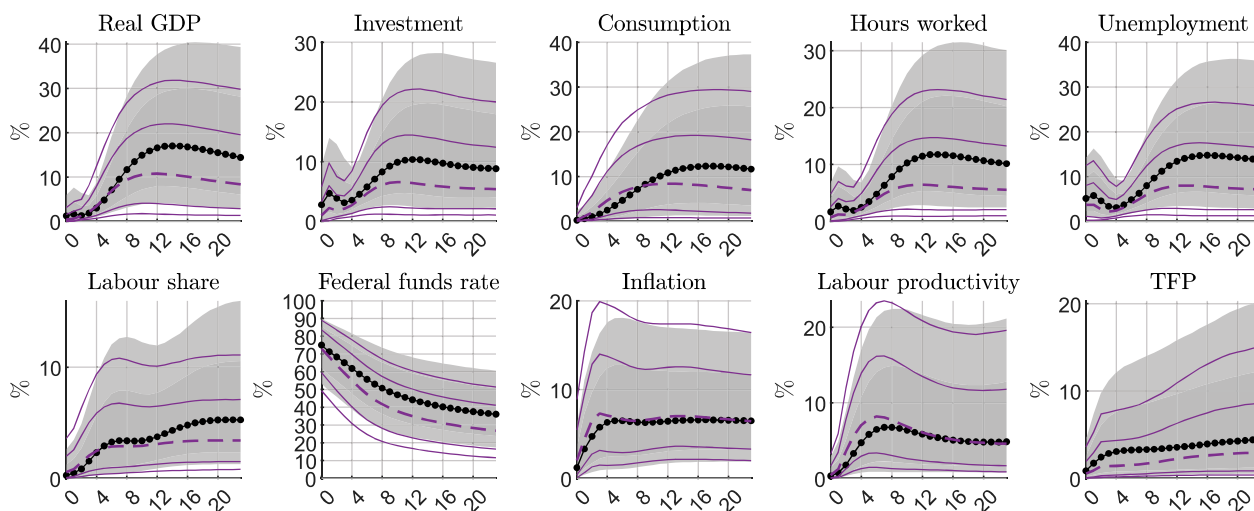
Note: The dotted line and shaded areas show the pointwise median and credible sets from the baseline specification. The solid lines show the pointwise credible sets from the alternative specification.

Figure G-23: 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.