Flexible prior beliefs on impulse responses in Bayesian vector autoregressive models*

Fabio Canova† Andrzej Kocięcki ‡ Michele Piffer§

November 23, 2025

Abstract

We design a prior for VAR coefficients that allows for flexible, non-dogmatic beliefs on the shape and the timing of the structural impulse responses. We achieve this with a particular setting of the moments of a Normal distribution. Posterior computations are no more demanding than with existing specifications; yet, the methodology provides shrinkage on impulse responses. We study the transmission of monetary policy shocks. Replacing standard priors with a prior that assumes that monetary policy shocks generate temporary but persistent effects leads to a hump-shaped posterior response of industrial production. The trough occurs eight months after the shock.

JEL classification: C32, E52.

Keywords: Non-dogmatic beliefs, impulse responses, structural shocks, identification, monetary policy.

---

*We thank Marta Bańbura, Luca Gambetti, Paolo Gelain, Michele Lenza, Christian Matthes, Giovanni Ricco, Giuseppe Ragusa, Raffaella Giacomini, Giorgio Primiceri, Domenico Giannone, Barbara Rossi, Herman van Dijk and Alessio Volpicella for helpful comments and suggestions. We also thank the participants of several seminars and conferences for comments and suggestions.

†BI Norwegian Business School, Oslo, Norway. e-mail: fabio.canova@bi.no

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

§Bank of England, UK, and King's Business School, King's College London, UK. e-mail: m.b.piffer@gmail.com

1 Introduction

Impulse response functions (IRFs) are one of the most popular tools in modern macroeconomics and have proved to be essential in exploring the dynamics induced by structural disturbances. Applications of impulse response analysis include, among many others, studying how the economy responds to policy interventions (Caldara and Kamps, 2017, Miranda-Agrippino and Ricco, 2021), to financial disruptions (Gilchrist and Zakrajšek, 2012), and to geopolitical and uncertainty-related risks (Piffer and Podstawski, 2018, Caldara and Iacoviello, 2022).

Researchers typically have strong views on what dynamic responses should be considered reasonable. Beliefs could be held on the timing, on the persistence, or on the shape of the responses. For example, it is generally believed that a contractionary monetary policy shock should persistently decrease output, although it is not clear when the largest effect occurs. The long standing debates about the 'price', the 'liquidity', and the 'exchange rate' puzzles provide other leading examples of situations where researchers have strong views about the likely path of certain variables in responses to structural shocks (Ramey, 2016, Gourinchas and Tornell, 2004).

Unfortunately, macroeconomists are severely constrained in their ability to support the estimation of impulse response functions with meaningful prior beliefs on their timing, their persistence, or their shape. The computational convenience of working with Vector Autoregressive (VAR) models has been widely acknowledged in the literature (Kilian and Lütkepohl, 2017). Yet, the existing practice of specifying a flat or Minnesota-like prior for VAR parameters does not allow a researcher to introduce, even indirectly, meaningful constraints on features of the estimated impulse response functions. In fact, prior beliefs about the shape and the timing of the responses can only be imposed dogmatically, by rejecting all posterior draws failing to imply the required pattern of responses (Canova and Pappa, 2011).

Beliefs on the shape of the impulse responses can be naturally introduced directly estimating Moving Average (MA) models (Plagborg-Møller, 2019). However, MA models are computationally demanding and may require approximation techniques to obtain parameter estimates (Barnichon and Matthes, 2018). An alternative is to work with Local Projections and impose priors directly on the projection coefficients. An unresolved challenge in working with a Bayesian Local Projections is the specification of a meaningful covariance structure for the residuals (Ferreira et al., 2023).

The contribution of the paper. This paper develops a prior for the VAR coefficient which can be used to sharpen inference via flexible, non-dogmatic beliefs on the timing, the persistence, or the shape of the impulse responses. We remain agnostic about the prior for the contemporaneous effect of the shocks, and allow for a wide range of options currently available in the literature. Rather than expressing a prior directly on the remaining response horizons, we work with the popular Normal prior for the reduced form VAR parameters. Our prior specifies the first moment of the Normal prior that differs from those of the Minnesota prior and achieves two goals. First, the first moments are selected so that the implied distribution of the IRFs is approximately centered around the path a researcher wants a-priori to introduce. Second, because the prior has a Normal format, it retains the computational convenience of using highly tractable posterior sampling. Hence, one can introduce non-dogmatic beliefs about features of the IRFs by simply replacing the Minnesota-like prior with our proposed specification. Our prior nests the flat and Minnesota-like priors, as well as the long run prior of Giannone et al. (2019) as special cases. Since it is imposed on the VAR coefficients, it vanishes asymptotically as the sample size increases.

Importantly, our contribution is not intended to provide a new approach to achieve identification of the structural shocks. Instead, we wish to provide shrinkage via prior beliefs on the dynamics produced by impulse responses.

There are four main advantages of our specification. It has been acknowledged in the Bayesian literature that shrinkage is more naturally introduced on endogenous functions of the parameters of the model (see, for instance, Van Dijk and Kloek, 1980, Harvey et al., 2007). In the SVAR literature, priors of this type are currently viable only when considering the unconditional properties of the observables (see Villani, 2009 and Jarociński and Marcet, 2019). The first advantage of our method is to allow for shrinkage of dynamic responses, which are undoubtedly a key function of SVAR parameters (Kilian, 2022). A second advantage is that, contrary to existing approaches, by suggesting the data to produce particular IRF shapes, it can sharpen inference without requiring additional identifying restrictions (Kilian and Murphy, 2012, Amir-Ahmadi and Drautzburg, 2021). The third advantage is that it remains computationally tractable, since it exploits the conditionally conjugate nature of the Normal prior for reduced form VAR parameters. The fourth advantage is that the specification is flexible and does not require formulating believes simultaneously on the features of *all* impulse responses. For example, one can select tighter beliefs for some responses and looser beliefs for other responses.

We illustrate the properties of our prior specification using data simulated from a conventional three-variable New Keynesian model. The model features a very persistent response of the output gap to a government spending shock and a hump-shaped response of the interest rate to a TFP shock. As expected, in large samples the prior is irrelevant and all the specifications we consider lead to the same responses. However, in a sample of a realistic size, a flat prior leads to very wide posterior credible sets, while a random walk or a white noise Minnesota prior lead to posterior IRFs where all variables in response to all shocks which are either persistent or not persistent. By contrast, our prior makes it possible to introduce the belief that the effect is relatively persistent for a subset of the shocks and the variables, while allowing for sizable uncertainty around the responses. As a result, posterior IRFs broadly mimic the half-life of the true responses, regardless of their persistence or shape.

We study the classic question of how output responds to a monetary policy surprises. The issue of at what horizon the largest negative response occurs has received considerable attention (Christiano et al., 1999, Uhlig, 2005, Antolín-Díaz and Rubio-Ramírez, 2018), but both the shape and the timing of the output responses remain uncertain. We take a five variable VAR model inspired by Jarociński and Karadi (2020) and identify policy disturbances using an instrumental variables (IV) approach.

A flat prior and two versions of the Minnesota priors produce output responses with no particular pattern. In the former case no hump is generated and, in the latter case, the strongest effect occurs on impact. We then introduce the belief that monetary shocks generate persistent mean output effects, a belief which is in line with a wide class of current New Keynesian models. We find that the posterior distribution of output responses displays hump-shaped dynamics, and the through is reached 8 months after the shock, regardless of how one treats the Covid-19 event. Finally, a one standard deviation shock that increases the federal funds rate on impact leads to a maximum output decrease of around 3%, which is in line with what Jarociński and Karadi (2020) obtain. Thus, our prior supports the widely-held view that a central bank is capable to affect real economic activity but it does so with long and, possibly, variable lags.

The relationship with the literature. There is a considerable body of literature dealing with Bayesian VARs, see Koop and Korobilis (2010) and Miranda-Agrippino and Ricco (2019) for a detailed discussion. Relative to Baumeister and Hamilton (2015, 2024), we design priors for VAR coefficients rather than for structural elasticities; and relative to Baumeister and Hamilton (2018), we focus attention on dynamic responses rather than the impact effects. The paper is also related to Barnichon and Matthes (2018), who work with functional approximations of IRFs for MA models, and to an earlier contribution of Kocięcki (2010), who works with recursive identification and a joint Normal prior distributions for the impulse responses. Our approach has the same flavor as Villani (2009) and Andrle and Benes (2013), who design priors for endogenous objects of a model. The need to design tools that explicitly introduce non-dogmatic beliefs on impulse responses was acknowledged early on by Gordon and Boccanfuso (2001) and Dwyer (1998), who nevertheless did not deliver usable priors.

Our approach is related to the one of Plagborg-Møller (2019), who obtains impulse responses directly estimating MA models. Because he works with the structural MA representation, his method is more general than ours with respect to the prior beliefs that can be designed. However, the generality comes with costs. First, the estimation of Structural Vector Moving Average models is complicated. Although he proposed an efficient Bayesian method, relying on Whittle likelihood approximation and a state-of-the-art variant of the Hamiltonian Monte Carlo principle, it is not obvious the approach can be scaled up to large models without numerical issues. Our specification can be easily employed in large SVARs since it exploits the properties of conjugate priors. Scalability is important since prior shrinkage matters primarily in larger models. Second, since impulse responses are more frequently estimated from SVAR models, his approach does not allow a direct comparison of the information content of the prior relative to existing specifications. In contrast, our prior allows this comparison as it nests popular choices used in the literature. Finally, the Plagborg-Møller (2019) approach needs a specification of how the matrix of impact coefficients is identified. Our method is consistent with a number of existing identification choices. In sum, since the two approaches take a different point of view in the flexibility-computational complexity frontier, they should be considered complementary rather than substitute.

Nowadays, most of the current Bayesian SVAR literature dealing with sign identification restrictions discusses the pros and cons of following a two-step approach to the estimation and the identification of the VAR (Baumeister and Hamilton, 2015), whether the failure to update contemporaneous prior beliefs is a problem or not (Inoue and Kilian, 2020), and whether contemporaneous beliefs implied by a standard two-step algorithm are informative or not (Arias et al., 2024). Our paper is not designed to deal with these issues and it does not take a stand in the discussion. In fact, it is consistent with both the methodology of Arias et al. (2018) and of Baumeister and

Hamilton (2015), and with any point identification approach. Finally, while we do not follow the method of Giacomini and Kitagawa (2021), such an approach can also be used in conjunction with our prior to robustify response inference.

The outline of the paper. The rest of the paper is organized as follows. section 2 discusses in details the prior we propose. section 3 compares the properties of our specification using data simulated from a standard DSGE model. section 4 studies how monetary policy disturbances are transmitted to the real economy. section 5 concludes. The Online Appendices contain the derivations of expressions presented in the paper, the computational details, and additional material and figures mentioned in the text.

2 The empirical methodology

This section explains how our prior for VAR coefficients is derived. We show how our selection can be combined with commonly employed identification strategies to produce posterior structural impulse responses.

2.1 The model

We write a Structural Vector Autoregressive (SVAR) model as

$$\boldsymbol{y}_t = \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{c} + B\boldsymbol{\epsilon}_t, \tag{1a}$$

$$\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, I_k), \tag{1b}$$

where $\boldsymbol{y}_t$ is a $k \times 1$ vector of observables, Π_l is a $k \times k$ matrix of autoregressive coefficients at horizon $l = 1, .., p$, $\boldsymbol{c}$ is a $k \times 1$ vector of constants, and B is a $k \times k$ non-singular matrix. The vector $\boldsymbol{\epsilon}_t$ contains the serially independent structural shocks, whose covariance matrix is normalized to the identity matrix. The model can also be written in other ways. A useful alternative is given by:

$$\boldsymbol{y}_t = \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{c} + \boldsymbol{u}_t, \tag{2a}$$

$$\boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma), \tag{2b}$$

$$\boldsymbol{u}_t = B\boldsymbol{\epsilon}_t, \tag{2c}$$

$$\Sigma = BB', \tag{2d}$$

$$B = \chi(\Sigma)Q, \tag{2e}$$

where $\chi(.)$ is a function capturing any square root factorization of Σ, and Q an orthonormal matrix. Equations (1a)-(1b) and (2a)-(2e) define the same SVAR, while equation (2e) highlights the correspondence between reduced form and structural representations. The mapping between model (1) and model (2) is discussed at length, for instance, in Arias et al. (2018). Below, we use notation:

$$\Pi = [\Pi_1, .., \Pi_p], \tag{3}$$

$$\boldsymbol{\pi} = \text{vec}(\Pi), \tag{4}$$

$$\tilde{\boldsymbol{\pi}} = (\boldsymbol{\pi}', \boldsymbol{c}')'. \tag{5}$$

For the rest of this section we assume, without loss of generality, that the data is demeaned so that $\boldsymbol{c} = \boldsymbol{0}$.

Let Ψ_h denote the impulse response function (IRF) h periods after the shocks, and let M be the maximum horizon of interest for which responses are computed. Ψ_h is of dimensions $k \times k$, with entry $i, j = 1, .., k$ capturing how variable i responds to shock j, h periods after the shock. It is well known that the mapping between SVAR objects $(B, \Pi_1, .., \Pi_p)$ and IRF objects $(\Psi_0, \Psi_1, .., \Psi_M)$ can be obtained recursively, and for

$M \geq p$ it is given by (Kilian and Lütkepohl, 2017):

$$\Psi_0 = B, \tag{6a}$$
$$\Psi_1 = \Pi_1 \Psi_0, \tag{6b}$$
$$\Psi_2 = \Pi_1 \Psi_1 + \Pi_2 \Psi_0, \tag{6c}$$
$$\Psi_3 = \Pi_1 \Psi_2 + \Pi_2 \Psi_1 + \Pi_3 \Psi_0, \tag{6d}$$
$$\dots$$
$$\Psi_p = \Pi_1 \Psi_{p-1} + \Pi_2 \Psi_{p-2} + \cdots + \Pi_p \Psi_0, \tag{6e}$$
$$\Psi_{p+1} = \Pi_1 \Psi_p + \Pi_2 \Psi_{p-1} + \cdots + \Pi_p \Psi_1, \tag{6f}$$
$$\dots$$
$$\Psi_M = \Pi_1 \Psi_{M-1} + \Pi_2 \Psi_{M-2} + \cdots + \Pi_p \Psi_{M-p}. \tag{6g}$$

If B is non-singular, (6) provides a one-to-one mapping between the SVAR parameters and the IRF parameters for any $M \geq p$. Thus, any prior beliefs on the SVAR coefficients imply prior beliefs on the IRFs elements via the system of equations (6).

2.2 Our approach

To provide intuition, we first illustrate the features of our approach in a simplified setting. We then generalize the derivation and relate it to the existing literature.

2.2.1 Illustration in a restricted environment

For the time being, suppose that:

a) The shocks are identified recursively;

b) An inverse-Wishart prior is used for Σ;

c) B is set equal to the Cholesky decomposition of Σ;

d) The prior for $\tilde{\boldsymbol{\pi}}$ is Normal independent of Σ,

$$\tilde{\boldsymbol{\pi}} \sim N(\boldsymbol{\mu}, V). \tag{7}$$

As is well known, in this case the joint posterior $p(\tilde{\boldsymbol{\pi}}, \Sigma | Y)$ can be conveniently explored with a standard Gibbs sampler (Koop and Korobilis, 2010).

We wish to know whether one can select the hyperparameters $(\boldsymbol{\mu}, V)$ so as to grant a researcher some flexibility over the implied prior for the IRFs, given that, conditional on B, $p(\tilde{\boldsymbol{\pi}})$ implies a prior for the IRFs via the system (6).

It is standard in the literature to set $(\boldsymbol{\mu}, V)$ according to:

$$E((\Pi_h)_{ij}) = \begin{cases} \delta_i, & j = i, h = 1 \\ 0 & \text{otherwise} \end{cases}, \qquad V((\Pi_h)_{ij}) = \begin{cases} \frac{\lambda}{h^2}, & j = i \\ \eta \frac{\lambda}{h^2} \frac{\sigma_i}{\sigma_j} & \text{otherwise} \end{cases}, \tag{8}$$

which is typically referred to as Minnesota prior. Note that a flat prior is obtained by letting λ to be large, while the random walk and the white noise specifications can be obtained choosing δ_i to be 1 or 0, respectively (see Canova, 2007, Bańbura et al., 2010 and Koop and Korobilis, 2010 for popular selections of the remaining hyperparameters η, σ_i, σ_j) [1] While the forecasting properties of a VAR endowed with such prior restrictions are well-documented, the prior allows for no flexibility in designing IRFs shapes. One could constrain the implied IRFs by adding to the algorithm an accept/reject step to ensure that the posterior draws do produce the required shapes (for example, that the response of economic activity to a monetary shock is larger in absolute value at horizon two than at horizon one). However, such a way of proceeding introduces restrictions *dogmatically*, which might not be in the intention of the researcher. In addition, the computations may turn out to be inefficient when most of the posterior draws fail to satisfy the candidate restrictions.

We view our beliefs as a tool to implement posterior Bayesian shrinkage on impulse responses. Thus, we endogenously select $\boldsymbol{\mu}$ so that the mean of the IRFs has certain a-priori features. Let $\bar{\Psi} = (\bar{\Psi}_0, .., \bar{\Psi}_H)$ be an array reflecting the researcher's prior mean for the impulse responses, where H is the maximum horizon up to where beliefs are formulated. Generally, $H \leq M$, as one need not have prior beliefs stretching as far as the horizon of interest. $\bar{\Psi}$ is a high dimensional object, and includes $k^2(H+1)$ entries. Assume that $H = p$ (the number of lags of the SVAR), and that $\bar{\Psi}_0$ is equal to the expected value of B implied by the prior on Σ and the identification approach used (here, a lower triangular matrix with positive diagonal entries). Set the remaining entries $(\bar{\Psi}_1, .., \bar{\Psi}_H)$ to capture prior beliefs about the dynamics of the IRFs to the structural shocks. Then, substitute $\bar{\Psi}$ into (6), eliminate the first equation and invert

[1] We will refer to the Minnesota or Minnesota-like prior only with reference to the prior for $\boldsymbol{\pi}$, while remaining intentionally silent about the priors on B or Σ.

the next p equations of system (6) to obtain:

$$\bar{\Pi}_1 = \bar{\Psi}_1 \bar{\Psi}_0^{-1}, \tag{9a}$$
$$\bar{\Pi}_2 = [\bar{\Psi}_2 - \bar{\Pi}_1 \bar{\Psi}_1] \bar{\Psi}_0^{-1}, \tag{9b}$$
$$\bar{\Pi}_3 = [\bar{\Psi}_3 - \bar{\Pi}_1 \bar{\Psi}_2 - \bar{\Pi}_2 \bar{\Psi}_1] \bar{\Psi}_0^{-1}, \tag{9c}$$
$$\dots$$
$$\bar{\Pi}_p = [\bar{\Psi}_p - \bar{\Pi}_1 \bar{\Psi}_{p-1} - \bar{\Pi}_2 \bar{\Psi}_{p-2} - \cdots - \bar{\Pi}_{p-1} \bar{\Psi}_1] \bar{\Psi}_0^{-1}. \tag{9d}$$

$\bar{\Pi} = [\bar{\Pi}_1, .., \bar{\Pi}_p]$ are the values of the VAR coefficients associated with the selected $\bar{\Psi}$.

In this simplified setting, replacing the specification of $\boldsymbol{\mu}$ in (8) with

$$E((\Pi_h)_{ij}) = (\bar{\Pi}_h)_{ij}, \qquad h = 1, .., p, \tag{10}$$

is enough to gain control over the implied prior for impulse responses. In fact, the prior $p(\tilde{\boldsymbol{\pi}}, \Sigma)$, the Cholesky identification of B and the equality $E(B) = \bar{\Psi}_0$ jointly imply, via system (6), that $p(\Psi_0, \Psi_1, ...)$ satisfies

$$E(\Psi_h) = \bar{\Psi}_h, \qquad h = 0, 1, \tag{11a}$$
$$\lim_{V \to 0} E(\Psi_h) = \bar{\Psi}_h, \qquad h = 2, .., H, \tag{11b}$$

where (11a) holds for $h = 0$ by assumption, while the remaining conclusions are derived in the Online Appendix.

Since $\bar{\Psi}$ is a researcher choice, replacing the Minnesota selection of $\boldsymbol{\mu}$ with an alternative selection which depends on $\bar{\Psi}$, gives the researcher leverage over prior expectation of the impulse responses. Note also that, since our prior on the VAR coefficients is Normal, standard posterior samplers can be used, even when $p(\Psi_0, \Psi_1, ...)$ is not Normal.

Note also that equation (11b) is exactly satisfied in the limit. However, the exercises we conduct in subsection 2.4, section 3 and section 4 document that the approximation error $E(\Psi_h) - \bar{\Psi}_h$ is negligible for the values of V used in the literature, for instance those of Bańbura et al. (2010).

One could consider adding correction terms that would eliminate the discrepancy between $\bar{\Psi}_h$, and implied IRFs prior means, $E(\Psi_h)$. Unfortunately, deriving these terms turns out to be a daunting task. Appendix B in the Online Appendix contains the derivations of the implied IRFs prior means up to $h = 5$ and shows that the underlying complexity is such that it prevents us from deriving moments for larger horizons and propose a generic bias-correction procedure. An alternative option would be to find bounds for the discrepancy using inequalities for functions of random variables (e.g. Jensen's type inequalities). Such an option, however, is also unfeasible. To understand why consider the univariate random variable X. Then $E(X^2) \geq [E(X)]^2$, but, in general, nothing is known about the relationship between, say, $E(X^3)$ and $[E(X)]^3$, unless X is positive or negative, which is not the case here. Furthermore, the formula for $E(\Psi_h)$, even for moderate h, involves large number of mixed moments (polynomials in multiple variables) of large order which display little regularity and the complexity of the computation increases with h (see the Online Appendix for a constructive demonstration).

Because no additional formal results can be explicitly derived, and although our examples show that little biases are present, one should routinely check the size of the approximation error before using our prior.

2.2.2 Generalization for $H > p$

The discussion in subsubsection 2.2.1 focuses on the case in which the researcher formulates IRFs beliefs for up to p horizons, where p is the number of VAR lags. We will refer to this as *Case a*. It is useful to extend the approach to the case in which beliefs are expressed up to horizon $H > p$. Since p is typically a researcher choice, one could, in principle, work with *Case a* and increase p up to the intended horizon H if it is initially selected to be smaller. However, in practice, larger values of p make the model dimensionality an issue. Thus, we prefer to derive the mapping for $H > p$.

Define $\bar{\Pi}$ like in subsubsection 2.2.1 as the value of the SVAR coefficients associated with the first $p+1$ blocks of $\bar{\Psi}$. It can happen, in principle, that when substituting $\bar{\Psi} = (\bar{\Psi}_0, .., \bar{\Psi}_H)$ and $\bar{\Pi} = [\bar{\Pi}_1, .., \bar{\Pi}_p]$ into system (6), all equalities hold despite the fact that $H > p$. In this case $\bar{\Psi} = (\bar{\Psi}_0, .., \bar{\Psi}_H)$ are functionally constrained in a way that a SVAR model with $p < H$ lags still generates $\bar{\Psi}$ via equation (6). We will refer to this as *Case b*.

When $H > p$ and $\bar{\Psi}$ are not functionally constrained, which we refer to as *Case c*, one could in principle still select $\boldsymbol{\mu}$ as in (10), effectively using only the first $p+1$ entries of $\bar{\Psi}$. This may produce smaller approximation errors $E(\Psi_h) - \bar{\Psi}_h$ at shorter horizons at the cost of introducing larger errors at longer horizons and introduce inefficiencies. Thus, we consider all horizons H and select $\boldsymbol{\mu}$ as discussed next.

Define $\boldsymbol{b} = \text{vec}(B)$, $\Psi = [\Psi_0, \Psi_F]$, $\Psi_F = [\Psi_1, .., \Psi_H]$, $\boldsymbol{\psi} = \text{vec}(\Psi)$, $\boldsymbol{\psi}_F = \text{vec}(\Psi_F)$ and $\boldsymbol{\psi}_h = \text{vec}(\Psi_h)$, for $h = 0, 1, .., H$. Vectorizing the first $H+1$ equations of (6) one obtains:

$$\boldsymbol{\psi}_0 = \boldsymbol{b}, \tag{12}$$

$$\boldsymbol{\psi}_F = R\boldsymbol{\pi}, \tag{13}$$

where

$$R = R_H \otimes I_k \equiv \underbrace{\begin{bmatrix} \Psi_0' & 0 & 0 & \dots & 0 \\ \Psi_1' & \Psi_0' & 0 & .. & 0 \\ \Psi_2' & \Psi_1' & \Psi_0' & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_{p-1}' & \Psi_{p-2}' & \Psi_{p-3}' & \dots & \Psi_0' \\ \Psi_p' & \Psi_{p-1}' & \Psi_{p-2}' & \dots & \Psi_1' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_{H-1}' & \Psi_{H-2}' & \Psi_{H-3}' & \dots & \Psi_{H-p}' \end{bmatrix}}_{R_H} \otimes I_k. \tag{14}$$

R_H is a function of Ψ, it is of dimension $Hk \times pk$, and is of full column rank as long as B is non-singular. Let $\bar{\boldsymbol{\psi}}_h = \text{vec}(\bar{\Psi}_h)$, $\bar{\boldsymbol{\psi}} = (\bar{\boldsymbol{\psi}}_0', \bar{\boldsymbol{\psi}}_F')' = \text{vec}(\bar{\Psi})$, $\bar{\Psi} = [\bar{\Psi}_0, \bar{\Psi}_F]$, $\bar{\Psi}_F = [\bar{\Psi}_1, .., \bar{\Psi}_H]$. Define the artificial random variables W_h, $h = 1, .., H$:

$$W_1 = \Pi_1 \bar{\Psi}_0 - \bar{\Psi}_1, \tag{15a}$$

$$W_2 = \Pi_1 \bar{\Psi}_1 + \Pi_2 \bar{\Psi}_0 - \bar{\Psi}_2, \tag{15b}$$

$$W_3 = \Pi_1 \bar{\Psi}_2 + \Pi_2 \bar{\Psi}_1 + \Pi_3 \bar{\Psi}_0 - \bar{\Psi}_3, \tag{15c}$$

$$\vdots$$

$$W_p = \Pi_1 \bar{\Psi}_{p-1} + \Pi_2 \bar{\Psi}_{p-2} + \dots + \Pi_p \bar{\Psi}_0 - \bar{\Psi}_p, \tag{15d}$$

$$W_{p+1} = \Pi_1 \bar{\Psi}_p + \Pi_2 \bar{\Psi}_{p-1} + \dots + \Pi_p \bar{\Psi}_1 - \bar{\Psi}_{p+1}, \tag{15e}$$

$$\vdots$$

$$W_H = \Pi_1 \bar{\Psi}_{H-1} + \Pi_2 \bar{\Psi}_{H-2} + \dots + \Pi_p \bar{\Psi}_{H-p} - \bar{\Psi}_H. \tag{15f}$$

The system of equations (15) is closely related to system (6) except that it drops the equation at horizon 0, replaces the matrices Ψ_h with the matrices of hyperparameters

$\bar{\Psi}_h$, and drops the last $M - H$ equations. The system (15) can be vectorized as

$$\boldsymbol{w} = \bar{R}\boldsymbol{\pi} - \bar{\boldsymbol{\psi}}_F, \tag{16}$$

where $\boldsymbol{w} = \text{vec}([W_1, .., W_H])$ and $\bar{R}$ is defined in equation (14) after replacing Ψ with $\bar{\Psi}$. If $\bar{\Psi}_0$ is non-singular, the matrix $\bar{R}$ has full column rank. Premultiplying both sides of (16) by $\bar{R}'$ and rearranging the terms gives:

$$\boldsymbol{\pi} = \left(\bar{R}'\bar{R}\right)^{-1}\bar{R}'\boldsymbol{w} + \left(\bar{R}'\bar{R}\right)^{-1}\bar{R}'\bar{\boldsymbol{\psi}}_F. \tag{17}$$

Note that in the system of equations (16) $\boldsymbol{\pi}$ is of lower dimension than $\boldsymbol{w}$ for $H > p$. In other words, given $(\bar{R}, \bar{\boldsymbol{\psi}}_F)$, whether a solution for $\boldsymbol{\pi}$ exists depends on $\boldsymbol{w}$. In this case, one can select $\boldsymbol{\mu}$ from (17), setting

$$\boldsymbol{\mu} = \left(\bar{R}'\bar{R}\right)^{-1}\bar{R}'\bar{\boldsymbol{\psi}}_F. \tag{18}$$

To summarize, the three cases of interest are:

1) When $H = p$ (*Case a*), a SVAR with p lags replicates exactly the pattern of $\bar{\Psi}$, making the system (16) consistent for $\boldsymbol{w} = \mathbf{0}$. This means that (18) is the unique solution at $\boldsymbol{w} = \mathbf{0}$, $\bar{R}$ is square and invertible, and $E(\boldsymbol{\pi})$ simplifies to $\boldsymbol{\mu} = \bar{R}^{-1}\bar{\boldsymbol{\psi}}_F$, which coincides with $\text{vec}(\bar{\Pi})$ defined in equation (9).

2) When $H > p$ but $\bar{\Psi}$ is selected such that it can be replicated by a VAR with p lags (*Case b*), then the system (16) is still consistent for $\boldsymbol{w} = \mathbf{0}$, (18) is still its unique solution, and $\boldsymbol{\mu}$ from (18) still coincides with $\text{vec}(\bar{\Pi})$ defined in equation (9) despite $\bar{R}$ not being square and invertible.

3) Lastly, when $H > p$ but no parametrization of a VAR with p lags exists that replicates $\bar{\Psi}$ (*Case c*), then $\boldsymbol{w} = \mathbf{0}$ makes the system (16) inconsistent. Yet, $E(\boldsymbol{\pi})$ from (18) is still the unique solution if $\boldsymbol{w} = [\bar{R}\left(\bar{R}'\bar{R}\right)^{-1}\bar{R}' - I]\bar{\boldsymbol{\psi}}_F$.

In the Online Appendix we show that (11) holds for *Case a* and *Case b*, but not necessarily for *Case c*. Thus, in the latter case it is imperative to check, application by application, that the approximation error is small.

2.3 Discussion

Several features of the prior specification need to be emphasized.

Our prior for $\boldsymbol{\pi}$ is Normal. For this reason, posterior sampling is not in any way more challenging than with existing methodologies. If an inverse-Wishart-Uniform prior on (Σ, Q) is employed, the joint posterior can be explored with a Gibbs sampler, when we specify the prior as in (7), or direct sampling, if the prior (7) is replaced with

$$\tilde{\boldsymbol{\pi}}|\Sigma \sim N(\boldsymbol{\mu}, V_s \otimes \Sigma). \tag{19}$$

Our method also works under this latter specification. However, when an inverse-Wishart-Uniform prior for (Σ, Q) is not assumed, posterior sampling requires more involved methods to compute the marginal posterior $p(B|Y)$, for example the Metropolis-Hastings algorithm as in Baumeister and Hamilton (2015), the two-step algorithm as in Bruns and Piffer (2023), or the methods discussed in Canova and Pérez Forero (2015) and Waggoner et al. (2016).

It is worth emphasizing that, while we work with a Normal prior for the VAR coefficients, the combination of such a prior, the prior on the impact effect of the shocks and the system (6) generally implies a non-Gaussian joint distribution for Ψ. Thus our approach is consistent with the idea that, a-priori, the distribution of impulse responses may be skewed and leptokurtic. section 3 shows this is indeed the case.

We illustrated the derivation of our prior under the assumption that identification is recursively achieved. Clearly, other assumptions could be used. For our approach to be feasible we only need that $E(B) = \bar{\Psi}_0$, an assumption that is not restrictive, given that $\bar{\Psi}_0$ is, in our setup, a hyperparameter. For example, sign restrictions can be used, and one may formulate prior beliefs on $A = B^{-1}$ (as advocated by Baumeister and Hamilton, 2015, Baumeister and Hamilton, 2024), on B (as suggested by Bruns and Piffer, 2023) or on a combination of A and B (see Baumeister and Hamilton, 2018). One can also combine sign and zero restrictions (Binning, 2013, Arias et al., 2018). Identification via external instruments can also be used by adding the instrument to the list of variables in a block recursive SVAR, as in Plagborg-Møller and Wolf (2021), and the specification can be extended to the VARX setting used by Paul (2020). While formulation of a prior for the relevant instantaneous objects is an active field of research, our method does not take a stand on the issue and leaves the researcher choose her preferred approach. Note also that because our prior is formulated on reduced form parameters, it vanishes asymptotically. Thus the potential issue raised by Baumeister and Hamilton (2015) does not apply here.

It is common in the literature to study the dynamics induced only by one or a subset of the structural shocks. The prior on $\boldsymbol{\pi}$ is designed for all the responses of all shocks. The identification of a subset of shocks is hard to implement simply by changing the specification for $\boldsymbol{\pi}$. However, one can adjust the specification for $p(B)$, since each column j of B affects the entire profile of the impulse responses to shock j. Thus, if only a subset of the shocks is identified and identification is achieved using, e.g., sign restrictions on the impact effect of the shocks, then one can specify $p(B)$ to feature a wide variance for the columns associated with unidentified shocks, and a smaller variance for the columns associated with the identified shocks. In subsection 2.4 we illustrate this principle, and refer the reader to Appendix C in the Online Appendix for further details on the implementation.

In applied work researchers might want to introduce sign or shape restrictions at horizons greater than one. Computationally, this is typically costly if the joint posterior distribution has a large mass on the part of the parameter space that violates the restrictions. Our approach helps in this respect, because it can induce impulse response shapes which are in line with the intended restrictions, hence reducing the number of draws required. It can significantly reduce the computational burden to draw from the restricted posterior distributions. See also Kilian and Murphy (2012) and Amir-Ahmadi and Drautzburg (2021) for related ways of obtaining shrinkage.

A key advantage of our approach is that it does not require deriving and integrating the joint prior distribution of Ψ, a procedure that involves complex computational techniques. Our analysis only requires the expectation operator of a multivariate system of equations, see Appendix A of the Online Appendix. This is a considerable advantage relative to the contributions that work with the transformation between the SVAR and impulse responses parametrization (Kocięcki, 2010, Arias et al., 2018).

Because of the way our prior is designed, a Minnesota-like selection of $\boldsymbol{\mu}$ emerges as a special case. In fact, the random walk prior, corresponding to $\delta_i = 1$ in equation (8), can be obtained by setting $\bar{\Psi}_h = \bar{\Psi}_0$, $h = 1,..,H$, $H = p$. The white noise prior, i.e. $\delta_i = 0$, can be obtained by setting $\bar{\Psi}_h = 0$, $h = 1,..,H$, $H = p$. Thus, when V is sufficiently small, the Minnesota-like selection given in (8) is consistent with the belief that the responses to the structural shocks are either very persistent (the random walk prior) or not persistent at all (the white noise prior). In contrast, intermediate persistence cases are possible with our prior, and shape beliefs of any form can also be formulated. Note also that one may specify mean patterns for the IRFs that are consistent with both the invertibility and the non-invertibility of the VAR. Nevertheless, in the latter case, care should be exercised as the approximation error may become large.

Our method enjoys the strengths and the limitations of all Bayesian analyses. It gives a new dimension along which to introduce shrinkage, namely the impulse responses, which are key in structural analyses. It goes without saying that if the researcher introduces beliefs that are inconsistent with the data, then the posterior may be driven away from the region where the most likely IRFs are present. As we stressed, our selection of V is in line with the literature. Hence, our prior beliefs that are necessarily tighter than those in standard applications. In addition, prior sensitivity can help to assess if the results are driven by the prior.

It is common in the literature to assess the "goodness" of a prior specification using the forecasting performance of the posterior. We warn against using such an exercise when the scope is structural analysis for at least two important reasons.[2] First, forecasting and structural performance are not two sides of the same coin. For example, suppose there are two shocks in the data, one that explains 90 percent and one that explains 10 percent of the variance of inflation. Suppose that one is interested in the dynamics induced by the latter shock. Good forecasting performance for inflation requires proper identification of the former shock and capturing well the dynamics it induces. But a prior that is tailored to that purpose will not tell us much about the dynamics in response to the second shock. In other words, a good forecasting performance is neither a necessary nor a sufficient condition for good structural inference. By the same token, a prior that flexibly accommodates prior beliefs on certain impulse responses need not have a good forecasting performance, but this should not be considered a defect of the specification.

Second, SVARs often suffer from deformation problems, see Canova and Ferroni (2022). Because systems tend to be small, structural shocks may be confounded, making structural analysis typically biased. Still, deformed systems may have good forecasting performance as long as enough lags are used. Thus, one may be able to produce decent forecasts even when the structural model is misspecified and the dynamics in response to the shocks distorted. For these two reasons, we find it inappropriate to judge a prior specification, which is specifically designed for structural objects, using the forecasting performance of the implied posterior model. If anything, introducing prior beliefs consistent with the true response of structural shocks in a deformed system increases the ability of the posterior distribution to reflect some true features of

[2]See also Todd (1992), Giannone et al. (2004) and Brignone and Piffer (2025) for an analysis of the relation between forecasting and structural analysis in vector autoregressive models.

the shock, despite the misspecification present in the model.

It is also useful to draw a short comparison with the approaches of Villani (2009), Baumeister and Hamilton (2024) and Andrle and Benes (2013). Villani (2009) writes the VAR in deviation from the steady states and designs priors for the steady states. These priors imply, given a prior for theV AR parameters, a prior specification for the constant. Our approach works the same way: we formulate prior beliefs on the VAR coefficients that reflect certain prior beliefs on the IRFs. Baumeister and Hamilton (2024) also impose priors on functions of the SVAR coefficients but work directly with the structural form. Andrle and Benes (2013) provide priors for endogenous objects of a structural (DSGE) model, such as the sacrifice ratio. These priors imply, in turn, priors on the structural parameters that enter the functions of interest. The main difference here is that a DSGE model rather than a VAR model is used in the exercise.

Our specification is also related to Jarociński and Marcet (2019), who formulate a prior on observable variables of a VAR. They rightly point out that standard priors for the VAR parameters may imply priors for observables, for example output growth, that are hard to defend and provide a way to translate subjective prior beliefs for observables into a prior for VAR coefficients. Similarly, we start from the premise that researchers are more comfortable with specifying priors on IRFs rather than VAR coefficients, making our method complementary to theirs.

Finally, our approach shares similarities with Del Negro and Schorfheide (2004). Their methodology is designed to provide structural IRFs, which are a-priori restricted by a theoretical DSGE model. Our prior restrictions are subjective and not necessarily linked to any DSGE model.

Giannone et al. (2019) have suggested a prior that describes beliefs on the long run properties of the data. Our setup can recover their specification. To see this consider equation (17). The prior for the long run is simply a prior on the sum of the VAR coefficients. Thus, by appropriately choosing the elements of $\bar{\boldsymbol{\psi}}_F$ one can either a-priori impose stationarity or unit roots. Note that while the prior of Giannone et al. (2019) is silent about the shape or the persistence of the implied IRFs, our version of the long run prior has built-in particular IRFs structures. We refer to Appendix D of the Online Appendix for a further discussion. Non-dogmatic prior restrictions on long run impulse responses were also considered by Baumeister and Hamilton (2015) within a model with sign restrictions on the contemporaneous relationship among variables.

2.4 The specification of $(\bar{\Psi}, V)$ and an illustration of our prior

To make our approach operational, we need to choose $\bar{\Psi}$. This can be difficult, due to the dimensionality of the matrix. To reduce the complexity, we found it convenient to represent $\bar{\Psi}$ via Gaussian basis functions:

$$\bar{\psi}_{ij,h} = a_{ij} \cdot e^{-\left(\frac{(h-b_{ij})^2}{c_{ij}^2}\right)+\frac{b_{ij}^2}{c_{ij}^2}}. \tag{20}$$

For each variable i and each shock j, the specification allows us to span $H+1$ dynamic responses with as few as three scalar parameters (a_{ij}, b_{ij}, c_{ij}). Here a_{ij} captures the impact effect of shock j on variable i, and thus regulates the (i,j) entry of $\bar{\Psi}_0$; b_{ij} pins down the horizon at which the peak effect is reached, and equals 0 if no hump-shaped response is desired; while c_{ij} controls for the persistence of the response. Equation (20) re-parametrizes the function used in Barnichon and Matthes (2018) to ensure that the impact effect is a free parameter, as this is needed to match $\bar{\Psi}_0$ with $E(B)$.[3]

As already mentioned, the prior mean for the VAR coefficients is consistent with a-priori beliefs on the mean of the IRFs. It is, however, silent about how the covariance matrix of the VAR coefficients V is set.

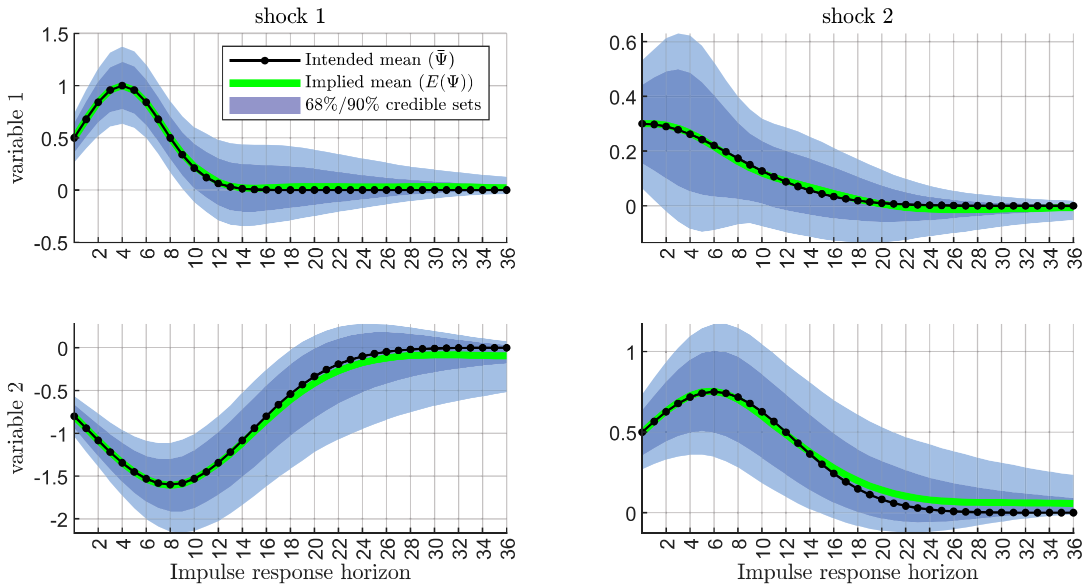
One could use the system (16) to back out a V consistent with the assumed uncertainty about the mean of the IRFs one wishes to impose. However, as it is clear from (16), such an approach implies that uncertainty increases with the lag, which is what the Minnesota-like prior does, and this is counterintuitive. Furthermore, it is difficult to formulate consistent prior variances for all impulse responses. Finally, normality is not guaranteed.

For this reason, we prefer to specify V directly so as to achieve an amount of shrinkage comparable to the Minnesota-like prior. This allows us to benefit from the same dimensionality reduction with a handful of hyperparameters, see equation (8).[4]

[3]If a prior is specified directly on B, $\bar{\Psi}_0$ can be treated as a free parameter and a_{ij} can be selected by the researcher. If a prior is specified on A or (Σ, Q), $\bar{\Psi}_0$ should be set equal to $E(B)$, which can be evaluated numerically. In this case, a researcher does not have a direct control on a_{ij}. As for c_{ij}, if $b_{ij} = 0$, we compute $c_{ij} = h_{hl,ij}/\sqrt{-\ln(2)}$ with $h_{hl,ij}$ being the horizon at which the IRF reaches its half-life relative to the impact effect, i.e. $a_{ij}/2$. If $b_{ij} > 0$, we compute $c_{ij} = b_{ij}/(\sqrt{\ln(\tilde{a}_{ij}/a_{ij})})$, with $\tilde{a}_{ij}$ capturing the maximum value of the response at horizon $h = b_{ij}$, with $\text{sign}(a_{ij}) = \text{sign}(\tilde{a}_{ij})$.

[4]One can also use hierarchical or empirical Bayes approaches, in the context of our conjugate version of the prior (19), setting V_s as in Giannone et al. (2015). We do not follow such an approach because, as mentioned, updating is generally driven by the forecasting performance of the model, which we view as conceptually separate from computing structural responses.

Figure 1: Our prior on IRFs: information on both shocks

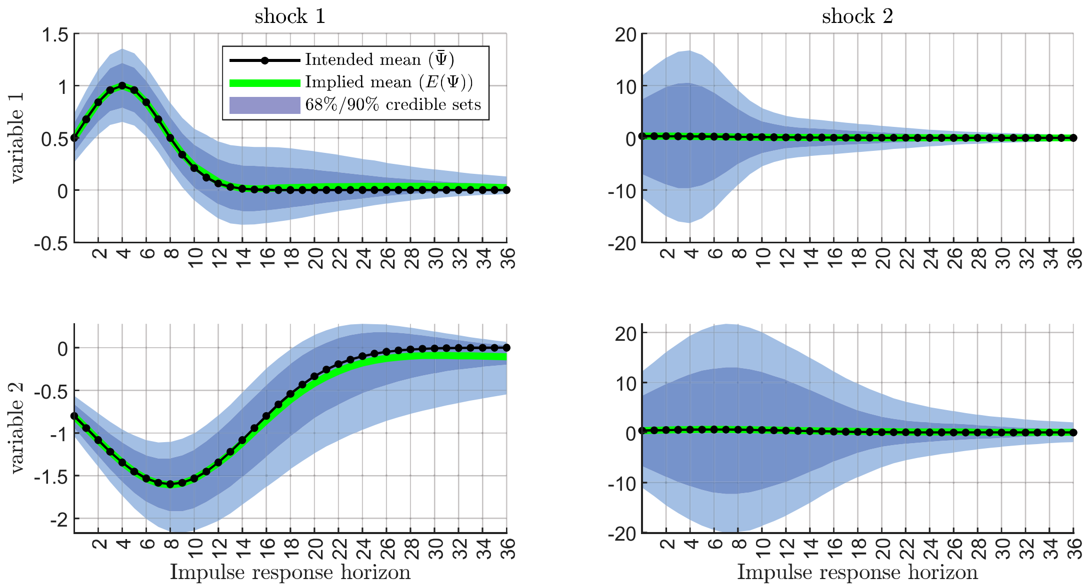


Note: The black dotted line corresponds to the prior mean $\bar{\Psi}$, specified over $H = 36$ horizons. The pointwise mean (green line) and 68%/90% credible sets (shaded areas) correspond to our prior in a VAR with $p = 12$ lags (*Case c*). The figure assumes that the prior $p(B)$ is relatively informative for both shocks.

We conclude this section with an illustration of our prior. We consider a system of 2 variables and 2 shocks, and a prior for the mean of the IRFs specified up to 36 horizons ($k = 2$, $H = 36$), so $\bar{\Psi}$ contains $k^2(H + 1) = 148$ entries. With Gaussian basis functions the selection of 148 parameters is replaced by the selection of $3k^2 = 12$ hyperparameters. The black dotted line in Figure 1 gives the $\bar{\Psi}$ selected using Gaussian Basis Function as in equation (20). The top left plot shows the case of a hump-shaped response that takes value 0.5 on impact, reaches the peak value of 1 after four horizons, before progressively declining to zero. The top-right plot, instead, shows the case of a response that equals 0.5 on impact, features no hump, and has half of the impact effect reached 9 periods after the shock.

Figure 1 also shows how our prior captures $\bar{\Psi}$ in a SVAR with $p = 12$ lags, only one third of the number of horizons of $\bar{\Psi}$ (*Case c*). We use $p(B, \boldsymbol{\pi}) = p(B) \cdot p(\boldsymbol{\pi})$, $\text{vec}(B) \sim N(\text{vec}(\bar{\Psi}_0), 0.02 \cdot I_{k^2})$ and $\boldsymbol{\pi} \sim N(\boldsymbol{\mu}, V)$, with $\boldsymbol{\mu}$ set as (18) given $\bar{\Psi}$, and $V = (0.01)^2 \cdot \text{diag}(1^{-2}, 2^{-2}, .., p^{-2}) \otimes I_{k^2}$. We draw 5,000 times from the prior for $(B, \boldsymbol{\pi})$ to compute responses. The figure reports the implied pointwise IRF means, as well as the 68% and 90% credible sets. The implied expected values of the IRFs (shown

Figure 2: Our prior on IRFs: information on shock one only



Note: The black dotted line corresponds to the prior mean $\bar{\Psi}$, specified over $H = 36$ horizons. The pointwise mean (green line) and 68%/90% credible sets (shaded areas) correspond to our prior in a VAR with $p = 12$ lags (*Case c*). The prior $p(B)$ is relatively informative on the first shock, and relatively uninformative on the second.

in the green solid lines) track $\bar{\Psi}$ (black dotted line) very well. This confirms that our prior can capture the timing and shape of the assumed responses without introducing important biases.

Figure 2 modifies Figure 1 by increasing the prior variance of the second column of B to 50, while leaving the specification for the first column unchanged. As it is clear, our prior can represent a loose prior variance on the marginal distribution of some of the impulse responses while maintaining a tighter prior variance on other responses.

3 A comparison of priors using simulated data

We illustrate the properties of our methodology using the data generated by the three-variate DSGE model of An and Schorfheide (2007). An additional illustration with simulated data from a bivariate VAR with supply and demand disturbances is in Appendix F of the Online Appendix.

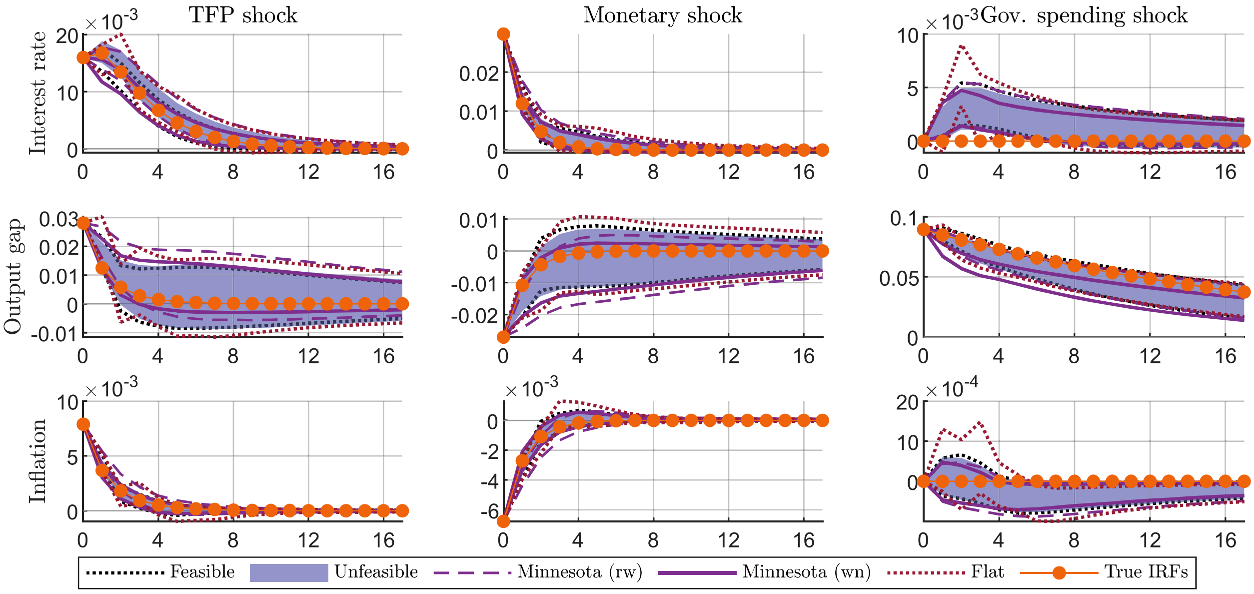
The model has three endogenous variables: the output gap, inflation, and the nominal interest rate, and their dynamics are driven by three structural shocks: a TFP shock, a monetary policy shock, and a government spending shock. We choose the structural parameters using the posterior mean estimates of An and Schorfheide (2007). Since all shocks are stationary, the model is solved in log deviation from the steady state. The data generating process generally produces short-lived impulse responses with two exceptions: the interest rate, which displays a hump-shaped response to the TFP shock, with peak one period after the shock, and the output gap, whose response to a government spending shock is very persistent (it takes 14 periods for the response to decline to half of its impact effect). Simulations are started from the steady state. We generate two datasets, one with 250 and one with 650 data points; we discard the first 100 observations and use the next 50 as a training sample. This means that we have samples of $T = 100$ or $T = 500$ to conduct inference.

We estimate five SVAR models, which are identical in the specification and include no constant and 4 lags. Employing $p > 1$ lags makes the model misspecified, given that the DGP features only one lag. However, the choice improves the visibility of the illustration, since under the working assumption that $H = p$, prior restrictions on the responses are introduced on 4 horizons, rather than only 1. The results, however, are not affected by this choice. To focus attention on the role of prior distributions for $\boldsymbol{\pi}$, we treat the impact matrix B as known and set it equal to the true values when computing posterior impulse responses.

The five models differ in the specification of $(\boldsymbol{\mu}, V)$, the Normal prior for the VAR coefficients in equation (7). The first model uses a flat prior, i.e. $V^{-1} = 0$. The second and third use a Minnesota-like prior for $\boldsymbol{\pi}$, with δ either set to zero or one, and $\lambda = 0.01$ in V via equation (8). The last two models use our prior, with $\bar{\Psi}$ specified over $H = 4$ horizons. The fourth model sets $\bar{\Psi}$ equal to the true responses up to horizon H, and describes a benchmark (unfeasible) scenario in which the researcher has true prior information about the data generating process. The fifth model specifies $\bar{\Psi}$ using the Gaussian Basis function, setting the parameters in equation (20) to ensure that the responses decline fast, reaching half of the impact effect after one period from the shock, except for the response of the output gap to the government spending shock, which is assumed to take 7 periods to reach half of the impact effect. We refer to the fifth model as the feasible specification of our prior, since it introduces some misspecification relative to the data generating process.

In the fourth and fifth model we choose the prior variance V as with the two Minnesota prior cases, to ensure comparability. Note that because $H = p$, the setup coincides with *Case a* discussed in subsection 2.2.

Figure 3: Large sample simulation, $T = 500$



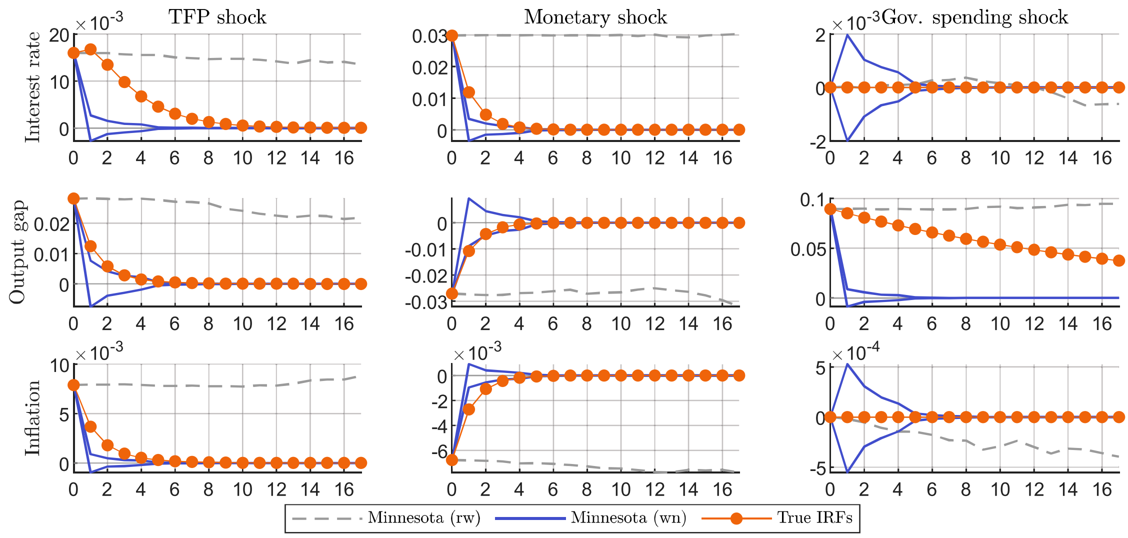
Note: The red solid-dotted line shows the true impulse responses to a one standard deviation shock. The remaining shaded areas and lines report the pointwise 90% posterior credible sets.

Figure 3 reports the pointwise 90% credible sets for the posterior IRFs in the five cases when $T = 500$. The red-dotted line displays the true impulse responses. Since $\boldsymbol{\pi}$ is identified from the data, a sufficiently large sample makes the prior for $\boldsymbol{\pi}$ irrelevant. This is clear in Figure 3, which suggests that all prior distributions lead to posterior responses that correctly replicate the true dynamics of the data generating process.

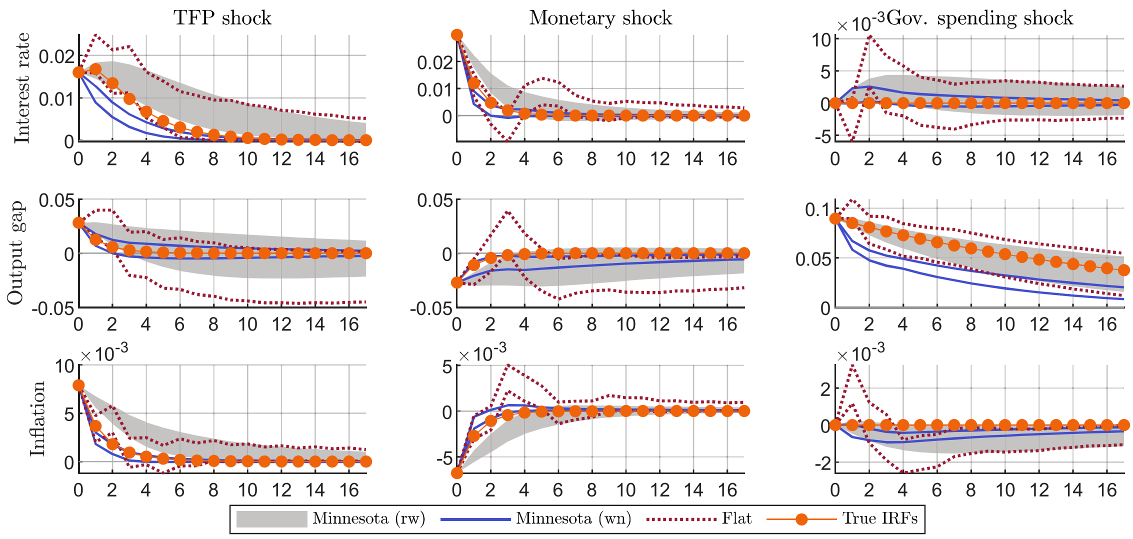
The results differ when a smaller sample is used. Figure 4 shows the results for $T = 100$ when a flat prior or the two Minnesota priors are used. The top panel reports the prior distribution (68% credible set for the white noise specification, and median for the random walk specification), while the 68% pointwise posterior credible sets are in the bottom panel. The white noise specification introduces the belief that all responses are short lived. Hence, in a small sample, the posterior fails to detect the more persistent responses of the interest rate to the TFP shock and of the output gap to the spending shock. By contrast, the random walk specification introduces the belief that all responses are very persistent. As a consequence, in a small sample, the posterior correctly detects the true responses of the interest rate to the TFP shock and of the output gap to the spending shock, at the cost of considerably overestimating the persistence of all the remaining responses. The flat prior instead leads to very

Figure 4: Minnesota and flat priors

A) Prior



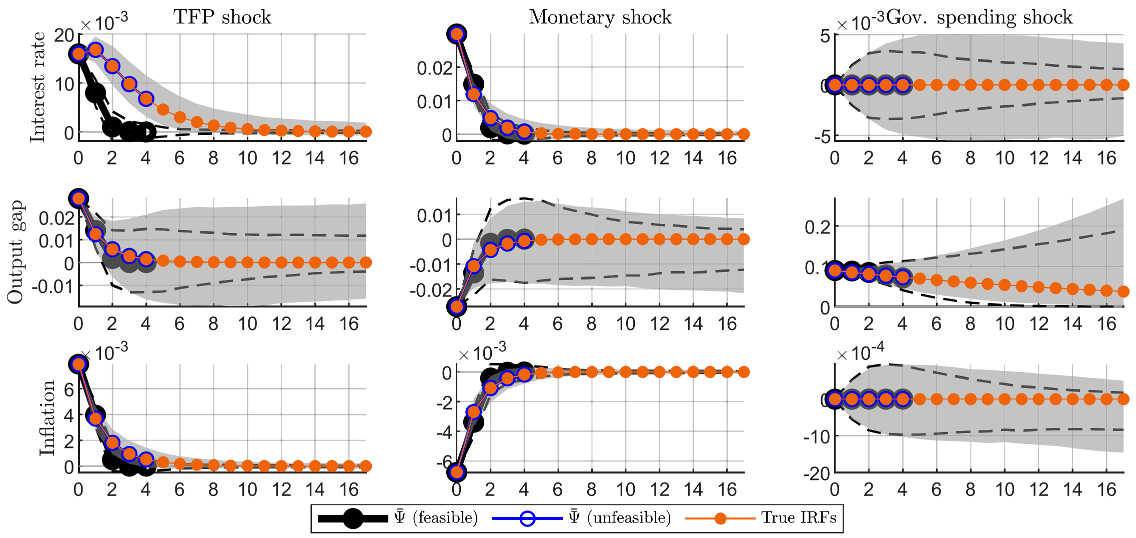
B) Posterior



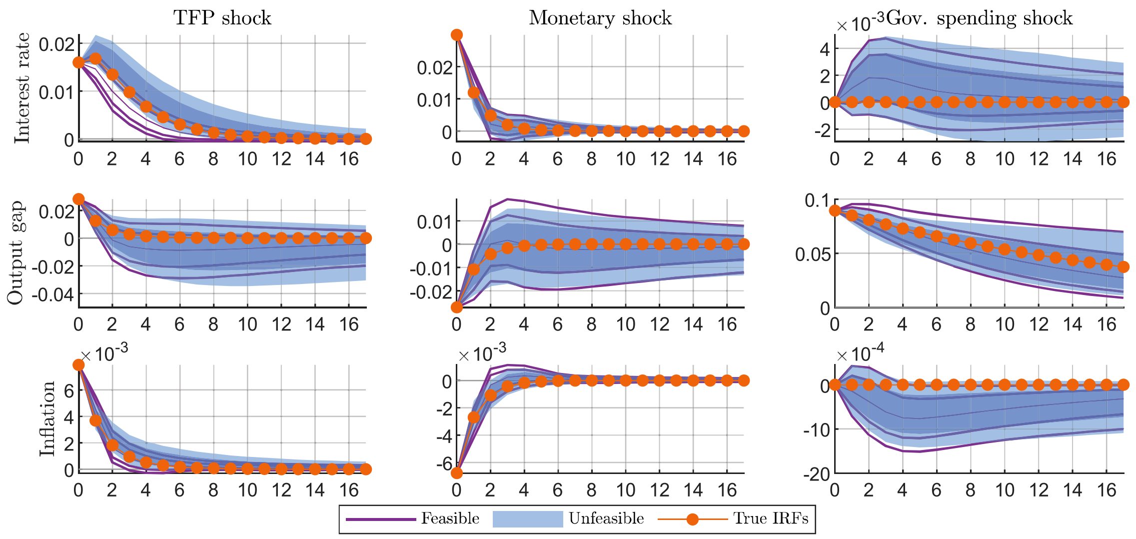
Note: The red solid-dotted line shows the true impulse responses to a one standard deviation shock. The upper panel reports the pointwise 68% prior credible sets from the white noise Minnesota prior (solid blue lines) and the prior pointwise median from the random walk Minnesota prior (grey dashed line). The lower panel, reports the 68% posterior credible sets from the white noise Minnesota prior (blue solid lines), the random walk Minnesota prior (grey shaded area) and the flat prior (red dotted lines).

Figure 5: Our prior

A) Prior



B) Posterior



Note: The red solid-dotted line shows the true impulse responses to a one standard deviation shock. The upper panel reports the target impulse response $\bar{\Psi}$ and prior 68% pointwise credible set for the unfeasible specification of our prior (blue circled line and grey shaded area) and the feasible specification (black circled line and black dashed line). The lower panel reports the posterior pointwise 68% and 90% credible sets from the unfeasible (blue shaded areas) and the feasible specification (purple solid lines).

volatile responses, due to the lack of prior shrinkage.

Our prior introduces beliefs that help to guide inference in a small sample. Figure 5 reports the results for both the unfeasible and feasible specifications. The circled lines in the top panel show the prior mean $\bar{\Psi}$, which coincides by construction with the true impulse responses up to horizon 4 in the unfeasible case (blue circled line), but not with the feasible case (black circled line). The latter underestimates the true persistence of all shocks, especially for the response of the interest rate to a TFP shock, and fails to introduce a hump in the response. Note that, even in this case, the prior distribution is well centered around $\bar{\Psi}$, suggesting that the numerical approximation implied by our computational method is small. The lower panel of the figure reports the pointwise 68% and 90% posterior credible sets. By construction, the unfeasible specification correctly reproducces the true responses, despite the short sample. Interestingly, also the feasible specification of the prior does a good job in uncovering the true impulse responses. Because the prior selection of $\bar{\Psi}$ is informed by the data generating process, it leads to posterior IRFs for relevant variables that are more persistent than the prior and mimic the true ones.

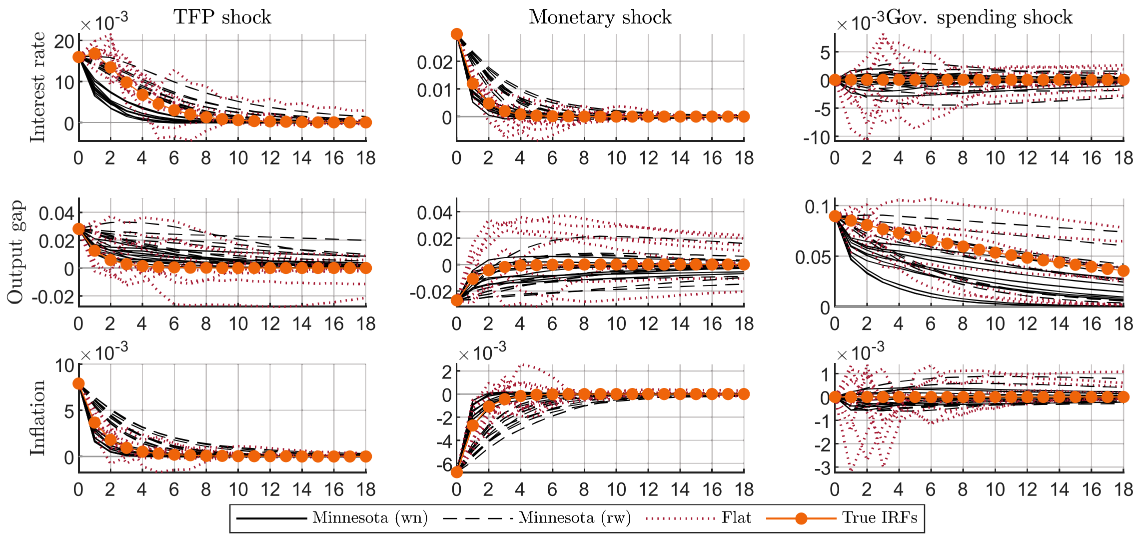
The analysis so far has considered a single dataset. We replicate the analysis for 10 datasets, holding the data generating process fixed and $T = 100$, and exploring how persistent the posterior distributions of IRFs are under the five priors for $\boldsymbol{\pi}$. As shown in Figure 6, the white noise and random walk specifications of the Minnesota prior lead one conclude that there are no persistent effects for all responses, or that all are persistent. By contract, our approach consistently detects that only two responses are relatively persistent, while the remaining ones are short-lived.

It is typically suggested that a relatively loose prior on the objects of interest should be used "to let the data speak". We find that the statement is not necessarily appropriate when performing structural analyses. As λ increases, V becomes larger, and the prior more diffuse. In our fourth and fifth models, such a change makes the prior loose its informational shape. Thus, a loose prior is less useful when the shape or persistence of a response is of interest. Figure E-6 in the Online Appendix provides an illustration of this fact. When λ increases, the posterior IRFs obtained with our specifications get closer to the posterior obtained with the flat prior. Thus, responses become more erratic and fail to reproduce the true dynamics to some of the shocks.

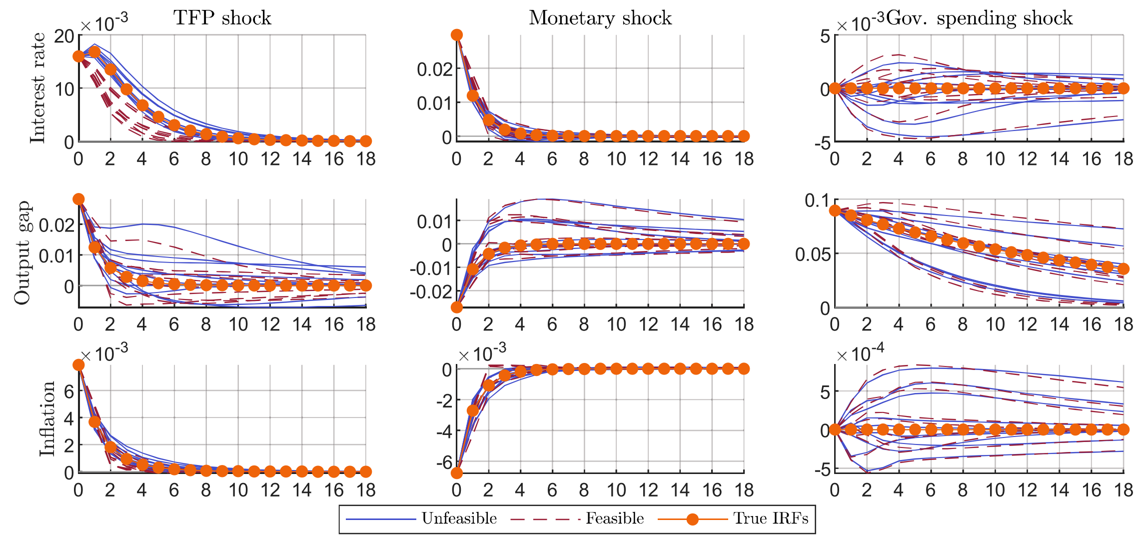
As discussed, our method controls the first moment of the responses without constraining their prior distribution to be jointly Normal. Figure E-8 in the Online Appendix shows the marginal prior distributions for the responses at horizon 8 and com-

Figure 6: Loop over 10 pseudo datasets

A) Flat and Minnesota priors



B) Our prior



Note: The red solid-dotted line shows the true impulse responses to a one standard deviation shock. The remaining lines report the pointwise posterior medians.

pare them to a Normal distribution with similar mean and standard deviation. Clearly, the marginal prior for the responses is not necessarily Normal and it can be strongly platykurtic.

On the other hand, Figure E-7 in the Online Appendix documents that using a tighter Minnesota prior exacerbates the problems we have highlighted with this specification in Figure 4. Instead, tightening our prior does not influence the relative persistence of the median posterior responses, and this is because it introduces features consistent with the data generating process.

4 The output effects of monetary policy shocks

We use our prior to quantify the effects of monetary policy shocks in the US. The VAR includes five variables: the one-year treasure interest rate, the log of industrial production, the log of the PCE deflator, the log of S&P500 index, and the GZ Excess Bond Premium. Identification of the monetary policy disturbance is achieved via (internal) instrumental variables using the 'poor man' instrument of Jarociński and Karadi (2020), which we downloaded from the authors' webpage. We follow the standard practice and include the instrument first in the list of variables and applying a recursive structure to B. The sample is 1990M2 - 2024M9. The end of the sample corresponds to the latest available date for the instrument. We use 1990M2 - 1996M12 as training sample and the remaining data for estimation. The VAR includes 12 lags, a constant and Covid dummies for 2020M2 to 2020M7 as in Castaldi-Garcia (2024).

We present the posterior distributions of IRFs associated with three different prior specifications for $\boldsymbol{\pi}$: a flat prior, a Minnesota prior, and our feasible prior. All three cases use an inverse Wishart prior for Σ with the hyperparameters set as in Kadiyala and Karlsson (1997). The Minnesota prior and our prior for $\boldsymbol{\pi}$ are chosen to be independent of Σ. The priors for the constant and the Covid dummies have mean zero and large variances. In the Minnesota prior the model is assumed to follow a random walk and the variance of the AR coefficients is set using equation (8), with $\lambda = 0.001$. Our prior has the same format for the variance of the autoregressive parameters, while $\bar{\Psi}$ is chosen as follows.

We set $\bar{\Psi}_0$ equal to the expected value of B implied by an inverse Wishart prior for Σ and the IV assumption. $E(B)$ is computed numerically using 1,000 draws from $p(\Sigma)$. We select $H = 24$, which is twice as high as the lag length p. On the other hand $\bar{\Psi}_h$, $h = 1, .., H$ is shaped as follows. The interest rate, the stock market index and the excess bond premium are assumed to increase at horizon $h = 1$, and then slowly revert back to 0, reaching half of the initial effect 6 months after the shock. The remaining variables are assumed to feature a hump-shaped response, with the through response occurring 7 months after the shock.

Figure 7 shows the prior distribution on the impulse responses to the monetary shock associated with the Minnesota prior (left column) and with our prior (middle column) [5]. The right column illustrates how $\bar{\Psi}$ from our prior (which is set over $H = 24$ horizons after the shock, *Case c*) is approximated by a VAR with $l = 3, 6, 12$ lags when using only the lags up to l of $\bar{\Psi}$ (this coincides with *Case a* of $H = p$). Figure H-10 and Figure H-11 in the Online Appendix complement the illustration by reporting the responses to the non-identified shocks.

Five facts are worth noticing about these priors. First, the prior for the responses to monetary shocks we employ is loose, and credible sets include both positive and negative values for all variables at all horizons. Second, as already mentioned, the Minnesota prior is completely uninformative about the responses to monetary shocks. Third, our prior for shocks other than the one of interest are clearly uninformative. Fourth, the pointwise prior mean of the impulse responses implied by our prior tracks $\bar{\Psi}$ well, with negligible approximation error despite the fact that $H >> p$. Finally, while the Minnesota prior implies larger uncertainty for responses at longer horizons, our prior is wider at shorter than at longer horizons. Under our prior, a VAR model with 12 lags would already approximate $\bar{\Psi}$ well under the more restrictive case of $H = p$, and implies negligible approximation errors when further allowing for $H > p$ (*Case c*, baseline specification, middle column).

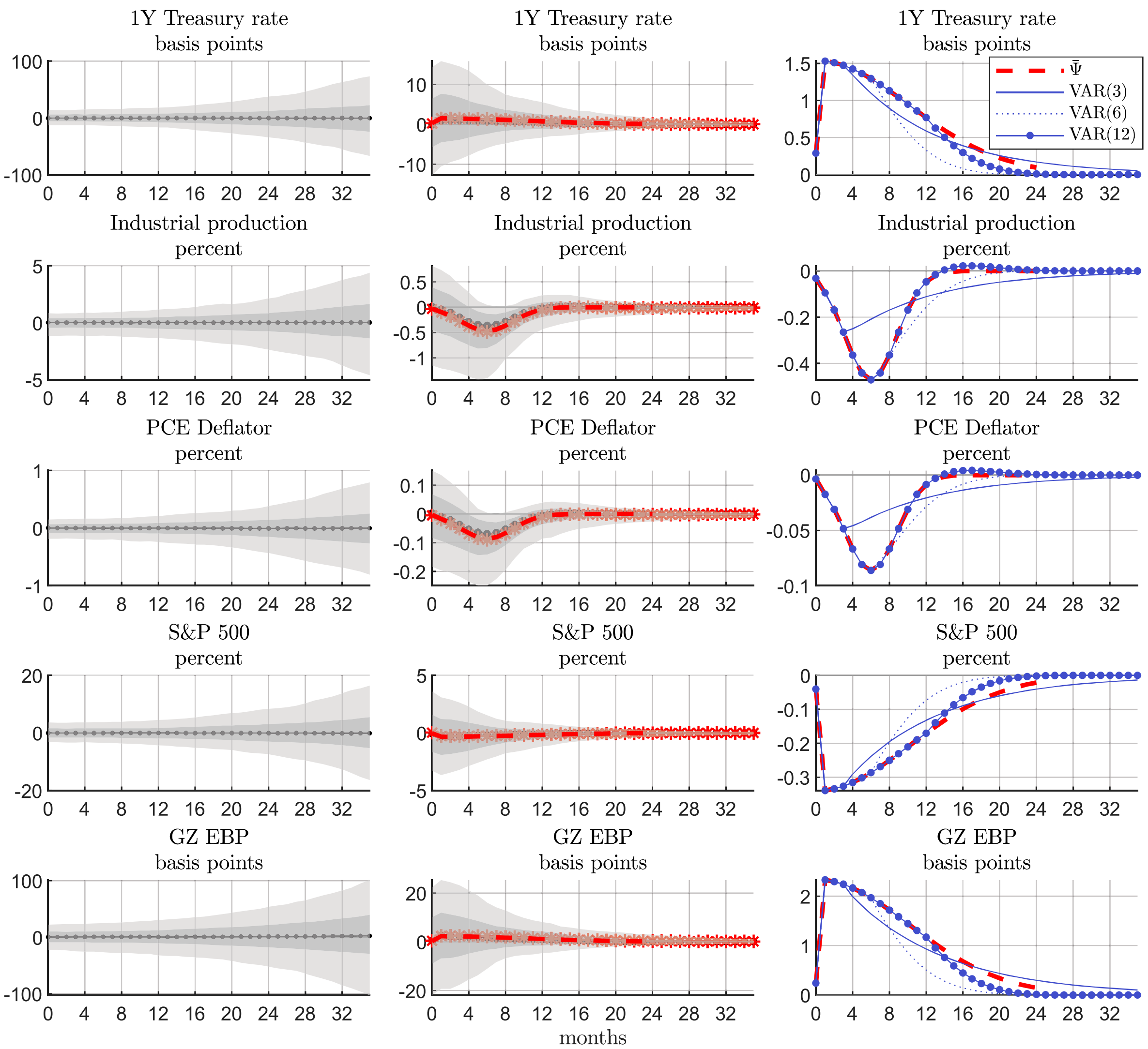
Panels a) and b) of Figure 8 show the posterior associated with the flat prior and the Minnesota prior. A flat prior leads to erratic posterior responses, while the Minnesota prior tends to produce smoother dynamics. Moreover, the posterior IRFs associated with the flat prior do not show any clear decline in industrial production nor in the PCE deflator. The posterior IRFs associated with Minnesota prior display some hump-shaped responses in industrial production. Nevertheless, the effect is ambiguous as the responses are mostly insignificant. In addition, there is also no evidence of a decrease in prices at any horizon.

Panel c) of Figure 8 shows the posterior IRFs associated with our prior. Contrary to the dynamics obtained with a flat prior or a Minnesota prior, now the responses make economic sense. The monetary shock generates an impact increase in the interest rate and the increase is highly persistent. Industrial production does not significantly respond on impact, but it progressively decreases and reaches a through eight months after the shock, before reverting back slowly to zero. The shock also generates a

[5] The expected value $\boldsymbol{\mu}$ associated with the above specification of $\bar{\Psi}$ is calculated drawing from the resulting prior $p(\boldsymbol{\pi}, \Sigma)$, and computing responses.

Figure 7: Prior Responses to monetary policy shocks

(a) Minnesota prior **(b)** Our prior **(c)** $\bar{\Psi}$ and VAR approximation

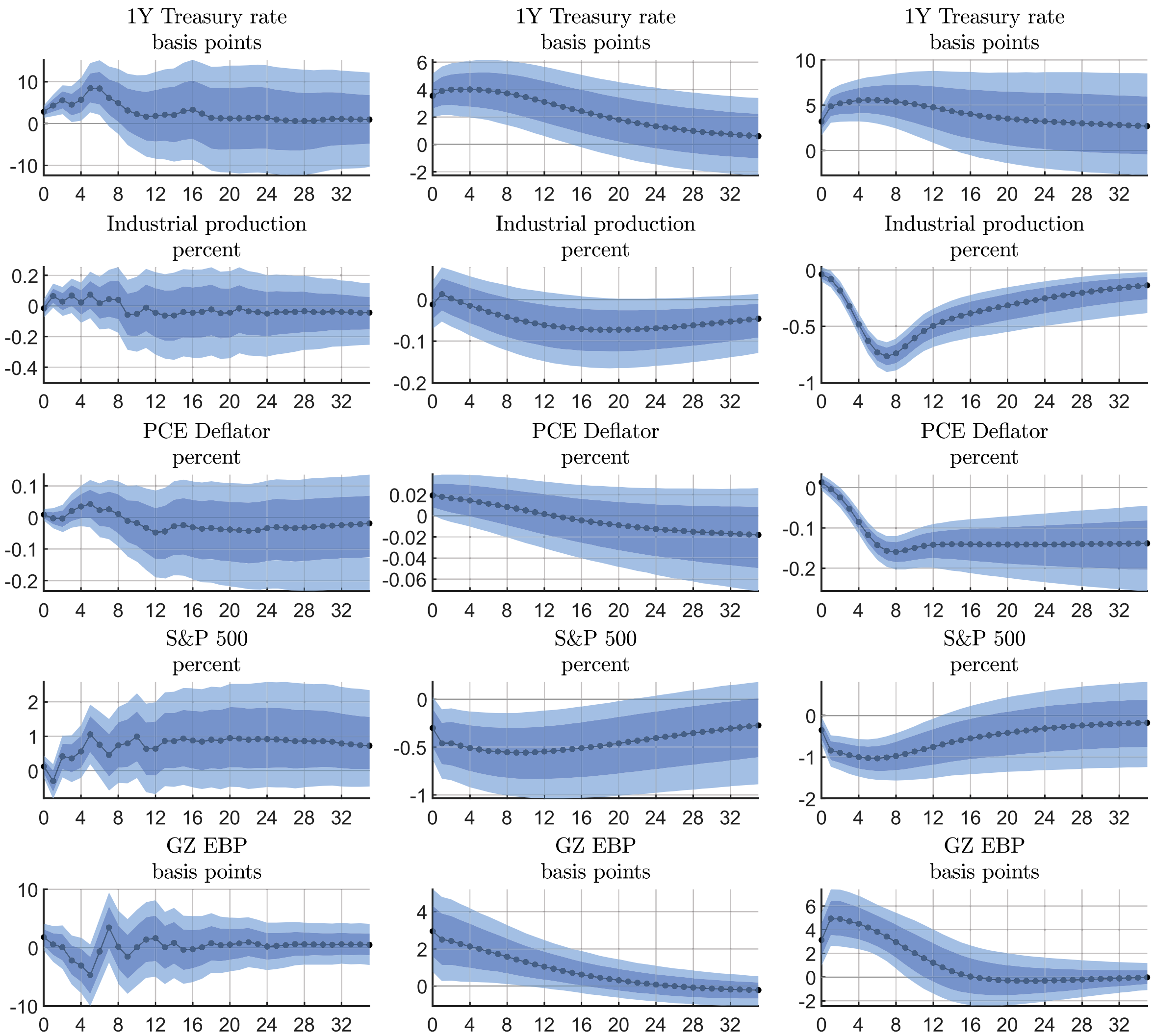


Note: In panel a) and b), shaded areas report the pointwise prior 68% and 90% credible sets while the dotted lines report the pointwise prior median. In panel b), the red dashed line presents $\bar{\Psi}$, while the star marks represent $E(\Psi)$, computed numerically from the prior draws. In panel c), we take entries $[\bar{\Psi}_0, .., \bar{\Psi}_l]$ of $\bar{\Psi}$ with $l = 3, 6, 12$, use equation (9) to compute the implied parameters of a VAR(l) model, and report the associated impulse responses up to horizon 36.

smooth decline in PCE deflator, which reaches a minimum about eight-nine months after the shock. Note also that the monetary contraction is associated with an initial

Figure 8: Posterior Responses to monetary policy shocks

(a) Flat prior **(b)** Minnesota prior **(c)** Our prior



Note: Shaded areas report the pointwise posterior 68% and 90% credible sets. Dotted lines report the pointwise median.

drop in the stock market index, which reverts after a year and a half. The excess bond premium increases on impact and then progressively reverts back to zero.

The hump-shaped response of industrial production (IP) is a posteriori quite common with our prior specification. Table 1 documents this fact. It reports the prior and posterior probability that the impulse response of industrial production reaches

Table 1: Prior and posterior probability of a hump-shaped impulse response of industrial production at selected horizons

		periods after the shock						
		$h = 5$	$h = 6$	$h = 7$	$h = 8$	$h = 9$	$h = 10$	$h = 7 - 9$
Our prior	prior	3.50%	18.30%	38.20%	23.40%	4.40%	0.10%	66.00%
	posterior	0.00%	0.00%	0.27%	96.19%	3.45%	0.00%	99.90%
Minnesota prior	prior	1.10%	1.00%	0.90%	0.70%	0.90%	1.00%	2.50%
	posterior	0.66%	0.52%	0.58%	0.71%	0.90%	1.84%	2.19%

Note: The last column reports the cumulative prior and posterior probability that a hump occurs between horizons 7 and 9.

a minimum at either a given horizon after the shock or within an interval of periods. With our prior, in over 96% of the posterior draws the through occurs eight periods after the shock, and in 99.9% of the draws it occurs between 7 and 9 periods after the shock. A-priori, the through generally occurs after 7 months and in only 23% of the draws the through appears exactly 8 months after the shock. In contrast, with the Minnesota prior, only 2.19% of the posterior draws imply a through between 7 and 9 periods after the shock, a probability that is virtually not updated (actually reduced) relative to the prior probability of 2.50%. Hence, contrary to the Minnesota prior, our prior helps to obtain a pattern of responses in line with the conventional wisdom.

In Section H of the Online Appendix we show that the results are unaffected if we exclude the Covid dummies from the VAR or end the estimation sample before Covid. Under the Minnesota prior, both specifications lead to equally inconclusive results. The results are also robust when we tighten the prior on $\boldsymbol{\pi}$ or shorten H, the horizon until which a prior on the responses is specified. Lastly, the Online Appendix shows that the humped shaped prior is important and that results are indeed affected if IP and price responses are a-priori assumed to be non-informative.

5 Conclusions

Bayesian VAR models are frequently used to estimate impulse response functions to structural shocks. This paper develops a tractable prior distribution for VAR coefficients that achieves two goals. First, it allows for an explicit introduction of prior beliefs on the shape or the persistence of the impulse responses. Second, it does so by working with a Normal prior distribution which ensures tractable posterior sampling.

We illustrate the properties of the methodology using simulated data from a small scale DSGE model, We discuss how key hyperparameters of the prior can be specified. We demonstrate the flexibility of our prior specification and its properties relative to

Minnesota-style and flat prior choices.

We then use the prior we design to investigate how long it takes for a monetary policy shock to generate its largest effect on industrial production. We show that the popular flat and Minnesota priors lead to posterior IRFs that feature no hump-shaped response for industrial production. By contrast, our prior, which is set to mimic the belief that monetary policy shocks generate persistent yet temporary effects, leads to a posterior that features a hump-shaped response of industrial production. We estimate that it takes around 8 months for the maximum effect to materialize.

Our work can be extended in many ways. For example, one can think about using a similar approach to deal with prior beliefs in multicountry VARs of the type studied by Canova and Ciccarelli (2009) or in local projection exercises. It is also possible to impose prior restrictions on partial and cumulative multipliers or on the contribution of certain shocks to the variance of the endogenous variables. We leave all these extensions to future work.

References

Amir-Ahmadi, P. and Drautzburg, T. (2021), 'Identification and inference with ranking restrictions', *Quantitative Economics* **12**(1), 1–39.

An, S. and Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', *Econometric reviews* **26**(2-4), 113–172.

Andrle, M. and Benes, J. (2013), 'System priors: Formulating priors about DSGE models' properties', *IMF working paper* (13/257).

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018), 'Narrative sign restrictions for SVARs', *American Economic Review* **108**(10), 2802–29.

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.

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2024), 'Uniform priors for impulse responses', *CEPR working paper 18836* .

Bańbura, M., Giannone, D. and Reichlin, L. (2010), 'Large Bayesian vector auto regressions', *Journal of Applied Econometrics* **25**(1), 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.

Baumeister, C. and Hamilton, J. D. (2024), 'Advances in using vector autoregressions to estimate structural magnitudes', *Econometric Theory* **40**(3), 472–510.

Baumeister, C. J. and Hamilton, J. D. (2018), 'Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations', *Journal of Monetary Economics* **100**, 48–65.

Binning, A. (2013), Underidentified SVAR models: A framework for combining short and long-run restrictions with sign restrictions, Technical report, Norges Bank working paper series.

Brignone, D. and Piffer, M. (2025), 'A structural var model for the uk economy', *Bank of England Macro Technical Paper* (3).

Bruns, M. and Piffer, M. (2023), 'A new posterior sampler for Bayesian structural vector autoregressive models', *Quantitative Economics* **4**(14), 1221–1250.

Caldara, D. and Iacoviello, M. (2022), 'Measuring geopolitical risk', *American Economic Review* **112**(4), 1194–1225.

Caldara, D. and Kamps, C. (2017), 'The analytics of SVARs: a unified framework to measure fiscal multipliers', *The Review of Economic Studies* **84**(3), 1015–1040.

Canova, F. (2007), *Methods for applied macroeconomic research*, Vol. 13, Princeton University Press.

Canova, F. and Ciccarelli, M. (2009), 'Estimating multicountry var models', *International economic review* **50**(3), 929–959.

Canova, F. and Ferroni, F. (2022), 'Mind the gap! Stylized dynamic facts and structural models', *American Economic Journal: Macroeconomics* **14**(4), 104–135.

Canova, F. and Pappa, E. (2011), 'Fiscal policy, pricing frictions and monetary accommodation', *Economic Policy* **26**(68), 555–598.

Canova, F. and Pérez Forero, F. J. (2015), 'Estimating overidentified, non-recursive, time varying coefficients structural VARs', *Quantitative Economics* **6**(2), 309–358.

Castaldi-Garcia, D. (2024), Pandemic priors, Technical report, Federal Reserve Board.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999), 'Monetary policy shocks: What have we learned and to what end?', *Handbook of Macroeconomics* **1**, 65–148.

Del Negro, M. and Schorfheide, F. (2004), 'Priors from general equilibrium models for VARs', *International Economic Review* **45**(2), 643–673.

Dwyer, M. (1998), 'Impulse response priors for discriminating structural vector autoregressions'.

Ferreira, L. N., Miranda-Agrippino, S. and Ricco, G. (2023), 'Bayesian local projections', *The Review of Economics and Statistics* pp. 1–45.

Giacomini, R. and Kitagawa, T. (2021), 'Robust Bayesian inference for set-identified models', *Econometrica* **89**(4), 1519–1556.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Giannone, D., Lenza, M. and Primiceri, G. E. (2019), 'Priors for the long run', *Journal of the American Statistical Association* **114**(526), 565–580.

Giannone, D., Reichlin, L. and Sala, L. (2004), 'Monetary policy in real time', *NBER macroeconomics annual* **19**, 161–200.

Gilchrist, S. and Zakrajšek, E. (2012), 'Credit spreads and business cycle fluctuations', *American Economic Review* **102**(4), 1692–1720.

Gordon, S. and Boccanfuso, D. (2001), 'Learning from structural vector autoregression models', *Manuscript, Universite Laval, Quebec City* .

Gourinchas, P.-O. and Tornell, A. (2004), 'Exchange rate puzzles and distorted beliefs', *Journal of International Economics* **64**(2), 303–333.

Gupta, A. K. and Nagar, D. K. (2000), *Matrix Variate Distributions*, Chapman and Hall/CRC.

Harvey, A. C., Trimbur, T. M. and Van Dijk, H. K. (2007), 'Trends and cycles in economic time series: A Bayesian approach', *Journal of Econometrics* **140**(2), 618–649.

Hudak, D. and Richter, G. (1996), 'Moments of Special Normally Distributed Matrices', *Statistics* **27**(3-4), 363–378.

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

Jarociński, M. and Karadi, P. (2020), 'Deconstructing monetary policy surprisesâthe role of information shocks', *American Economic Journal: Macroeconomics* **12**(2), 1–43.

Jarociński, M. and Marcet, A. (2019), 'Priors about observables in vector autoregressions', *Journal of Econometrics* **209**(2), 238–255.

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.

Kilian, L. (2022), 'Comment on Giacomini, Kitagawa, and Read's "narrative restrictions and proxies"', *Journal of Business & Economic Statistics* **40**(4), 1429–1433.

Kilian, L. and Lütkepohl, H. (2017), *Structural vector autoregressive analysis*, Cambridge University Press.

Kilian, L. and Murphy, D. P. (2012), 'Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models', *Journal of the European Economic Association* **10**(5), 1166–1188.

Kocięcki, A. (2010), 'A prior for impulse responses in Bayesian structural VAR models', *Journal of Business & Economic Statistics* **28**(1), 115–127.

Kollo, T. and von Rosen, D. (2005), *Advanced Multivariate Statistics with Matrices*, Springer.

Koop, G. and Korobilis, D. (2010), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**(4), 267–358.

Miranda-Agrippino, S. and Ricco, G. (2019), Bayesian vector autoregressions: Estimation, *in* 'Oxford Research Encyclopedia of Economics and Finance'.

Miranda-Agrippino, S. and Ricco, G. (2021), 'The transmission of monetary policy shocks', *American Economic Journal: Macroeconomics* **13**(3), 74–107.

Paul, P. (2020), 'The time-varying effect of monetary policy on asset prices', *Review of Economics and Statistics* **102**(4), 690–704.

Piffer, M. and Podstawski, M. (2018), 'Identifying uncertainty shocks using the price of gold', *The Economic Journal* **128**(616), 3266–3284.

Plagborg-Møller, M. (2019), 'Bayesian inference on structural impulse response functions', *Quantitative Economics* **10**(1), 145–184.

Plagborg-Møller, M. and Wolf, C. K. (2021), 'Local Projections and VARs Estimate the Same Impulse Responses', *Econometrica* **89**(2), 955–980.

Ramey, V. A. (2016), 'Macroeconomic shocks and their propagation', *Handbook of macroeconomics* **2**, 71–162.

Todd, R. M. (1992), 'Algorithms for explaining forecast revisions', *Journal of Forecasting* **11**(8), 675–685.

Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**(2), 381–419.

Van Dijk, H. K. and Kloek, T. (1980), 'Further experience in Bayesian analysis using monte carlo integration', *Journal of Econometrics* **14**(3), 307–328.

Villani, M. (2009), 'Steady-state priors for vector autoregressions', *Journal of Applied Econometrics* **24**(4), 630–650.

Waggoner, D. F., Wu, H. and Zha, T. (2016), 'Striated Metropolis–Hastings sampler for high-dimensional models', *Journal of Econometrics* **192**(2), 406–420.