Bayesian Structural VAR models: a new approach for prior beliefs on impulse responses

Martin Bruns† Michele Piffer‡

September 17, 2018

PRELIMINARY. Most recent version here

Abstract

Structural VAR models are frequently identified using restrictions on the sign of the impulse response to selected structural shocks. However, the existing methods either provide only limited flexibility on the prior beliefs that can be imposed on impulse responses, or are computationally demanding. We propose an approach that allows for extensive flexibility on the contemporaneous impulse responses while still allowing for a tractable posterior sampling. We illustrate the new methodology on simulated data in a controlled experiment exercise. We then show an application to US data and document that monetary policy shocks and financial shocks have quantitatively similar effects on economic activity.


Keywords: Vector autoregressions, sign restrictions, Bayesian inference, set identification.

*We are thankful to Lutz Kilian, Helmut Lütkepohl, Haroon Mumtaz and Eduardo Palombo for helpful comments and suggestions. Many thanks to Taiki Yakamura for excellent research assistance. Michele Piffer thanks for the financial support received from the European Union’s Horizon 2020 research and innovation program, Marie Skłodowska-Curie grant agreement number 744010.
†Freie Universität Berlin and German Institute for Economic Research (DIW Berlin).
‡Queen Mary, University of London. Corresponding author e-mail: m.b.piffer@gmail.com


1 Introduction

Structural Vector Autoregressive (VAR) models are extensively used in applied work. It is well known that these models are not identified unless restrictions are introduced (Kilian and Lütkepohl, 2017). One popular approach consists of imposing sign restrictions on selected parameters of interest, either within a frequentist set-identified setting (Moon et al., forthcoming), or, as in this paper, using Bayesian informative prior beliefs that reflect the intended signs (Uhlig, 2005, Baumeister and Hamilton, 2015, Arias et al., forthcoming).

This paper relates to the literature that studies how to best implement sign restrictions on a particular statistic, namely on impulse responses. Write the structural and the reduced form models as

\[ Ay_t = A_w + \epsilon_t, \quad \epsilon_t \sim N(0, I_k), \]  
\[ y_t = \Pi w_t + u_t, \quad u_t \sim N(0, \Sigma), \]

respectively, and the structural moving average form as

\[ y_t = B_0 \epsilon_t + B_1 \epsilon_{t-1} + B_2 \epsilon_{t-2} + ..., \quad \epsilon_t \sim N(0, I_k). \]

Contemporaneous impulse responses are captured by \( B_0 = A^{-1} \), while future responses are given by \( \{B_h\}_{h=1}^H \) and can be computed recursively using \( A^{-1} \) and \( \Pi \). The literature frequently expresses prior beliefs on the reduced form parameters and maps reduced form draws into draws of the impulse responses using draws of orthogonal matrices \( Q \) obtained from the algorithm by Rubio-Ramirez et al. (2010), see for example Arias et al. (forthcoming) and in part Uhlig (2005). While computationally convenient due to the existence of prior beliefs for the reduced form parameters that make posterior sampling tractable, this approach provides limited flexibility on the priors implied on the actual parameters of interest, namely the impulse responses, as remarked by Baumeis-
These shortcomings are partly addressed by Plagborg-Møller et al. (forthcoming), who departs from prior beliefs directly on the moving average form. However, his method makes the analysis computationally more demanding, as a Monte Carlo sampler for the posterior is required for all parameters.

The first contribution of this paper consists of developing an approach offering a compromise between the methods by Arias et al. (forthcoming) and by Plagborg-Møller et al. (forthcoming). We first note that in the applied literature, sign restrictions on impulse responses are typically imposed on contemporaneous responses, because the sign for these responses appear to be more robust across DSGE models (Canova and Pina, 2005). Accordingly, we parametrize the model in $A^{-1}, \Pi$ rather than in $\Pi, \Sigma, Q$ (as in Arias et al., forthcoming) or in $\{B_h\}_{h=0}^H$ (as in Plagborg-Møller et al., forthcoming). By expressing prior beliefs directly on $A^{-1}$, our approach benefits from extensive flexibility where it is needed the most, offering the researcher flexibility on the prior beliefs on the contemporaneous impulse responses while still exploiting conventional priors for $\Pi$ that make posterior sampling tractable. Accordingly, a Monte Carlo sampler is required for only a subset of parameters.

The second contribution of the paper relates to the sampling of the marginal posterior $p(A^{-1}|Y)$. To account for applications to potentially large models, we follow Waggoner et al. (2016) and use a sequential approach that does not sample the posterior directly, but samples a more tractable kernel distribution that is progressively modified to equal the posterior distribution of interest. Yet, we depart from Waggoner et al. (2016) and do not initiate the sampling from the prior, but from the the posterior distribution that the approach by Arias et al. (forthcoming) implies for $A^{-1}$, which we derive analytically. As such, our approach allows for a wider flexibility on prior beliefs on impulse responses relative to Arias et al. (forthcoming) while still making use of the computational convenience of their approach.

We first develop the intuition of our approach using an application to data simulated from the An and Schorfheide (2007) model. We use this framework to discuss in
details the selection of the priors and the sampler. We then apply our method to US data. We estimate a SVAR model identifying at the same time a monetary shock and a financial shock. The aim of the analysis is to assess whether the effect of a monetary shock on output is quantitatively comparable to the effect on output exerted by a financial shock. The stronger the effect on output of a financial shock relative to a monetary shock, the less the monetary authority can use monetary policy to offset financial shocks. Indeed, the 2007 crisis has shed extensive light on how far reaching the consequences of financial shocks can be on the economy. We document that monetary shocks affect output quantitatively in a similar way relative to financial shocks, suggesting room for the monetary authority to potentially offset financial shocks.

The paper is strongly related to Baumeister and Hamilton (2015). As in their contribution, we work in a framework in which prior beliefs are expressed directly on the structural parameters of interest rather than indirectly via priors on reduced form parameters and orthogonal matrices. We find this important, since the latter approach potentially introduces restrictions which go beyond the intention of the researcher, and on key parameters of the model. Contrary to Baumeister and Hamilton (2015), we study the case of prior beliefs on contemporaneous impulse responses ($A^{-1}$) rather than on the contemporaneous relation among variables ($A$). We find it important to investigate this alternative case because, as remarked for example by Kilian and Lütkepohl (2017), the case of prior beliefs being available on $A$ rather than $A^{-1}$ is relatively rare in applied work. One step in this direction is provided by Baumeister and Hamilton (forthcoming), where prior beliefs on $A$ are combined with prior beliefs on $A^{-1}$. Contrary to their paper, we do not combine information on both $A$ and $A^{-1}$ since in their approach this implies that the prior on either $A$ or $A^{-1}$ is ultimately available only numerically. In addition, we propose an alternative approach for posterior sampling, which can adapted to their framework.

The paper also relates to Kociecki (2010), who departs from explicit prior beliefs on impulse responses. Contrary to his approach, we apply identification through sign
restrictions and hence avoid the more restrictive recursive structure. As we document in the appendix, our approach can be extended to the imposition of prior beliefs on impose responses beyond the contemporaneous effect, making the analysis more in line with the work by Kociecki (2010). However, we document that when a non-recursive structure is used, in his framework posterior sampling becomes computationally much more demanding, whereas the approach that we propose reaches a more balanced compromise that still exploits a tractable posterior distribution conditioning on $A^{-1}$.

The paper is organized as follows. Section 2 outlines the methodology proposed and discuss its relation to the existing literature, while leaving the discussion of the posterior sampler mainly to the appendix. Section 3 shows the illustrative example on simulated data in order to convey additional intuition on the method proposed. Section 4 reports the application to US data and the identification of monetary and financial shocks. Section 5 concludes.

2 The model

In this section we discuss in details the methodology that we propose, and relate it to the existing literature.

2.1 Different parametrizations of the structural model

Structural VAR models can be written in different forms, and the difference is crucial for the present paper. The most popular specification of structural VAR models is

$$Ay_t = a_{+0} + \sum_{l=1}^{p} A_{+l}y_{t-l} + \epsilon_t, \quad \epsilon_t \sim N(0, I_k),$$

$$= A_{+}w_t + \epsilon_t, \quad \epsilon_t \sim N(0, I_k), \quad (4)$$

where $y_t$ is a $k \times 1$ vector of endogenous variables, $\epsilon_t$ is a $k \times 1$ vector of structural shocks, and $w_t = (1, y'_{t-1}, \ldots, y'_{t-p})'$ is an $m \times 1$ vector of the constant and $p$ lags
of the variables, with \( m = kp + 1 \). The \( k \times k \) matrices \( A, A_1, \ldots, A_p \) are gathered in \( A_+ = [a_{+,0}, A_1, \ldots, A_p] \), which is of dimension \( k \times m \). For simplicity, we normalize the covariance matrix of \( \epsilon_t \) to the identity matrix rather than to a more general diagonal matrix.\(^1\)

Model (4) highlights the structural nature of the equations. Yet, as is well known, one can rewrite the structural form after premultiplying both sides of the equation by \( B = A^{-1} \), obtaining

\[
y_t = \pi_{+,0} + \sum_{l=1}^{p} \Pi_{+,l} y_{t-l} + B \epsilon_t, \quad \epsilon_t \sim N(0, I_k),
\]

\[
y_t = \Pi w_t + B \epsilon_t, \quad \epsilon_t \sim N(0, I_k),
\]

with \( \pi_{+,0} = A^{-1} a_{+,0} \), \( \Pi_l = A^{-1} A_{+,l} \) and \( \Pi = [\pi_{+,0}, \Pi_1, \ldots, \Pi_p] = A^{-1} A_+ \). Matrix \( B \) captures the contemporaneous effects of one standard deviation shocks, while future horizons of the impulse responses can be calculated using model (5) recursively. Alternatively, one can compute impulse responses using the structural moving average representation of the data,

\[
y_t = b_0 + \sum_{l=1}^{\infty} B_l \epsilon_{t-l} + B_0 \epsilon_t, \quad \epsilon_t \sim N(0, I_k),
\]

where it holds that \( B_0 = B \). For the full mapping of \( (\Pi, B) \) into \( \{B_l\}_{l=0}^{\infty} \) see, for example, Kilian and Lütkepohl (2017), Chapter 2.

The difference between specification (4) and (5) of the structural model is crucial for the analysis of this paper. Equation (4) highlights the contemporaneous relations among the variables in the system, as captured by matrix \( A \). Because of these relations, a structural shock that hits one variable potentially affects contemporaneously

\(^1\)In doing so, we depart from Baumeister and Hamilton (2015), who exploit conjugate priors for
the variance of the structural shocks. We apply this normalization because it is frequently used in
applications that employ sign restrictions on the effects of the shocks, see for example Uhlig (2005)
and Arias et al. (forthcoming).
all variables, in a way that is captured by matrix $B$. Whether the model is more conveniently expressed as (4) or (5) (or even in other forms) depends on whether the identifying restrictions used by the researcher to identify the model are more naturally expressed on $A$ or on $B$, as further discussed below.\textsuperscript{2}

The reduced form representation of the data is

$$y_t = \pi_{+,0} + \sum_{l=1}^{p} \Pi_{+,l} y_{t-l} + u_t, \quad u_t \sim N(0, \Sigma),$$

$$= \Pi w_t + u_t, \quad u_t \sim N(0, \Sigma),$$

(7)

with $u_t = B \epsilon_t$ and $\Sigma = BB'$. Matrices $A$, $A_+$ and $B$ contain structural parameters, while $\Pi$ and $\Sigma$ represent reduced form parameters. Orthogonal matrices $Q$, which by construction satisfy $QQ' = I_k$, allow for the mapping from reduced form to structural parameters, with

$$B = h(\Sigma)Q,$$

(8)

and $h(\Sigma)$ any decomposition of $\Sigma$ satisfying $h(\Sigma)h(\Sigma)' = \Sigma$, for example the Choleski decomposition.

Arias et al. (forthcoming) refer to $(A_0, A_+)$ from equation (4) as the ‘structural’ parametrization, to $(\Pi, \Sigma, Q)$ from equation (7) and equation (8) as the ‘orthogonal

\textsuperscript{2}In the terminology used by Lütkepohl (2005), model (4) represents the $A$ form of the model while model (5) represents the $B$ form of the model. To appreciate the importance of the distinction, note that restrictions imposed on one form might not be apparent in the other form, due to the nonlinearities in the mapping from one to another. As an example, the $A$ model by Sims and Zha (2006) imposes the non-recursive zero restrictions $A = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix}$, which implies the $B$ form $A^{-1} = B = \begin{pmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$. A well-known special case is the recursive structure of $A$, which implies a recursive structure of $B$. Going through the publications of all top-five journals and the Journal of Monetary Economics since 1998, we found that around 10% employs Structural Vector autoregressive models. Of this 10%, approximately 21% specifies the model in the $A$ form, 72% specifies the model in the $B$ form, and 5% specifies the model in the hybrid $AB$ form. The details list is available here [ADD HYPERLINK]. Amisano and Giannini (2012) had previously referred to models (4) and (5) as the $K$ and $C$ form of the SVAR, respectively.
reduced form’ parametrization, and to \((B_0, B_1, ..., B_p)\) from equation (6) as the ‘impulse response function’ parametrization. Building on this terminology, \((\Pi, B)\) from equation (5) could be referred to as the ‘reduced form structural’ parametrization. In the rest of the paper, we refer to the direct approach when we use the ‘reduced form structural parametrization’, while we refer to the indirect approach when using the ‘orthogonal reduced form parametrization’. We use this terminology to highlight whether prior beliefs are directly expressed on the structural parameters of interest or not.

2.2 The direct approach proposed in this paper

This paper studies the case in which the researcher has prior beliefs on the contemporaneous effect of selected structural shocks of interest. For this type of restrictions, the most natural parametrizations of the model are the ‘reduced form structural’ parametrization from model (5) and the ‘impulse response function’ parametrization from model (6), because the parameters in \(B\) or \(B_0\) directly capture these effects. We follow the ‘reduced form structural’ parametrization and use model (5) since it makes the analysis considerably more tractable, as we discuss in ???. For an analysis that departs from model (6) see Plagborg-Møller et al. (forthcoming).

Define \(\pi = \text{vec}(\Pi)\) as the vector of dimension \(km \times 1\) that stacks the columns of \(\Pi\). The parameters of model (5) are then \(\pi, B\). Prior beliefs are captured by the joint prior distribution

\[
p(\pi, B) = p(\pi | B) \cdot p(B).
\]

Since we aim to contribute to the literature on identification, flexibility on \(p(B)\) is more valuable than flexibility on \(p(\pi | B)\). Hence, as common in the literature, we restrict \(p(\pi | B)\) to

\[
\pi \sim N(\mu_\pi, V_\pi).
\]

By contrast, \(p(B)\) can fall within a wide range of prior distributions, granting the
researcher flexibility on the prior beliefs used to express sign restrictions on structural parameters. Note that by not imposing a Kronecker structure on $V_\pi$, $p(\pi|B)$ allows for the popular prior by Litterman (1986), treating the variance on ‘own lags’ and ‘lags on other variables’ differently. Parametrizing the model in $\Pi$ rather than in $A_+ = B\Pi$ makes it straightforward to apply the prior by Litterman (1986) and simplifies the analysis compared to Sims and Zha (1998) and Baumeister and Hamilton (2015), who depart from model (4). In addition, it simplifies the comparison of the direct and the indirect approach, as $\Pi$ appears in both parametrizations.\footnote{The selection of parametrization in $\Pi$ rather than $A_+$ does not affect the number of parameters for which the MCMC algorithm is required, because both $p(\Pi|B,Y)$ (when parametrizing the model in $\Pi$) and $p(A_+|B,Y)$ (when parametrizing the model in $A_+$) have a common form that can be drawn from using available random number generators. In the special case of the $A$ form, Sims and Zha (1998) and Baumeister and Hamilton (2015) show that the parametrization in $A,A_+$ together with prior independence across the parameters in the different structural equations of the model allow for posterior simulation equation by equation. In the more general framework considered in this paper, the presence of $B \neq I_k$ prevents from breaking the analysis equation by equation.} Last, to facilitate the comparison of the direct and indirect approach, we treat $\mu$ and $V_\pi$ as independent on $B$, implying $p(\pi|B) = p(\pi)$. This assumption can be removed without further complications.

In Section A of the Appendix we show that, given equations (9) and (10), the joint posterior distribution

$$p(\pi,B|Y) = p(\pi|B,Y) \cdot p(B|Y),$$

satisfies

$$\pi|B,Y \sim N(\mu_\pi^*, V_\pi^*),$$

$$p(B|Y) \propto p(B) \cdot |\det(B)|^{-\frac{1}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^\frac{1}{2} \cdot e^{-\frac{1}{2}\{\tilde{y}'(I_T \otimes (BB')^{-1})\tilde{y} - \mu_\pi'V_\pi^{-1}\mu_\pi + \mu_\pi'V_\pi^*\mu_\pi\}},$$

\text{(13)}
with

\begin{align*}
V^*_{\pi} &= [V_{\pi}^{-1} + (W W' \otimes (B B')^{-1})]^{-1}, \quad (14) \\
\mu^*_{\pi} &= V^*_{\pi} \cdot \left[V_{\pi}^{-1} \mu_{\pi} + [W \otimes (B B')^{-1}] \bar{y}\right]. \quad (15)
\end{align*}

and \( \tilde{y} = \text{vec}([y_1, ..., y_t, ..., y_T]) \), \( W = [w_1, ..., w_t, ..., w_T] \). Accordingly, the analysis of the joint posterior distribution requires an MCMC algorithm for the \( k^2 \) elements in \( p(B|Y) \), or for fewer parameters in case zero (or other point) restrictions are imposed on \( B \). Conditioning on \( B \), posterior draws from \( \pi \) can be obtained with a standard random number generator.

To make the analysis feasible for more than a limited number of parameters, we explore the posterior distribution \( p(B|Y) \) using an extension of the Dynamic Striated Metropolis-Hastings sampler by Waggoner et al. (2016). The general principle is that one does not immediately sample \( p(B|Y) \), which might feature an irregular shape and multiple peaks, but a simpler function, and then progressively converts the draws from this other function into draws from \( p(B|Y) \). Define \( \theta \) the vector including the parameters of interest and the tempering function

\[ f_\lambda(\theta) = f^s(\theta) \cdot \left(f^i(\theta)\right)^\lambda \cdot \left(f^d(\theta)\right)^{1-\lambda}, \quad (16) \]

with tempering parameter \( \lambda \in [0, 1] \). The functions \( f^s(\theta) \), \( f^i(\theta) \) and \( f^d(\theta) \) are selected such that \( f_{h=0}(\theta) \) is the kernel of a distribution from which one can obtain draws easily, and \( f_{h=1}(\theta) \) coincides with the kernel of the probability distribution that one ultimately wants to sample. As \( \lambda \) increases from 0 to 1, the sampler gradually explores the function of interest using the draws at each stage to select the starting point of the Metropolis-Hastings for the next stage. For example, in sampling the posterior \( p(B|Y) \) from equation (13) with \( \theta \) containing the entries of \( B \) that are not restricted...
to zero, one could use functions

\[ f^s(\theta) = p(B), \]  
\[ f^i(\theta) = |\det(B)|^{-T} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^\frac{1}{2} \cdot e^{-\frac{1}{2} \left\{ \tilde{y} \left( I_T \otimes (BB^*)^{-1} \right) \tilde{y} - \mu_{\pi}^* V^{-1}_\pi \mu_{\pi}^* + \mu_{\pi}^* V_{\pi \pi}^{-1} \mu_{\pi} \right\} }, \]
\[ f^d(\theta) = 1. \]

Alternative specifications of \( f^s(\theta), f^i(\theta), f^d(\theta) \) are also possible as long as \( f_{h=1}(\theta) \) equals equation (13).

In a representative model with 3 variables and 13 lags, out of the total of 141 parameters, the MCMC algorithm is required for only 21 of them, well within the range in which the Dynamic Striated Metropolis-Hastings sampler performs efficiently. In Section C of the Appendix we provide a more detailed discussion of the sampler, and document the performance and running time for the applications pursued in this paper.

We conclude this section by discussing the prior distribution \( p(B) \). \( p(B) \) reflects the researcher’s prior beliefs regarding the contemporaneous effects of the shocks, and gives room for the introduction of sign restrictions on the effect of the shocks. The approach discussed in this paper is general enough to work for a wide range of prior beliefs \( p(B) \). Yet, we aim to make the approach operational to the applied researcher by proposing the following specification of \( p(B) \). Call \( b_{ij} \) the entry of \( B \) capturing the effect of a one standard deviation shock \( j \) to variable \( i \). Define two hyperparameters \( \psi_1 \) and \( \psi_2 \), and call \( \gamma_i \) a positive scalar that summarizes a realistic scale for variable \( i \). \( p(B) \) can be constructed as \( p(B) = \prod_i \prod_j p(b_{ij}) \), with \( p(b_{ij}) \) departing from untruncated normal distributions \( N(\mu_{ij}, \sigma_{ij}) \) as follows: if no sign restriction is imposed on \( b_{ij} \), set \( \mu_{ij} = 0 \) and \( \sigma_{ij} = \psi_2 \cdot \gamma_i / 1.96 \), so that the distribution is symmetric around 0 and 95% of the

---

\(^4\)Waggoner et al. (2016) apply their sampler to all structural parameters of their model, showing that the sampler is successful also when exploring jointly more than 100 parameters. A back-of-the-envelope calculation shows that we could run our approach on a model with up to 11 variables and still apply the posterior sampler on fewer parameters than in Waggoner et al. (2016).
prior mass is in the space $(-\psi_2 \gamma_i, \psi_2 \gamma_i)$; if $b_{ij}$ is restricted to be positive (or negative), depart set from a normal distribution with $\mu_{ij} = \psi_1 \gamma_i$ (or $\mu_{ij} = -\psi_1 \gamma_i$) and calibrate the variance such that the distribution truncated on the positive (or negative) support has 95% prior mass in the space $(0, \psi_2 \gamma_i)$ (or $(-\psi_2 \gamma_i, 0)$).

The convenience of the above approach is that the researcher sets a reasonable scale for the effect of the shocks by selecting $\gamma_i$, and given this scale, he controls the location and tightness of the prior using the hyperparameters $\psi_1$ and $\psi_2$. Since it can be shown that $-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}$ with $\Sigma_{ii}$ the $i$-th element of the diagonal of $\Sigma$, we set $\gamma_i = \bar{\Sigma}_{ii}^{0.5}$, where $\bar{\Sigma}$ is an estimate of $\Sigma$ based on a training sample that employs 20% of the sample, as in Primiceri (2005). $\psi_1$ and $\psi_2$ can either be set by the researcher or treated randomly within a hierarchical approach. For the latter, it suffices to replace $p(B)$ with $p(B, \psi_1, \psi_2) = p(B|\psi_1, \psi_2)p(\psi_1, \psi_2)$ in equation (13), and jointly explore the posterior $p(B, \psi_1, \psi_2|Y)$ using the same procedure discussed above (see Section A in the Appendix). We will explore both the non-hierarchical and the hierarchical structure. Within our framework, a hierarchical approach does not imply additional technical challenges, contrary to, for example, Giannone et al. (2015).

Roughly speaking, the above procedure extends to the structural parameters the procedure used within the Minnesota Prior for the variance of the reduced form parameters. With the Minnesota Prior, one selects a scale of each variable by computing the variance $\sigma_i$ of the residual on univariate AR processes on each variable, and then combines such scaling factor with a small number of hyperparameters to achieve Bayesian shrinkage. In our framework, the scaling factor is computed by exploiting the information on the covariance restrictions $\Sigma = BB'$. An alternative approach is to set $\gamma_i = \sigma_i$. Alternative specifications of $p(B)$ are also possible and can be used within our framework.

---

5Given $\Sigma = BB'$, the restrictions corresponding to the diagonal elements of $\Sigma$ are $\Sigma_{ii} = b_{i1}^2 + b_{i2}^2 + \ldots + b_{ik}^2$. Since $\Sigma_{ii}$ is nonnegative and since $b_{ij}^2 \geq 0$, each element $b_{ij}$ must satisfy $-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}$. 

---
2.3 The indirect approach used in the literature

We now summarize the indirect approach frequently used in applied work, and relate it to our approach. Working on the ‘orthogonal reduced form’ parametrization, the indirect approach departs from prior distributions on \((\pi, \Sigma, Q)\) rather than \((\pi, B)\). To make the analysis comparable to the direct approach discussed in Section 2.2, we do not restrict the indirect approach to the more tractable case of conjugate priors, but consider the independent Normal-Inverse-Wishart prior specification.

As already known in the literature and as summarized in Section B of the Appendix, the prior beliefs

\[
\begin{align*}
\pi & \sim N(\mu_\pi, V_\pi), \\
\Sigma & \sim iW(\nu, \Phi), \\
Q & \sim U,
\end{align*}
\]

(20a) (20b) (20c)

with \(p(\pi, \Sigma) = p(\pi)p(\Sigma)\) lead to a joint posterior distribution \(p(\pi, \Sigma|Y)\) that can be explored using a Gibbs sampler, given the conditional posterior distributions

\[
\begin{align*}
\pi|Y, \Sigma & \sim N(\mu^*_\pi, V^*_\pi), \\
\Sigma|Y, \Pi & \sim iW(\nu^*, \Phi^*),
\end{align*}
\]

(21) (22)

with

\[
\begin{align*}
V^*_\pi &= [V^{-1}_\pi + (WW' \otimes \Sigma^{-1})]^{-1}, \\
\mu^*_\pi &= V^*_\pi \cdot [V^{-1}_\pi \mu_\pi + (W \otimes \Sigma^{-1})\hat{y}], \\
d^* &= d + T, \\
S^* &= S + (Y - \Pi X)(Y - \Pi X)',
\end{align*}
\]

(23) (24) (25) (26)

The corresponding marginal prior and posterior distribution for \(B\) can then be explored.
numerically by drawing $\Sigma$ from $p(\Sigma)$ or $p(\Sigma|Y)$, drawing an orthogonal matrix $Q$ from the algorithm by Rubio-Ramirez et al. (2010), computing the candidate matrix $B = h(\Sigma)Q$ and keeping the draw if sign restrictions are satisfied.\textsuperscript{6} Note that we use the same notation $V^*_\pi$ for the direct and the indirect approach, under the understanding that $\Sigma = BB'$ implies equivalence between equation (14) and equation (23). The same holds for $\mu^*_\pi$.

Arias et al. (forthcoming) work within the conjugate Normal-Inverse-Wishart prior (which imposes the additional restriction $V_\pi = V \otimes \Sigma$) and compute the distribution that the joint prior $p(\mu_\pi, \Sigma, Q)$ implies for matrix $A$ given model (4). They consider two separate cases, depending on whether zero restrictions are introduced or not. Building on their contribution, in Section B of the Appendix we show that the prior distribution (20a)-(20c) imply the following prior and posterior marginal distributions for $B$,

\begin{equation}
\begin{align*}
p^i(B) &\propto I\{\text{sign}\} \cdot |\text{det}(B)|^{-(d+k)} \cdot e^{-\frac{1}{2} \left\{ \text{vec}(B^{-1})'(S \otimes I_k)\text{vec}(B^{-1}) \right\}}, \\
p^i(B|Y) &\propto I\{\text{sign}\} \cdot |\text{det}(B)|^{-(d+k+T)} \cdot e^{-\frac{1}{2} \left\{ \text{vec}(B^{-1})'(S \otimes I_k)\text{vec}(B^{-1})+\tilde{y}' \left( I_T \otimes (BB')^{-1} \right)\tilde{y} - \mu^*_\pi V^{-1}_\pi \mu^*_\pi \right\}},
\end{align*}
\end{equation}

with $I\{\text{sign}\}$ an indicator functions equal to unity if $B$ satisfies the sign restrictions. For simplicity, we do not consider zero restrictions too.

Equation (27a) allows appreciating how the direct approach from Section 2.2 provides additional flexibility on the identifying restrictions used in the analysis. While the direct approach specifies a generic prior $p(B)$, the indirect approach restricts prior beliefs on $B$ to equation (27a), and provides flexibility on those only up to what granted by the selection of the hyperparameters $d, S$. By construction, the indirect approach provides a special case of the direct approach.

\textsuperscript{6}To ensure a higher efficiency of the algorithm, the sign restrictions are assessed for each candidate matrix $B$ after considering all possible permutations of its columns, and taking into account that shocks are identified only up to sign of the shock.
It remains to discuss the selection of the hyperparameters \( d, S \) for the indirect approach. We follow Arias et al. (forthcoming), Uhlig (2005) and several others and set \( d = 0, S = 0 \). This implies that the implicit prior and posterior distributions for \( B \) are given by

\[
p^i(B) \propto I\{\text{sign}\} \cdot |\text{det}(B)|^{-(d+k)},
\]

\[
p^i(B|Y) \propto I\{\text{sign}\} \cdot |\text{det}(B)|^{-(d+k+T)} \cdot e^{-\frac{1}{2} \left\{ + \bar{y}' \left( I_T \otimes (BB')^{-1} \right) \bar{y} - \mu_y^* V_y^{-1} \mu_y^* \right\}}.
\]

\[ (28) \]

\[ (29) \]

### 3 An illustrative example based on data simulated from the model by An and Schorfheide (2007)

We first apply the methodology discussed in Section 2.1 to a simplified scenario in which the data generating process admits a VAR representation, and in which the model is estimated on all the relevant variables. We use this scenario, for which we simulate from the model by An and Schorfheide (2007), as an ideal setting in which to highlight the intuition behind the approach proposed in the paper.

#### 3.1 The data generating process and the model

For the first data generating process we use the linearized DSGE model by An and Schorfheide (2007), which is given by the following set of equations:

\[
x_t = E_t x_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau} (r_t - E_t \pi_{t+1} - E_t z_{t+1}),
\]

\[ (30a) \]

\[
\pi_t = \beta E_t \pi_{t+1} + \kappa(x_t - g_t),
\]

\[ (30b) \]

\[
r_t = \rho_r r_{t-1} + (1 - \rho_r) \eta_1 \pi_t + (1 - \rho_r) \eta_2 (x_t - g_t) + \epsilon_{rt},
\]

\[ (30c) \]

\[
g_t = \rho_g g_{t-1} + \epsilon_{gt},
\]

\[ (30d) \]

\[
z_t = \rho_z z_{t-1} + \epsilon_{zt}.
\]

\[ (30e) \]
The variables of the model are the output gap \( x_t \), inflation \( \pi_t \) and the interest rate \( r_t \), which are driven by the technology shock \( \epsilon_{zt} \sim N(0, \sigma_z^2) \) (which drives the evolution of productivity \( z_t \)), the government spending shock \( \epsilon_{gt} \sim N(0, \sigma_g^2) \) (which drives the evolution of government spending \( g_t \)), and an interest rate shock \( \epsilon_{rt} \sim N(0, \sigma_r^2) \).

Equations (30a) to (30e) summarize a model economy of a representative household, perfectly competitive intermediate good producers, a final good producer, a fiscal authority, and the central bank. Households gain utility from consumption and real money balances, and disutility from labour. They supply labour on competitive markets to intermediate good producers, who face a production function subject to technology shocks and adjust prices after incurring a cost. The fiscal authority consumes a stochastic fraction of final output, raises lump-sum taxes and issues bonds. The interest rate is set by the central bank according to a Taylor rule subject to monetary policy shocks.

The parameters of the model are \( \iota = (\tau, r^A, \kappa, \rho_r, \rho_g, \rho_z, \psi_1, \psi_2, \sigma_r, \sigma_g, \sigma_z)' \), with \( \beta = \frac{1}{1+r^A/400} \).\(^7\) We calibrate \( \iota \) using the parameter values that An and Schorfheide (2007) employ for their data generating process, as summarized in Table E1 in the Appendix. We then use the solution method by Sims (2002) to solve the model and the factorization by Fernandez-Villaverde et al. (2007) to compute the associated VAR representation. Last, we rewrite the model such that the structural shocks have identity covariance matrix.\(^8\) The DGP has the following exact VAR(1) form,

\[
\begin{bmatrix}
  r_t \\
  x_t \\
  \pi_t \\
  y_t
\end{bmatrix}
= \begin{bmatrix}
  0.6048 & 0 & 0.9017 \\
  0.0616 & 0.95 & -1.9302 \\
  0.0103 & 0 & 0.4452 \\
  \Pi(\iota)
\end{bmatrix}
\begin{bmatrix}
  r_{t-1} \\
  x_{t-1} \\
  \pi_{t-1} \\
  y_{t-1}
\end{bmatrix}
+ \begin{bmatrix}
  0.0160 & 0 & 0.0298 \\
  0.0281 & 0.0894 & -0.0271 \\
  0.0079 & 0 & -0.0068 \\
  B(\iota)
\end{bmatrix}
\begin{bmatrix}
  \epsilon_{zt} \\
  \epsilon_{gt} \\
  \epsilon_{rt}
\end{bmatrix}
\]

\[(31)\]

\(^7\)To calibrate the model as in An and Schorfheide (2007), we treat \( \beta \) as a function of the fundamental parameter \( r^A \), which determines the steady state interest rate.

\(^8\)Given the solution in \( y_t = \Pi(\iota) y_{t-1} + B(\iota) \epsilon_t \) with \( \epsilon_t \sim N(0, D(\iota)) \), we compute \( B(\iota) = \bar{B}(\iota) D(\iota)^{-0.5} \) and generate the data from \( y_t = \Pi(\iota) y_{t-1} + B(\iota) \epsilon_t \) with \( \epsilon_t \sim N(0, I_k) \).
with $\epsilon_t \sim N(0, I_3)$, while the reduced form covariance matrix equals

$$
\Sigma(\iota) = B(\iota)B(\iota)' = \begin{pmatrix}
0.0011 & -0.004 & -0.0001 \\
-0.0004 & 0.0095 & 0.0004 \\
-0.0001 & 0.0004 & 0.0001
\end{pmatrix}.
$$

(32)

On impact, a technology shock of positive size increase the interest rate, output and inflation, a monetary shock of positive sign increases the interest rate and decreases output and inflation, while a government spending shock only increases spending. The data generating process from the previous section is used to draw 80 observations, as in An and Schorfheide (2007).

We use the generated data to estimate a SVAR model identified with sign restrictions. We include a constant and 4 lags of the dependent variables in the model, and hence do not employ the information that the true model features no constant and only one lag. We normalize the covariance matrix of the structural shocks to the identity matrix.

### 3.2 Update of prior beliefs - non-hierarchical approach

As a prior for $\pi$, the reduced form autoregressive parameters, we use the prior distribution from equation (10), setting $\mu_\pi$ and $V_\pi$ as applied by Litterman (1986) for the Minnesota Prior. We set $\mu_\pi$ to imply that each variable follows a white noise process, in accordance with the nontrending nature of the pseudo data generated. We then set $V_\pi$ as discussed in Canova (2007), allowing each lagged coefficient to have different variance depending on whether it enters the equation of the variable itself or of another variable. Since this prior on $\pi$ is well established in the literature, we refer the reader to Bańbura et al. (2010), Canova (2007), Koop et al. (2010) and Kilian and Lütkepohl (2017) for a discussion.

Last, we specify the prior beliefs on $B$ as explained in Section 2.2. We distinguish
two cases. In the first case, considered in this section, we parametrize the hyperparameters $\psi_1$ and $\psi_2$ to arbitrary values, which we set here equal to 0.8 and 1.2, respectively. In the second discussed in the next section, we employ a hierarchical approach and set a prior distribution on $\psi_1$ and $\psi_2$.

Figure 1 shows the update on prior beliefs on $B$. The line ports the prior beliefs, the white histogram the distribution we use to initialize the posterior sampling, and the blue histogram the posterior draws of $B$. The figure also reports a variety of statistics to assess the scaling of the volatility of the variables. As shown, the posterior distribution becomes tighter, and includes the true values of $B$, which is indicated by
the blue dot.

Last, Figure 2 shows the corresponding impulse responses. To help develop the intuition, we report the confidence band corresponding to prior beliefs, as well as to the distribution used to initialize the sampler. The posterior distribution, reported in red, tends to include the true impulse response, especially for inflation.
3.3 Update of prior beliefs - hierarchical approach

In this section we show the same exercise when treating the hyperparameters $\psi_1$ and $\psi_2$ hierarchically. We specify a gamma prior distribution on them, setting their expected value equal to the calibrated parameters from the previous distribution. The goal of the exercise is to assess the robustness of the results for a wider set of values of the hyperparameters.

Figure 3 and Figure 4 show the results for the case of a hierarchical approach. As should be expected, the prior beliefs become wider. The posterior distribution still
includes the true values. The results on impulse responses vary only a little.

4 An application to monetary policy and financial shocks

We now show an application to US data.

The 2007 financial crisis shows how far reaching the effects of financial shocks can be on the real economy. While monetary and fiscal authorities do control powerful
economic tools that can be used to mitigate the effects of financial shocks, it remains unknown whether such tools are powerful enough to offset the effect of financial shocks.

We apply the methodology discussed in this paper to study to what extent the effects of monetary and financial shocks differ with regard to their impact on output. The stronger the effect of financial shocks relative to monetary policy shocks, the less powerful the monetary authority is to counteract such financial shocks.

We estimate a structural VAR model on the following 4 variables: the federal funds rate, the log of CPI, the log of industrial production, and the S&P 500 stock market index. We estimate the model using data from 1987M1 until 2007M1. We add 4 lags to the model and we specify the hyperparameters of the prior as in the application to simulated data.

For the identification of the model, we apply the following sign restrictions. For the monetary shock, we impose that a monetary expansion increases CPI, industrial production and the S&P 500. For the financial shock, we impose that an adverse financial shock increases the S&P 500, decreases output and CPI, and does not lead to a tightening of monetary policy.

The results of the analysis are shown in Figure 5 and Figure 6. Prior beliefs are update as shown in the figure. Impulse responses document that the monetary policy shock affects output by an amount similar to the financial shock. This suggests that the monetary authority exerts similar pressure on output relative to standard financial shocks.

5 Conclusions

Structural Vector Autoregressive models are frequently identifies using sign restrictions on the impulse response of selected structural shocks of interest. However, it is not clear how this identification approach should be implemented in practice.

On the one hand, it is convenient to depart from a specification on reduced form
parameters, as this makes posterior sampling highly tractable. However, this approach implies very limited flexibility on the structural parameters of interest. On the other hand, one can depart from prior specification on the moving average representation of the data. While providing maximum flexibility on the prior on the statistics of interest, this alternative approach is technically more challenging.

We propose a compromise between the above two extreme approaches. We propose to parametrize the model such that flexibility on prior specification is needed the most, namely on the specification of prior beliefs on the contemporaneous impulse responses. The remaining parameters are set such that posterior sampling is tractable.
We apply this approach to data simulated from the An and Schorfheide (2007) model, in order to provide intuition behind the method proposed. We then show an application to monetary policy and financial shocks. We show that monetary policy and financial shocks of comparable size exert similar effects on the real economy. This suggests that monetary authorities do have tools that potentially offset the effect of financial shocks.
References


Baumeister, C. J. and Hamilton, J. D. (forthcoming), ‘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*.


