

A Gibbs Sampler for Efficient Bayesian Inference in Sign- Identified SVARs

Jonas E. Arias

Federal Reserve Bank of Philadelphia

Juan Rubio-Ramírez

Emory University, Federal Reserve Bank of Atlanta, and Visiting Scholar,
Federal Reserve Bank of Philadelphia Research Department

Minchul Shin

Federal Reserve Bank of Philadelphia

WP 25-19

PUBLISHED

May 2025

ISSN: 1962-5361

Disclaimer: This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in these papers are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Philadelphia Fed working papers are free to download at: <https://www.philadelphiafed.org/search-results/all-work?searchtype=working-papers>.

DOI: <https://doi.org/10.21799/frbp.wp.2025.19>

A Gibbs Sampler for Efficient Bayesian Inference in Sign-Identified SVARs

Jonas E. Arias

Juan Rubio-Ramírez*

Federal Reserve Bank of Philadelphia

Emory University

Federal Reserve Bank of Atlanta

Minchul Shin

Federal Reserve Bank of Philadelphia

May 30, 2025

Abstract

We develop a new algorithm for inference based on structural vector autoregressions (SVARs) identified with sign restrictions. The key insight of our algorithm is to break from the accept-reject tradition associated with sign-identified SVARs. We show that embedding an elliptical slice sampling within a Gibbs sampler approach can deliver dramatic gains in speed and turn previously infeasible applications into feasible ones. We provide a tractable example to illustrate the power of the elliptical slice sampling applied to sign-identified SVARs. We demonstrate the usefulness of our algorithm by applying it to a well-known small SVAR model of the oil market featuring a tight identified set, as well as to a large SVAR model with more than 100 sign restrictions.

JEL classification: C11, C15, C32.

Keywords: large structural vector autoregressions, sign restrictions, slice elliptical sampling.

We thank Frank Schorfheide for helpful comments. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

*Corresponding author: Juan F. Rubio-Ramírez <juan.rubio-ramirez@emory.edu>, Economics Department, Emory University, Rich Memorial Building, Room 306, Atlanta, Georgia 30322-2240.

1 Introduction

The conventional method of Bayesian inference in SVARs identified with sign restrictions—proposed by [Faust \(1998\)](#); [Canova and De Nicoló \(2002\)](#); [Uhlig \(2005\)](#) and extended by [Rubio-Ramírez, Waggoner, and Zha \(2010\)](#)—is one of the most popular tools available for assessing the dynamic causal effects of structural shocks in macroeconomics. To date, thousands of studies have adopted this standard accept-reject algorithm. Despite recent advances—[Chan, Matthes, and Yu \(2025\)](#) have proposed an improved accept-reject algorithm by exploiting permutations and sign alternation of the standard draw from the uniform prior density over the set of orthogonal matrices—the big data era has strained the capacity of this class of algorithms to the point that in some cases it becomes infeasible as identified sets become tight. Similar challenges arise even in lower-dimensional SVARs when incorporating rank restrictions, such as elasticity bounds, as discussed in [Kilian and Murphy \(2012\)](#); [Amir-Ahmadi and Drautzburg \(2021\)](#); these ranking restrictions can meaningfully sharpen inference at the cost of tight identified sets. The adoption of large-scale SVARs has become increasingly common in macroeconomic research following the foundational work of [Leeper, Sims, and Zha \(1996\)](#), who introduced medium-sized systems to analyze monetary policy shocks. This shift toward higher-dimensional models was significantly influenced by [Bańbura, Giannone, and Reichlin \(2010\)](#), who highlighted the advantages of including a broad set of variables for improving both predictive accuracy and structural interpretation. Subsequent contributions have demonstrated the practicality and value of large SVARs in various contexts, including the studies by [Carriero, Kapetanios, and Marcellino \(2009\)](#); [Koop \(2013\)](#); [Ellahie and Ricco \(2017\)](#); [Crump, Eusepi, Giannone, Qian, and Sbordone \(2021, 2025\)](#), which illustrate how such models can be effectively utilized in empirical macroeconomic analysis.

In this paper, we break with the accept-reject tradition and show that embedding an elliptical slice sampling within a Gibbs sampler approach can deliver dramatic gains in speed and turn previously infeasible applications into feasible ones. The objective is to obtain draws from the posterior distribution of the orthogonal reduced-form parameters conditional on the sign restrictions. To accomplish such a goal, the conventional approach draws from the posterior distribution of the orthogonal reduced-form parameters and uses an

accept-reject algorithm to discard the draws that do not satisfy the restrictions. Consequently, as more restrictions are incorporated to sharpen inference, this approach is destined to fail. The innovations in [Chan, Matthes, and Yu \(2025\)](#) will certainly expand the set of models that researchers can consider. Notwithstanding, it will eventually fail as the sets become even tighter. In contrast, we use a Gibbs sampler to iteratively draw from the posterior distributions conditional on the sign restrictions, making the accept-reject step unnecessary.

We use an example similar to the one used in [Granziera, Moon, and Schorfheide \(2018\)](#) in order to illustrate that the efficiency of the accept-reject algorithm hinges heavily on the size of the identified set. As the identified set becomes tighter, the accept-reject algorithm is destined to slow down significantly. In contrast, our Gibbs sampler exploits an elliptical slice sampling that exponentially shrinks candidate impulse responses to the identified set.

We illustrate our algorithm with two applications. In the first application, we replicate [Kilian and Murphy \(2014\)](#), a model of the world oil market where the standard application of the accept-reject algorithm fails. To get around this infeasibility, [Kilian and Murphy \(2014\)](#) consider an approach similar to the one in [Chan, Matthes, and Yu \(2025\)](#) by exploiting permutations and sign alternation. As we will show below, our algorithm can efficiently handle this application. In the second application, we re-visit the structural analysis in [Chan, Matthes, and Yu \(2025\)](#) who use [Crump et al.’s \(2025\)](#) large SVAR model of the U.S. economy to identify 8 structural shocks. We show that as we augment the number of shocks under analysis, the efficient accept-reject proposed by [Crump, Eusepi, Giannone, Qian, and Sbordone \(2025\)](#) slows down noticeably until it becomes impractical. In contrast, the computational time of our Gibbs sampler is largely insensitive of the number of identified structural shocks.

Finally, we have to highlight [Read and Zhu \(2025\)](#). This paper is contemporaneous to ours and it proposes an algorithm based on the slice sampling of [Neal \(2003\)](#) that is more efficient than the accept reject sampling. However, their approach relies on an approximation argument and it is restricted to using the conditionally uniform prior described in [Del Negro and Schorfheide \(2010\)](#); [Uhlig \(2017\)](#); [Amir-Ahmadi, Matthes, and Wang \(2020\)](#).

2 The Model

Consider the SVAR with the general form,

$$\mathbf{y}_t' \mathbf{A}_0 = \mathbf{x}_t' \mathbf{A}_+ + \boldsymbol{\varepsilon}_t' \text{ for } 1 \leq t \leq T, \quad (1)$$

where $\mathbf{A}_+ = [\mathbf{A}_1' \cdots \mathbf{A}_p' \mathbf{c}']$ and $\mathbf{x}_t' = [\mathbf{y}_{t-1}' \cdots \mathbf{y}_{t-p}' \mathbf{1}]$ for $1 \leq t \leq T$ and where \mathbf{y}_t is an $n \times 1$ vector of endogenous variables, $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ vector of exogenous structural shocks, \mathbf{A}_ℓ is an $n \times n$ matrix of parameters for $0 \leq \ell \leq p$ with \mathbf{A}_0 invertible, \mathbf{c} is a $1 \times n$ vector of parameters, p is the lag length, and T is the sample size. Hence, the dimension of \mathbf{A}_+ is $m \times n$, where $m = np + 1$. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \dots, \mathbf{y}_{1-p}$, is Gaussian with mean zero and covariance matrix \mathbf{I}_n , the $n \times n$ identity matrix. The reduced-form representation implied by Equation (1) is

$$\mathbf{y}_t' = \mathbf{x}_t' \mathbf{B} + \mathbf{u}_t' \text{ for } 1 \leq t \leq T, \quad (2)$$

where $\mathbf{B} = \mathbf{A}_+ \mathbf{A}_0^{-1}$, $\mathbf{u}_t' = \boldsymbol{\varepsilon}_t' \mathbf{A}_0^{-1}$, and $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \boldsymbol{\Sigma} = (\mathbf{A}_0 \mathbf{A}_0')^{-1}$. The matrices \mathbf{B} and $\boldsymbol{\Sigma}$ are the reduced-form parameters, while \mathbf{A}_0 and \mathbf{A}_+ are the structural parameters. While \mathbf{B} is a $m \times n$ matrix, $\boldsymbol{\Sigma}$ belongs to the set $\mathcal{S}(n)$, which is the set of $n \times n$ positive definite matrices.

It is well-known that, for the linear Gaussian models of the type studied in this paper, $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if they have the same reduced-form representation. This implies that the structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if $\mathbf{A}_0 = \tilde{\mathbf{A}}_0 \mathbf{Q}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+ \mathbf{Q}$ for some $\mathbf{Q} \in \mathcal{O}(n)$, which is the set of all $n \times n$ orthogonal matrices.

To solve the identification problem, one often imposes sign restrictions on either the structural parameters or some function of the structural parameters, like the impulse responses. To simplify the notation, we will summarize the sign restrictions by $\mathbf{S}(\mathbf{A}_0, \mathbf{A}_+) > 0$ and we let $[\mathbf{S}(\mathbf{A}_0, \mathbf{A}_+) > 0]$ be an indicator function that equals 1 if the sign restrictions are satisfied and 0 otherwise.

2.1 The Orthogonal Reduced-Form Parameterization

Equation (1) represents the SVAR in terms of the structural parameterization, which is characterized by \mathbf{A}_0 and \mathbf{A}_+ . The SVAR can alternatively be written in what we call the orthogonal reduced-form parameterization, see [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#). This parameterization is characterized by the reduced-form parameters \mathbf{B} and $\mathbf{\Sigma}$ together with an orthogonal matrix \mathbf{Q} and is given by the following equation

$$\mathbf{y}'_t = \mathbf{x}'_t \mathbf{B} + \boldsymbol{\varepsilon}'_t \mathbf{Q}' h(\mathbf{\Sigma}) \text{ for } 1 \leq t \leq T, \quad (3)$$

where the $n \times n$ matrix $h(\mathbf{\Sigma})$ is any decomposition of the covariance matrix $\mathbf{\Sigma}$ satisfying $h(\mathbf{\Sigma})' h(\mathbf{\Sigma}) = \mathbf{\Sigma}$. We will take h to be the Cholesky decomposition, though any differentiable decomposition would do. Given Equations (1) and (3), we can define a mapping between $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ and $(\mathbf{A}_0, \mathbf{A}_+)$ by

$$f(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = (\underbrace{h(\mathbf{\Sigma})^{-1} \mathbf{Q}}_{\mathbf{A}_0}, \underbrace{\mathbf{B} h(\mathbf{\Sigma})^{-1} \mathbf{Q}}_{\mathbf{A}_+}).$$

This mapping makes clear how the structural parameters depend on the reduced-form parameters and orthogonal matrices. Given the reduced-form parameters, one can consider each value of $\mathbf{Q} \in \mathcal{O}(n)$ as a particular choice among observationally equivalent structural parameters.

We can also define the impulse responses. Let $\mathbf{u}_t = \mathbf{L}_0 \boldsymbol{\varepsilon}_t$ for $1 \leq t \leq T$, where \mathbf{L}_0 is an $n \times n$ invertible matrix that represents impulse responses at horizon zero. Given \mathbf{L}_0 and \mathbf{B} , it is possible to obtain the impulse responses beyond horizon zero recursively, as

$$\mathbf{L}_\ell = \sum_{k=1}^{\min\{\ell, p\}} \mathbf{B}'_k \mathbf{L}_{\ell-k}, \text{ for } \ell > 0. \quad (4)$$

We combine the impulse responses from horizons one through p and the constant term \mathbf{c} into a single matrix, $\mathbf{L}_+ = [\mathbf{L}'_1 \cdots \mathbf{L}'_p \ \mathbf{c}']'$, where the maximum horizon of the impulse response in \mathbf{L}_+ is exactly the same as the lag length in Equation (1). We call $(\mathbf{L}_0, \mathbf{L}_+)$ the impulse

response parameters.

Given the function f and Equation (4), we can also define a mapping from $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ to $(\mathbf{L}_0, \mathbf{L}_+)$ by

$$\phi(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \left(\underbrace{h(\boldsymbol{\Sigma})' \mathbf{Q}}_{\mathbf{L}_0}, \underbrace{[\mathbf{L}_1(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})' \cdots \mathbf{L}_p(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})' \quad \mathbf{Q}'(h(\boldsymbol{\Sigma})^{-1})' \mathbf{d}']'}_{\mathbf{L}_+} \right), \quad (5)$$

where $\mathbf{L}_\ell(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ for $1 \leq \ell \leq p$ is implicitly defined in Equation (4). The functions f and ϕ are invertible, and f , ϕ and their inverses are differentiable. The results in this section show that, regardless of where we place the sign restrictions, we can always write the sign restrictions in terms of the orthogonal reduced-form parameterization. Hence, let $[\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0]$ be an indicator function in terms of the orthogonal reduced-form parameterization that equals 1 if the sign restrictions are satisfied and 0 otherwise.

2.2 Conjugate Priors and Posteriors

We will illustrate them using conjugate distributions. For the reduced-form representation in Equation (2), the normal-inverse-Wishart family of distributions is conjugate. A normal-inverse-Wishart distribution over the reduced-form parameters is characterized by four parameters: a scalar $\nu \geq n$, an $n \times n$ symmetric and positive definite matrix $\boldsymbol{\Phi}$, an $m \times n$ matrix $\boldsymbol{\Psi}$, and an $m \times m$ symmetric and positive definite matrix $\boldsymbol{\Omega}$. We denote this distribution by $NIW(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ and its density by $NIW_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{B}, \boldsymbol{\Sigma})$. Furthermore,

$$NIW_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{B}, \boldsymbol{\Sigma}) \propto \underbrace{|\det(\boldsymbol{\Sigma})|^{-\frac{\nu+n+1}{2}} e^{-\frac{1}{2} \text{tr}(\boldsymbol{\Phi} \boldsymbol{\Sigma}^{-1})}}_{\text{inverse-Wishart}} \underbrace{|\det(\boldsymbol{\Sigma})|^{-\frac{m}{2}} e^{-\frac{1}{2} \text{vec}(\mathbf{B} - \boldsymbol{\Psi})'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})^{-1} \text{vec}(\mathbf{B} - \boldsymbol{\Psi})}}_{\text{conditionally normal}}.$$

If the prior distribution over the reduced-form parameters is $NIW(\bar{\nu}, \bar{\boldsymbol{\Phi}}, \bar{\boldsymbol{\Psi}}, \bar{\boldsymbol{\Omega}})$, then the posterior distribution over the reduced-form parameters is $NIW(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, where

$$\begin{aligned} \tilde{\nu} &= T + \bar{\nu}, \\ \tilde{\boldsymbol{\Omega}} &= (\mathbf{X}'\mathbf{X} + \bar{\boldsymbol{\Omega}}^{-1})^{-1}, \end{aligned}$$

$$\begin{aligned}\tilde{\Psi} &= \tilde{\Omega}(\mathbf{X}'\mathbf{Y} + \bar{\Omega}^{-1}\bar{\Psi}), \\ \tilde{\Phi} &= \mathbf{Y}'\mathbf{Y} + \bar{\Phi} + \bar{\Psi}'\bar{\Omega}^{-1}\bar{\Psi} - \tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\Psi},\end{aligned}$$

for $\mathbf{Y} = [\mathbf{y}_1 \ \cdots \ \mathbf{y}_T]'$ and $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_T]'$.

The conjugate normal-inverse-Wishart prior is widely used in Bayesian VARs due to its computational convenience (see [Faust, 1998](#); [Uhlig, 2005](#); [Rubio-Ramírez, Waggoner, and Zha, 2010](#); [Kilian and Murphy, 2012, 2014](#); [Arias, Rubio-Ramirez, and Waggoner, 2025](#)). When combined with the conventional accept-reject approach, it produces independent draws from the posterior, which makes it especially attractive for simulation-based methods. However, it also imposes a Kronecker structure on the prior distribution of \mathbf{B} , thereby constraining its covariance matrix. This structure induces a dependence, within each equation, between the variance of the residuals and the variance of the VAR coefficients—an assumption that may be undesirable in many empirical applications.

An alternative is the independent normal-inverse-Wishart prior, which avoids this covariance structure and allows for greater flexibility. However, implementing this prior requires a Gibbs sampler, as it does not yield conditionally independent draws. Since our algorithms rely on a Gibbs sampler, the computational advantage of the conjugate prior vanishes in our context. Nonetheless, we will present the methodology using the conjugate normal-inverse-Wishart prior, given its prevalence in the literature, but our algorithms could also be applied to the independent normal-inverse-Wishart prior.

We will take $\pi(\mathbf{Q}|\mathbf{B}, \mathbf{\Sigma})$ to be the uniform density. This choice can be motivated by the fact that it assigns equal prior weight to observationally equivalent models or vectors of impulse responses (see [Arias, Rubio-Ramirez, and Waggoner, 2025](#)). We call this the uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization; denote it by $UNIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$, and denote its density over the orthogonal reduced-form parameterization by $UNIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$. It is the case that

$$UNIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \frac{NIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma})}{\int_{\mathcal{O}(n)} 1 d\mathcal{O}(n) \mathbf{Q}}.$$

Because $\mathcal{O}(n)$ is compact, $\int_{\mathcal{O}(n)} 1 d_{\mathcal{O}(n)} \mathbf{Q}$ is finite.

2.3 Inference Based on Sign Restrictions

Given results in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#), our objective will be to draw from the posterior of the orthogonal reduced-form parameters conditional on the sign restrictions

$$p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q} \mid (\mathbf{y}_t)_{t=1}^T, \mathcal{S}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) > 0) = \frac{[\mathcal{S}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) > 0] \text{UNIW}(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})}{\text{Pr}(\mathcal{S}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) > 0 \mid (\mathbf{y}_t)_{t=1}^T)} \quad (6)$$

and then use f and ϕ to transform the draws to the desired vector of objects of interests such as the structural parameters or impulse responses. The traditional approach to obtain draws from Equation (6) uses the following accept-reject algorithm:

Algorithm 1. *The following algorithm independently draws from the $\text{UNIW}(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$ distribution over the structural parameterization conditional on the sign restrictions.*

1. Draw $(\mathbf{B}, \mathbf{\Sigma})$ independently from the $\text{NIW}(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$ distribution.
2. Draw \mathbf{Q} independently from the uniform distribution over $\mathcal{O}(n)$.
3. Keep $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ if $[\mathcal{S}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) > 0] = 1$.
4. Return to Step 1 until the required number of draws has been obtained.

While this algorithm has been widely adopted, it is well known that there are cases in which the identified set is narrow limiting the efficiency of the algorithm (see e.g. [Kilian and Murphy, 2014](#); [Baumeister and Hamilton, 2024](#); [Chan, Matthes, and Yu, 2025](#); [Read and Zhu, 2025](#)). In the next section, we use a simple example to show its shortcomings. We will also show how a carefully designed ESS algorithm is not subject to this limitation and delivers dramatic speed gains. Importantly, [Chan, Matthes, and Yu \(2025\)](#) show a new numerically efficient version of Algorithm 1 that facilitates the drawing for a large number of structural restrictions. When comparing our algorithm to the traditional accept-reject approach we will use this efficient version as the benchmark.

3 The Problem with Accept-Reject Sampling

For our purposes, it suffices to work with a simple example similar to the one explored by [Granziera, Moon, and Schorfheide \(2018\)](#). Thus, consider the following SVAR, with $n = 2$ and $m = 0$, written under the orthogonal reduced-form parameterization:

$$\mathbf{y}'_t = (y_{t,1}, y_{t,2}) = \boldsymbol{\varepsilon}'_t(\boldsymbol{\Sigma}_{tr}\mathbf{Q})',$$

where we let $\boldsymbol{\Sigma}_{tr} = h(\boldsymbol{\Sigma})'$. Initially, we assume $\boldsymbol{\Sigma}_{tr}$ is known, but later we will relax this assumption. Let $\sigma_{tr,ij}$ denote the i -th row and j -th column entry of $\boldsymbol{\Sigma}_{tr}$. For simplicity, we set $\sigma_{tr,11} = \sigma_{tr,22} = 1$ and $\sigma_{tr,21} = -0.9$. Note that the contemporaneous impact matrix \mathbf{L}_0 is defined as $\mathbf{L}_0 = \boldsymbol{\Sigma}_{tr}\mathbf{Q}$. Henceforward, we focus on the impulse responses to the first shock—it is straightforward to extend our analysis to the second shock.

Given the above, it is easy to see that the impact of the first shock on $y_{t,1}$ and $y_{t,2}$ can be written as $\ell_{11} = q_{11}$ and $\ell_{21} = -0.9q_{11} + q_{21}$, where ℓ_{i1} and q_{i1} are the $(i, 1)$ entries of \mathbf{L}_0 and \mathbf{Q} , respectively. Let us now impose sign restrictions requiring that ℓ_{11} and ℓ_{21} are nonnegative. The imposed sign restrictions imply $q_{11} \geq 0$ and $q_{21} \geq 0.9q_{11}$. Figure [1a](#) illustrates this setup graphically. The green circle represents the domain of $\mathbf{q}_1 = (q_{11}, q_{21})'$, while the red arc highlights the identified set that satisfies the imposed sign restrictions.

When using the popular accept-reject sampling approach described in Algorithm [1](#), obtaining a draw from the posterior distribution of impulse responses of interest satisfying the sign restrictions involves drawing vector a 2×1 vector \mathbf{x}_1 from a $N(\mathbf{0}, \mathbf{I}_2)$ distribution and converting it into a unit vector \mathbf{q}_1 via the following normalization: $\mathbf{q}_1 = \mathbf{x}_1 / \|\mathbf{x}_1\|$. The draw is accepted only if \mathbf{q}_1 satisfies the sign restrictions. Unrestricted draws $(q_{11}, q_{21})'$ lie uniformly on the entire unit circle (depicted in green), whereas the accepted draws are uniformly distributed only over the subset of the unit circle that meets the sign restrictions (the red arc).

The efficiency of the posterior simulator based on this type of acceptance-rejection algorithm depends heavily on the size of the identified set. As the identified set becomes tighter, we naturally expect to discard a larger number of draws. Indeed, the expected

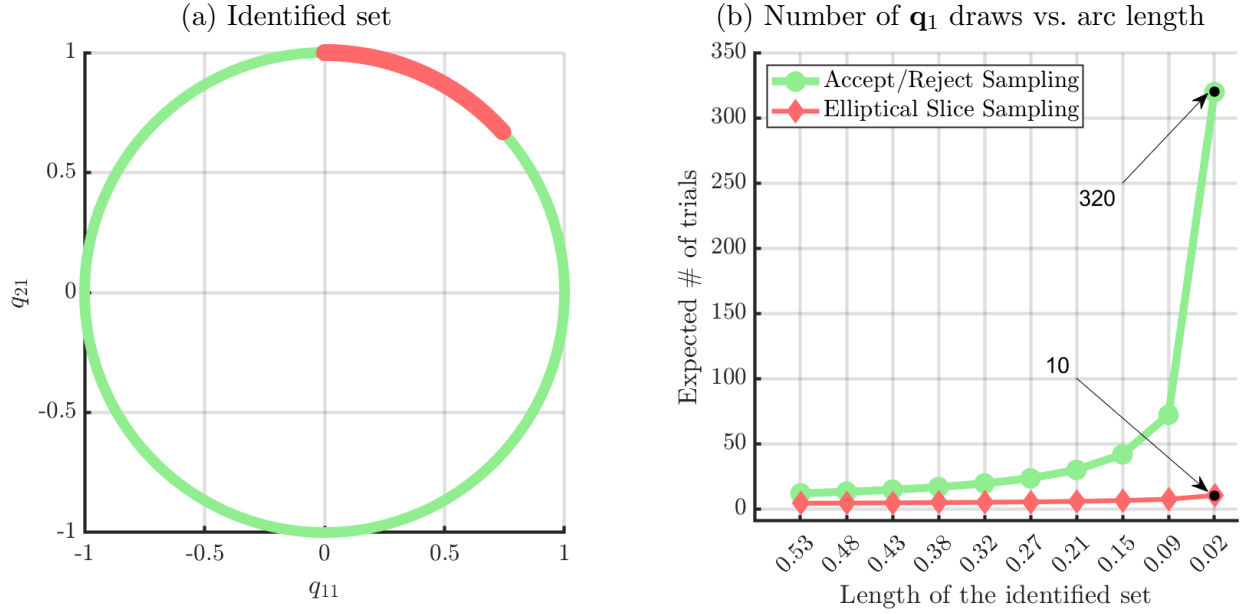


Figure 1: (a) Identified set (Red) and domain of $(q_{11}, q_{21})'$ (Green). (b) Expected number of draws required to meet the sign restriction as a function of the identified set size (arc length).

number of draws required to satisfy the sign restrictions is inversely proportional to the probability of meeting those restrictions. Figure 1b illustrates this relationship: the green line plots the expected number of draws needed to satisfy the sign restrictions as a function of the size of the identified set (i.e., the length of the red arc). More specifically, we generate smaller identified sets by gradually moving the left endpoint of the red arc toward its right endpoint. As shown in the figure, the expected number of draws required to meet the sign restrictions increases hyperbolically as the identified set shrinks. In realistic scenarios, as illustrated later in our empirical applications, the number of draws required can become quite large, rendering the algorithm inefficient.

In this paper, we propose a Gibbs sampling algorithm based on Elliptical Slice Sampling (ESS), which draws from the identified set more efficiently. This method can be viewed as an adaptive Metropolis-Hastings algorithm that transitions from the previous draw $\mathbf{x}_1^{(0)}$ to a new draw using the following elliptical proposal:

$$\mathbf{x}_1^{(*)} = \boldsymbol{\nu} \sin(\theta) + \mathbf{x}_1^{(0)} \cos(\theta) \text{ with } \theta \in [0, 2\pi],$$

where $\boldsymbol{\nu}$ is a 2×1 vector drawn from $N(\mathbf{0}, \mathbf{I}_2)$. The scalar parameter θ controls the step size of the proposed move. For instance, when θ is close to 0, the proposal is closer to the previous draw $\mathbf{x}_1^{(0)}$, whereas when θ approaches $\pi/2$, the proposal is closer to the newly drawn random vector $\boldsymbol{\nu}$. Unlike a conventional Metropolis-Hastings algorithm, ESS adaptively searches for a suitable step size to guarantee acceptance of the proposed draw at every iteration. Intuitively, given that the previous draw lies within the identified set, ESS ensures that the new proposal $\mathbf{x}_1^{(*)}$ also remains within the identified set by exponentially shrinking the candidate set for θ . Under appropriate regularity conditions, this procedure ensures the validity and convergence of the Gibbs sampling algorithm as long as the random variable of interest, in this case, \mathbf{q}_1 , can be written as a transformation of a normally distributed random variable, here denoted \mathbf{x}_1 . See [Murray, Adams, and MacKay \(2010\)](#) and [Natarovskii, Rudolf, and Sprungk \(2021\)](#) for details.

The fact that the candidate set for θ shrinks exponentially is an appealing feature, as it significantly reduces the number of candidate draws \mathbf{q}_1 needed to satisfy the restrictions. This efficiency gain becomes particularly important when the dimension of the model increases since generating new draws of \mathbf{q}_1 is computationally costly. Figure 1b (red line) displays the average number of trials required by ESS to generate an accepted draw of \mathbf{q}_1 within the identified set as a function of the length of the identified set. The number of required trials using ESS grows only logarithmically, whereas the conventional acceptance-rejection sampler grows hyperbolically.¹

In the following section, we extend this simple example into a more realistic and useful setting by: (1) identifying multiple shocks simultaneously rather than just a single shock; (2) allowing sign restrictions to take a general form; and (3) developing a Gibbs sampling algorithm that uses the ESS and that targets the posterior of the orthogonal reduced-form parameters conditional on the sign restrictions.

¹To make a fair comparison, one should account for the serial correlation introduced by the Gibbs sampling algorithm, since the accept-reject algorithm generates i.i.d. draws. As demonstrated in a later section, we compute the effective sample size and find that the number of draws required to obtain one i.i.d.-equivalent sample ranges from 1.04 to 1.35. Therefore, this adjustment does not alter the main conclusion illustrated in the figure.

4 An Algorithm

In this section, we propose a Gibbs sampler algorithm that uses the ESS to draw from the posterior of the orthogonal reduced-form parameters conditional on the sign restrictions defined in Section 2.3. Our algorithm requires Assumption 1.

Assumption 1. *The following conditions hold:*

- 1.1 *For almost all (\mathbf{B}, Σ) , the set $\{\mathbf{Q} \in \mathcal{O}(n) | \mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0\}$ has positive measure.*
- 1.2 *For almost all (\mathbf{B}, \mathbf{Q}) , the set $\{\Sigma \in \mathcal{S}(n) | \mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0\}$ has positive measure.*
- 1.3 *For almost all (Σ, \mathbf{Q}) , the set $\{\mathbf{B} \in \mathbb{R}^{m \times n} | \mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0\}$ has positive measure.*

As mentioned above, the algorithm will be written using a conjugate prior over the reduced-form parameters, but it also could be written using an independent conjugate prior. Because of the choice of uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameters prior distribution, the posterior can be written as:

$$p(\mathbf{B}, \Sigma, \mathbf{Q} | (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0) \propto [\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0] NIW_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B}, \Sigma). \quad (7)$$

Using Equation (7), we first obtain conditional posterior distributions satisfying the sign restrictions for each of the components of the orthogonal-reduced-form parameterization. Importantly, as it will become clear below, we will draw from each of these conditional distributions using the ESS. This is feasible because each of the conditional posteriors can be represented by a distribution featuring a Gaussian kernel allowing the use of the ESS.

Conditional Posterior for \mathbf{Q} . Let us begin by deriving the posterior for \mathbf{Q} given the reduced-form parameters and the sign restrictions. Equation (7) implies that

$$\begin{aligned} p(\mathbf{Q} | \mathbf{B}, \Sigma, (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0) &= \frac{[\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0] NIW_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B}, \Sigma)}{\int_{\mathcal{O}(n)} [\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0] NIW_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B}, \Sigma) d\mathbf{Q}} \\ &= \frac{[\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]}{\int_{\mathcal{O}(n)} [\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0] d\mathbf{Q}} \propto [\mathbf{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]. \end{aligned}$$

The first equality follows from Bayes' rule. The second equality is straightforward because $NIW_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B}, \Sigma)$ does not depend on \mathbf{Q} . The proportionality follows from Assumption 1.1.

Let us now explain how to draw the conditional posterior $p(\mathbf{Q} \mid \mathbf{B}, \Sigma, (\mathbf{y}_t)_{t=1}^T, \mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0)$. We will exploit the mapping from $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ to \mathbf{Q} , where $\mathbf{x}_j \in \mathbb{R}^{n+1-j}$ with each entry of \mathbf{x}_j being distributed according to a standard normal distribution, for $j = 1, \dots, n$, implies that \mathbf{Q} is uniform under the Haar measure. More specifically, this mapping is related to the QR -decomposition of \mathbf{X} and we denote it by $\mathbf{Q} = \gamma(\mathbf{X})$. Hence, if one draws \mathbf{X} from $[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]p(\mathbf{X})$ and then transforms the draws using $\mathbf{Q} = \gamma(\mathbf{X})$, one obtains draws from the desired distribution, which is proportional to $[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]$. Because \mathbf{X} is normal, we can sample from $p(\mathbf{X})$ using the ESS.

Conditional Posterior for Σ . Next, we derive a useful expression for the posterior of Σ given (\mathbf{B}, \mathbf{Q}) and the sign restrictions. In this case, Equation (7) implies that

$$\begin{aligned} p(\Sigma \mid \mathbf{B}, \mathbf{Q}, (\mathbf{y}_t)_{t=1}^T, \mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0) &= \frac{[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]N_{(\tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B})IW_{(\tilde{\nu}, \tilde{\Phi})}(\Sigma)}{\int_{\mathcal{S}(n)} [\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]N_{(\tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B})IW_{(\tilde{\nu}, \tilde{\Phi})}(\Sigma)d\Sigma} \\ &\propto [\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]N_{(\tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B})IW_{(\tilde{\nu}, \tilde{\Phi})}(\Sigma). \end{aligned}$$

The first equality follows from Bayes' rule and the following factorization of the normal-inverse Wishart density: $NIW_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B}, \Sigma)$ as $N_{(\tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B})IW_{(\tilde{\nu}, \tilde{\Phi})}(\Sigma)$. As we demonstrate below, this factorization is convenient to demonstrate that the posterior under analysis can be induced by a kernel that contains a Gaussian density among its components. The proportionality follows from Assumption 1.2.

Let us now explain how to draw from $p(\Sigma \mid \mathbf{B}, \mathbf{Q}, (\mathbf{y}_t)_{t=1}^T, \mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0)$. We will exploit the mapping between an $n \times \tilde{\nu}$ matrix $\mathbf{R} \sim N(\mathbf{0}, \tilde{\Phi}^{-1})$ and Σ . More specifically, if $\Sigma = (\mathbf{R}\mathbf{R}')^{-1}$ we have that $\Sigma \sim IW(\tilde{\nu}, \tilde{\Phi})$. Let us denote such a mapping by $\Sigma = \varsigma(\mathbf{R})$. Hence, if one draws \mathbf{R} from $[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]N_{(\tilde{\Psi}, \varsigma(\mathbf{R}) \otimes \tilde{\Omega})}(\mathbf{B})N_{(\mathbf{0}, \tilde{\Phi}^{-1})}(\mathbf{R})$ and then transforms the draws using $\Sigma = \varsigma(\mathbf{R})$, one obtains draws from the desired distribution $[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}) > 0]N_{(\tilde{\Psi}, \Sigma \otimes \tilde{\Omega})}(\mathbf{B})IW_{(\tilde{\nu}, \tilde{\Phi})}(\Sigma)$. As before, because \mathbf{R} is normal, we can sample from $[\mathcal{S}(\mathbf{B}, \Sigma, \mathbf{Q}^i) > 0]N_{(\tilde{\Psi}, \varsigma(\mathbf{R}) \otimes \tilde{\Omega})}(\mathbf{B})N_{(\mathbf{0}, \tilde{\Phi}^{-1})}(\mathbf{R})$ using the ESS.

Conditional Posterior for \mathbf{B} The third and last conditional posterior corresponds to the posterior of \mathbf{B} given $(\boldsymbol{\Sigma}, \mathbf{Q})$ and the sign restrictions. In this case, we use Equation (7) to obtain:

$$\begin{aligned} p(\mathbf{B} \mid \boldsymbol{\Sigma}, \mathbf{Q}, (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0) &= \frac{[\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma} \otimes \tilde{\Omega})}(\mathbf{B}) IW_{(\tilde{\nu}, \tilde{\Phi})}(\boldsymbol{\Sigma})}{\int_{\mathbb{R}^{m \times n}} [\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma} \otimes \tilde{\Omega})}(\mathbf{B}) IW_{(\tilde{\nu}, \tilde{\Phi})}(\boldsymbol{\Sigma}) d\mathbf{B}} \\ &\propto [\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma} \otimes \tilde{\Omega})}(\mathbf{B}). \end{aligned}$$

The proportionality comes from Assumption 1.3. Notice that in this case, it is immediately clear that the conditional posterior of interest contains a Gaussian kernel. Hence, we can draw from $[\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma} \otimes \tilde{\Omega})}(\mathbf{B})$ using the ESS.

Having defined the three conditional posteriors described above, we now can write a Gibbs Sampler of the following form:

Algorithm 2. *This algorithm draws from $p(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q} \mid (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) > 0)$.*

1. Let $I > 1$ and set $i = 1$ and assign initial values to $(\mathbf{B}^{i-1}, \boldsymbol{\Sigma}^{i-1})$.

2. Draw \mathbf{Q}^i from

$$p(\mathbf{Q} \mid \mathbf{B}^{i-1}, \boldsymbol{\Sigma}^{i-1}, (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}^{i-1}, \boldsymbol{\Sigma}^{i-1}, \mathbf{Q}) > 0) \propto [\mathbf{S}(\mathbf{B}^{i-1}, \boldsymbol{\Sigma}^{i-1}, \mathbf{Q}) > 0].$$

3. Draw $\boldsymbol{\Sigma}^i$ from

$$p(\boldsymbol{\Sigma} \mid \mathbf{B}^{i-1}, \mathbf{Q}^i, (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}^{i-1}, \boldsymbol{\Sigma}, \mathbf{Q}^i) > 0) \propto [\mathbf{S}(\mathbf{B}^{i-1}, \boldsymbol{\Sigma}, \mathbf{Q}^i) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma} \otimes \tilde{\Omega})}(\mathbf{B}^{i-1}) IW_{(\tilde{\nu}, \tilde{\Phi})}(\boldsymbol{\Sigma}).$$

4. Draw \mathbf{B}^i from

$$p(\mathbf{B} \mid \boldsymbol{\Sigma}^i, \mathbf{Q}^i, (\mathbf{y}_t)_{t=1}^T, \mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}^i, \mathbf{Q}^i) > 0) \propto [\mathbf{S}(\mathbf{B}, \boldsymbol{\Sigma}^i, \mathbf{Q}^i) > 0] N_{(\tilde{\Psi}, \boldsymbol{\Sigma}^i \otimes \tilde{\Omega})}(\mathbf{B}).$$

5. If $i < I$, let $i = i + 1$ and return to Step 2.

5 Applications

We now illustrate the performance of our algorithm using two empirical applications. The first is a small-scale SVAR of the global oil market based on the model in [Kilian and Murphy \(2014\)](#), which identifies flow supply, flow demand, and speculative demand shocks using a combination of sign and ranking restrictions. The tight identifying assumptions in this model render traditional accept-reject methods computationally intensive, whereas our algorithm improves efficiency while replicating the main results. The efficiency gains increase with the tightness of the identified set as we demonstrate in the second application. This revisits the large-scale SVAR model of the U.S. economy developed by [Crump, Eusepi, Giannone, Qian, and Sbordone \(2025\)](#) and analyzed structurally by [Chan, Matthes, and Yu \(2025\)](#). This model includes 35 macroeconomic and financial variables and identifies up to eight structural shocks. We demonstrate that our algorithm remains computationally stable as the number of restrictions increases, in contrast to the exponential rise in computation time exhibited by the accept-reject method. Both applications highlight the accuracy and scalability of our approach in different settings. For both applications, we first demonstrate that our approach replicates the main results reported in each of the original papers. We then analyze the computational timing to show that our method can be more efficient than the traditional accept-reject algorithm.

5.1 Small SVAR of the World Oil Market

In our first application, we replicate the results of [Kilian and Murphy \(2014\)](#), who extend the framework of [Kilian and Murphy \(2012\)](#) by incorporating oil inventories to identify speculative demand shocks. The identification strategy in [Kilian and Murphy \(2014\)](#) relies on tight sign and ranking restrictions, which result in a small identified set and may render standard accept-reject algorithms slow. As we demonstrate below, our algorithm can implement this identification strategy faster the time required by the accept-reject method used by [Chan, Matthes and Yu \(2025\)](#). Notably, [Kilian and Murphy \(2014\)](#) adopt an approach similar to that of [Chan, Matthes, and Yu \(2025\)](#), relying on permutations and sign alternations.

Therefore, the computation times we report for the accept-reject algorithm are comparable to those in the original study.

Model Specification and Impulse Responses

Let us begin by describing the model specification in [Kilian and Murphy \(2014\)](#). They model the global market for crude oil using a 4-variable SVAR featuring the percent change in global crude oil production, a measure of global real activity, the real price of crude oil, and the change in global crude oil inventories above ground. The SVAR is specified at monthly frequency and the estimation sample is 1973:M2–2009:M8. The model includes 24 lags, a constant, and seasonal dummies to remove seasonal variation. [Kilian and Murphy \(2014\)](#) adopt a weak Gaussian-inverse Wishart prior distribution (see e.g. [Uhlig, 2005](#)) for the reduced-form parameters.

Turning to identification, the goal of [Kilian and Murphy \(2014\)](#) is to identify 3 structural shocks by using a combination of sign restrictions on impact impulse responses, on impulse responses at horizons 1 through 12, and elasticity bounds (also known as ranking restrictions following the work of [Amir-Ahmadi, Matthes, and Wang, 2020](#)). Table 1 summarizes the identifying assumptions. The structural shocks are labeled flow supply shock, flow demand shock, and speculative demand shock.

Variable/Shock	Sign Restrictions on Impact Impulse Responses		
	Flow supply	Flow demand	Speculative demand
Oil production	−1	+1	+1
Real activity	−1	+1	−1
Real price of oil	+1	+1	+1
Inventories			+1
Elasticity Bounds			
	Flow supply shock	Flow demand shock	Speculative demand shock
Price Elasticity of Oil Supply		(0,0.025)	(0,0.025)
Sign Restrictions on Impulse Responses at Horizons 1 through 12			
	Flow supply shock	Flow demand shock	Speculative demand shock
Real activity	−1		
Real price of oil	+1		

Table 1: Sign and Ranking Restrictions

Note: All shocks raise the real oil price. ± 1 indicates positive or negative sign restrictions; blanks mean no restriction.

We adopt the exact same specification and use Algorithm 2. We obtain 1 million draws and save one every 100. Figure 2 shows the results. The results are broadly identical to those

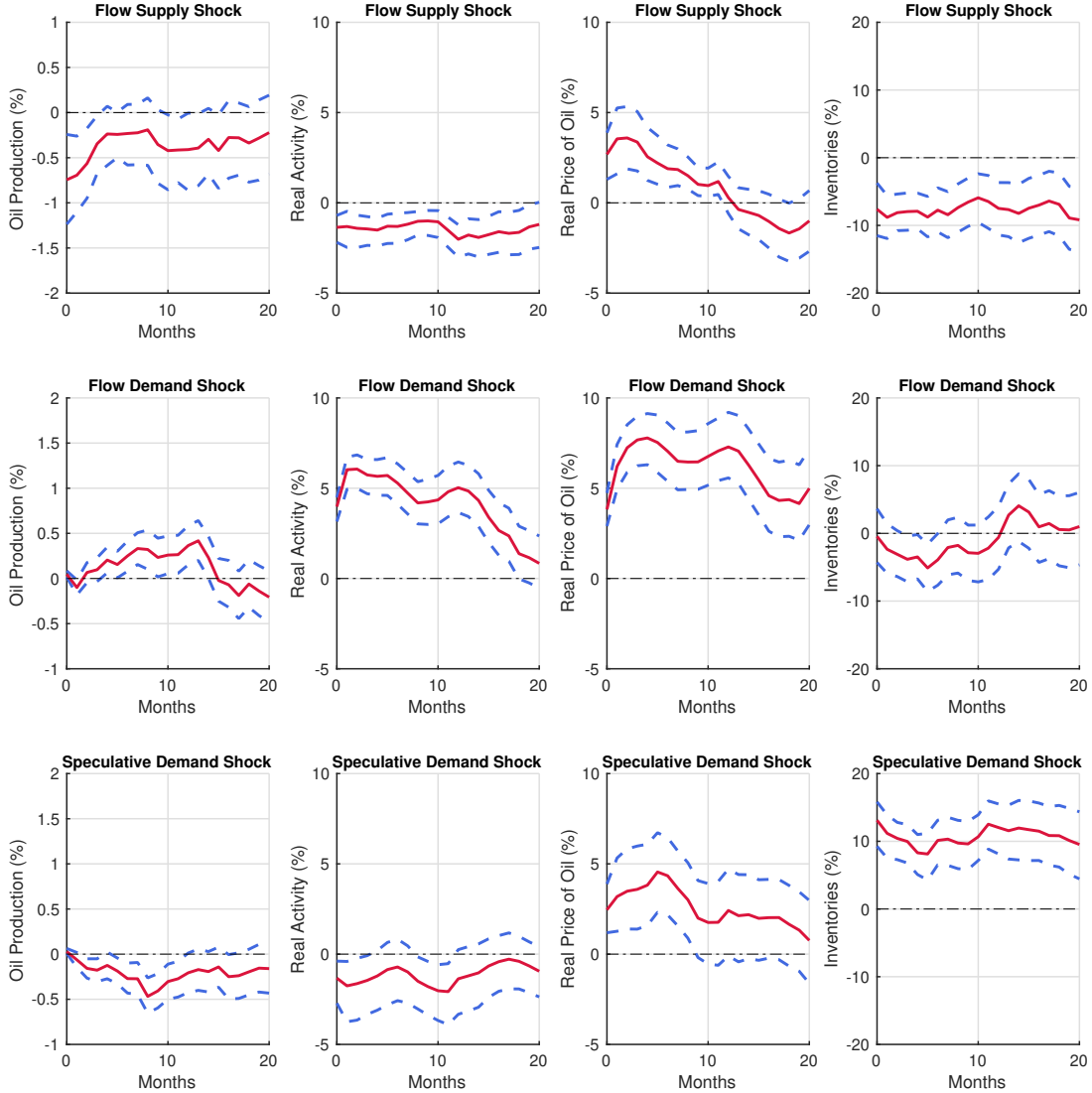


Figure 2: Impulse Responses

Note: The solid red lines depict the point-wise posterior median. The dashed blue lines depict the point-wise 68 percent posterior probability bands.

in Kilian and Murphy (2014). In particular, a negative flow supply shock causes a persistent decline in global economic activity and oil inventories, and a persistent increase in the real price of oil. The response of oil production is persistently negative. A positive flow demand shock is associated with a persistent increase in global economic activity and it causes a persistent increase the real price of oil as well as a positive response of oil production. Oil

production increases sluggishly given the imposed elasticity bounds and it peaks at about one year after the shock before declining to pre-shock levels. Finally, a positive speculative demand shock causes a persistent increase in the real price of oil and large increase in inventories. Global real activity and oil production persistently decline in response to this shock but the effects are small.

Timing

Next we compare the computational time of applying our Gibbs sampler relative to the accept-reject algorithm. Table 2 reports the time (in minutes) per 1,000 effective draw (i.e., time/effective sample size) using Algorithm 2 and the accept-reject algorithm. As highlighted above, we use Chan, Matthes and Yu’s variant of the accept-reject algorithm. To compute the time per effective draw using the Gibbs sampler, first we generate 100,000 draws and save one every 100. This yields 1,000 draws that we store along with the computational time. Second, we use the resulting 1,000 draws to approximate the effective sample size and we use it to compute the time per effective draw. Since the object of interest in SVARs analysis is typically multi-dimensional, that is, researchers are interested in vector of impulse responses, we compute the multivariate effective sample size proposed by Vats, Flegal, and Jones (2019) based on the vector of contemporaneous impulse response functions.²

The column “Benchmark Model” compares the effective draws per minute under Kilian and Murphy’s specification. As can be seen the results are comparable: The numbers reported in the table imply that in this case both algorithms take the same time: for both the Gibbs and the accept-reject sampler it would take about 3 hours to obtain 1,000 effective sample draws.³ In order to demonstrate the advantages of our proposed Gibbs sampler algorithm, we consider a researcher that intends to add a bound on the elasticity of oil price demand to the flow supply shock. As demonstrated by Caldara, Cavallo, and Iacoviello (2019) this is an

²Multivariate effective sample size calculations generally require a longer MCMC sequence as dimensionality increases; therefore, we focus on contemporaneous impulse responses. Our results are robust to including impulse responses beyond horizon 0. In fact, in our application the MCMC draws of impulse responses are less correlated at longer horizons, so the reported effective sample size should be viewed as conservative. We estimate the multivariate effective sample size based on a multivariate batch mean estimator as in Vats, Flegal, and Jones (2019), with a batch size of 100.

³The computations were performed using MATLAB on an Intel Xeon Platinum 8488C processor with 16 active cores running at 2.4 GHz on an x86_64 architecture.

important consideration in SVARs models of the oil market. Following their work we add the ranking restriction that the price elasticity of crude oil demand is restricted to be in a tight interval around the point estimate of -0.08 reported by [Caldara, Cavallo, and Iacoviello \(2019\)](#); in particular, we restrict this elasticity to the interval $(-0.09, -0.07)$.

As column “Benchmark Model+Additional Ranking Restriction” in [Table 2](#) shows, the efficiency of the Gibbs sampler algorithm remains approximately unchanged as we add the ranking restriction. On the contrary, the performance of the accept-reject algorithm declines significantly to the point where we had to extrapolate the times: it would take more than 80 hours to obtain 1,000 effective draws.

Specification	Benchmark Model	Benchmark Model + Additional Ranking Restriction
Gibbs Sampler	≈ 3	≈ 3
Accept-Reject	≈ 3	≈ 80

Table 2: Time (Minutes) Per 1,000 Effective Draws

These results show that, for this model, the accept-reject algorithm is near its maximum capacity under the benchmark specification. Our attempt to impose one additional restriction renders the algorithm infeasible. In contrast, the Gibbs sampler preserves its performance even as additional new ranking restriction is introduced.

5.2 Large SVAR of the U.S. Economy

In our second application, we replicate and extend the analysis of [Chan, Matthes, and Yu \(2025\)](#), who build on the large-scale SVAR framework of [Crump, Eusepi, Giannone, Qian, and Sbordone \(2025\)](#) to study the structural dynamics of the U.S. economy. Their model incorporates 35 macroeconomic and financial variables commonly monitored by the Federal Reserve and identifies 8 structural shocks using an extensive set of sign and ranking restrictions. [Chan, Matthes, and Yu \(2025\)](#) employ an accept-reject algorithm for inference, which becomes computationally intensive as the number of identifying restrictions increases. As we show below, our algorithm can implement this identification strategy more efficiently.

In particular, we extend the baseline model by identifying two additional shocks—an oil price shock and a consumer sentiment shock—bringing the total number of sign restrictions from 105 to 129. This provides a stringent test of our algorithm’s performance relative to the accept-reject method.

Model Specification and Impulse Responses

Table 3 provides a summary of the variables and the sign restrictions imposed on their contemporaneous impulse responses to 10 structural shocks. [Chan, Matthes, and Yu \(2025\)](#) consider only the first 8 shocks. We have added the last two to assess the performance of our algorithm. These last two shocks are labeled oil price and consumer sentiment shock, respectively. The SVAR is specified at quarterly frequency, it includes a constant and 5 lags, and the estimation sample is 1973:Q2–2019:Q4. We assume a Minnesota prior for the reduced-form parameters and we set the hyper-parameters following [Giannone, Lenza, and Primiceri \(2015\)](#). Importantly, [Chan, Matthes, and Yu \(2025\)](#) use the asymmetric priors defined in [Chan \(2022\)](#) for the reduced-form parameters, hence we will use these same priors when considering the accept-reject approach.⁴

Turning to the identification, [Chan, Matthes, and Yu \(2025\)](#) identify 8 structural shocks (demand, investment, financial, monetary, government spending, technology, labor supply and wage bargaining) by means of sign restrictions on the contemporaneous impulse responses as well as by ranking restrictions. In total, there are 105 sign restrictions imposed in their baseline specification. When considering our two added shocks, the oil price and consumer sentiment shocks mentioned above, we increase the number of sign restrictions to 129. When using the Gibbs sampler we obtain 1 million draws and save one every 100.

Let us begin by describing the selected impulse responses to a unit standard deviation expansionary demand shock, shown in Figure 3a. Red lines depict point-wise posterior medians. Shaded areas represent point-wise 68 percent posterior probability bands. The sign of the impact responses of real GDP, the PCE price index, the federal funds rate, and the unemployment rate are restricted. The remaining horizons as well the response of the non-residential investment and the real wage are unrestricted. As can be seen, the demand

⁴We thank Christian Matthes for sharing their replication files with us.

Sign restrictions	Dem	Inv	Fin	Mon	Gov	Tec	Lab	Wag	Oil	Con
GDP	+1	+1	+1	-1	+1	+1	+1	+1	+1	+1
PCE	0	0	0	0	0	+1	0	0	+1	+1
Residential investment	0	0	0	0	0	0	0	0	0	+1
Nonresidential investment	0	+1	0	0	0	+1	0	0	+1	+1
Exports	0	0	0	0	0	0	0	0	0	0
Imports	0	0	0	0	0	0	0	0	0	0
Government spending	0	0	0	0	+1	0	0	0	0	0
Fed. budget surplus/deficit	0	0	0	0	-1	0	0	0	0	0
Fed. tax receipts	0	0	0	0	+1	0	0	0	0	0
GDP deflator	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
PCE index	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
PCE index less F&E	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
CPI index	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
CPI index less F&E	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
Hourly wage	0	0	0	0	0	+1	-1	-1	+1	0
Labor productivity	0	0	0	0	0	+1	0	0	+1	0
Utilization-adjusted TFP	0	0	0	0	0	+1	0	0	+1	0
Employment	0	0	0	-1	0	0	-1	0	0	0
Unemployment rate	-1	-1	-1	+1	-1	-1	+1	-1	+1	+1
Industrial production index	+1	+1	+1	-1	0	0	0	0	0	0
Capacity utilization	+1	+1	+1	-1	0	0	0	0	0	0
Housing starts	0	0	0	0	0	0	0	0	0	0
Disposable income	0	0	0	0	0	0	0	0	0	0
Consumer sentiment	0	0	0	0	0	0	0	0	0	0
Fed funds rate	+1	+1	+1	+1	+1	0	0	0	0	0
3-month T-bill rate	+1	+1	+1	+1	+1	0	0	0	0	0
2-year T-note rate	0	0	0	+1	0	0	0	0	0	0
5-year T-note rate	0	0	0	+1	0	0	0	0	0	0
10-year T-note rate	0	0	0	+1	0	0	0	0	0	0
Prime rate	+1	+1	+1	+1	+1	0	0	0	0	0
Aaa corporate bond yield	0	0	0	+1	0	0	0	0	0	0
Baa corporate bond yield	0	0	0	+1	0	0	0	0	0	0
Trade-weighted US index	0	0	0	0	0	0	0	0	0	0
S&P 500	0	-1	+1	-1	0	0	0	0	0	+1
Spot oil price	0	0	0	0	0	0	0	0	-1	0
Ranking restrictions										
Nonresidential investment/GDP	-1	+1	+1	0	0	0	0	0	0	0
Government spending/GDP	-1	-1	-1	0	+1	0	0	0	0	0
N ⁰ of restrictions	14	15	15	19	14	12	8	8	13	11
Cum. N ⁰ of restrictions	14	29	44	63	77	89	97	105	118	129

Table 3: Sign restrictions, ranking restrictions and identified shocks for the 35-variable VAR.

Note: The mnemonics for the shocks are as follows. Dem: demand, Inv: Investment, Fin: Financial, Mon: Monetary Policy, Gov: Government Spending, Tec: Technology, Lab: Labor Supply, Wag: Wage Bargaining, Oil: Oil Price, Con: Consumer Sentiment.

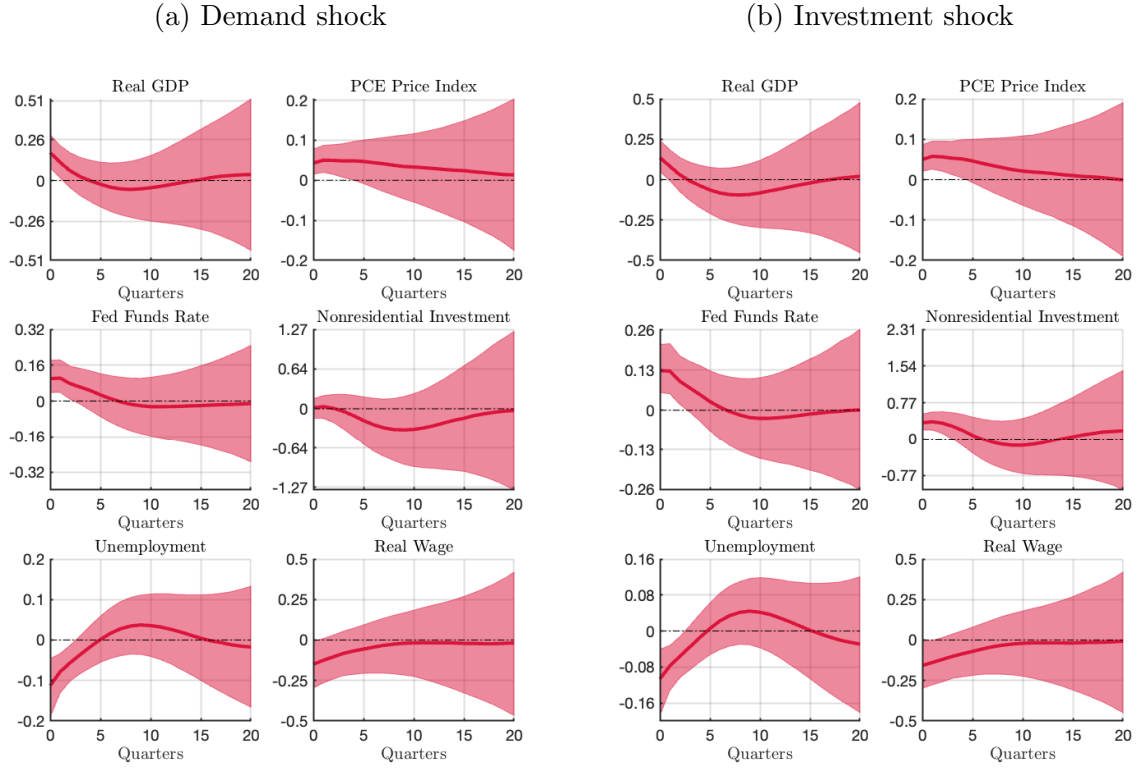


Figure 3: Impulse responses to demand-side structural shocks.

shock causes a transient increase in output and prices, and a decrease in the unemployment rate. The federal funds rate increases in response to the shock. The restrictive stance of monetary policy eventually lowers economic activity as can be seen for example in the decline of non-residential investment. The real wage decreases in the short-run in response to the shock as nominal wage increases are not enough to offset higher prices—possibly due to sluggish nominal wage adjustment.

The investment shock shown in Figure 3b looks similar to the demand shock in terms of the economic consequences for real GDP, the federal funds rate, the price level, and the unemployment rate. However, the impulse response of non-residential investment is substantially different. In particular, the investment shock causes a short-run boom in non-residential investment. In part, this finding is a consequence of the ranking restriction requiring that the impact response of non-residential investment be larger than the impact response of real GDP. As in the case of the demand shock, the investment shock causes a persistently negative response of the real wage.

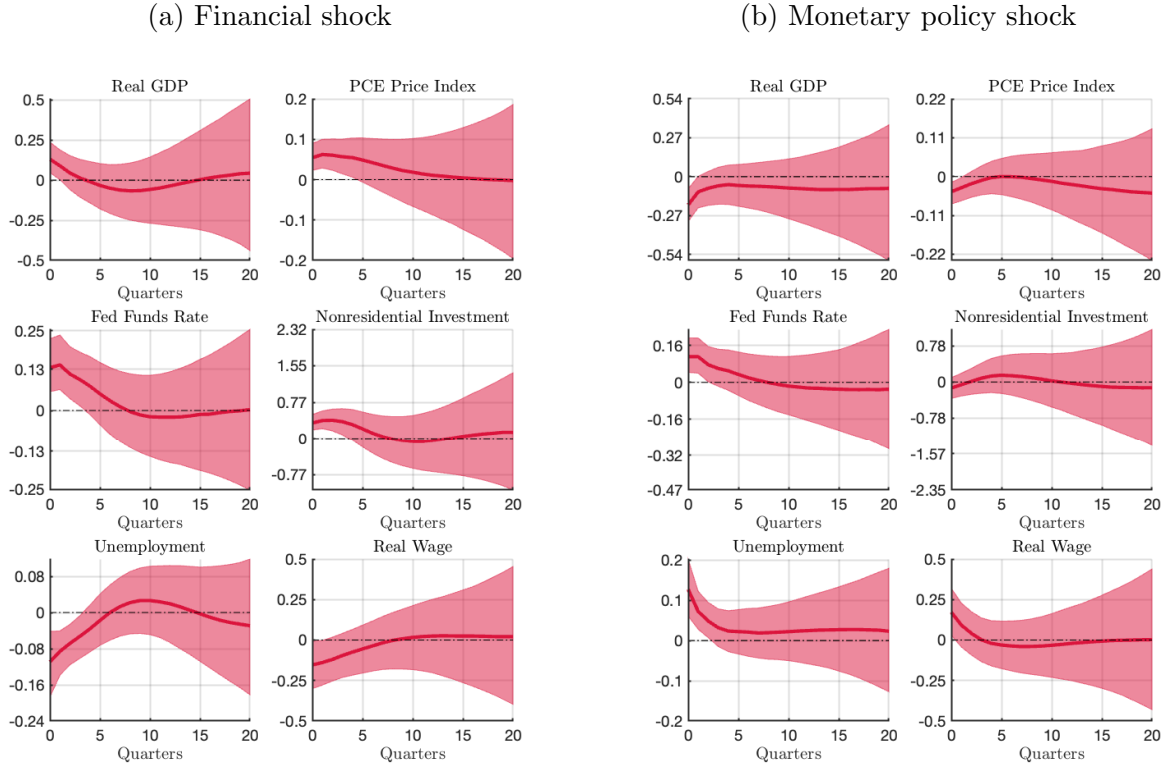
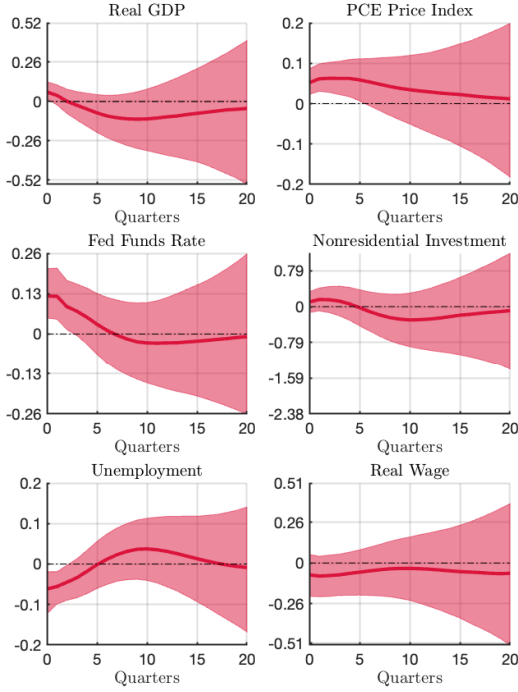


Figure 4: Impulse responses to financial and monetary policy shocks.

Turning to the financial shock, shown in Figure 4a, it is worth highlighting that this shock is identified with the same sign restrictions used to identify the investment shock except for the impact response of the S&P 500, which in the case of the financial shock is assumed to be positive instead of negative. Overall, the impulse responses are similar except that the decline in non-residential investment after 5 quarters is slightly less pronounced in the case of a financial shock consistent with the positive response of asset prices.

The impulse responses to a unit standard deviation contractionary monetary policy shock are depicted in Figure 4b. This shock causes the federal funds rate to remain above zero for more than 2 years, reflecting inertia in the conduct of monetary policy. Real GDP and prices decline persistently and the unemployment rate jumps upon impact before slowly returning to baseline. Non-residential investment drops upon impact and rounds one year after the shock in line with a less restrictive monetary policy stance. The real wage increases, driven by a decrease of the price level. A notable aspect of these responses is that they suggest that monetary policy can operate with shorter-lags relative to the long-and-variable lags wisdom.

(a) Government spending shock



(b) Technology shock

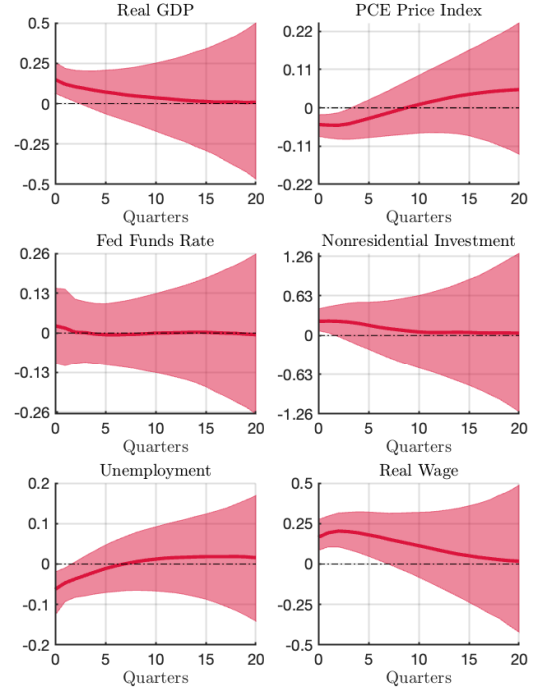


Figure 5: Impulse responses to government spending and technology shocks.

The government spending shock is shown in Figure 5a. An expansionary one unit standard deviation government spending shock leads to an increase of real GDP for about two quarters and to a long-lasting increase in the price level. To conclude, we discuss the impulse responses to the supply-related structural shocks, that is, the technology, labor supply, and wage bargaining shock. A unit standard deviation positive technology shock leads to a protracted increase in real GDP and non-residential investment, see Figure 5b. The higher level of output is accompanied by a sustained decline in the unemployment rate and a sustained increase in the real wage. The federal funds rate rises marginally, indicating that monetary policy remains roughly neutral in response to technology shocks.

The responses to a unit standard deviation positive labor supply shock are shown in Figure 6a. This shock induces a hump-shaped response of real GDP and leads to persistently lower prices. The responses to a unit standard deviation negative wage bargaining shock are shown in Figure 6b. The identifying assumptions for this shock are identical to an expansionary labor supply shock except that the unemployment rate is assumed to decrease upon impact.

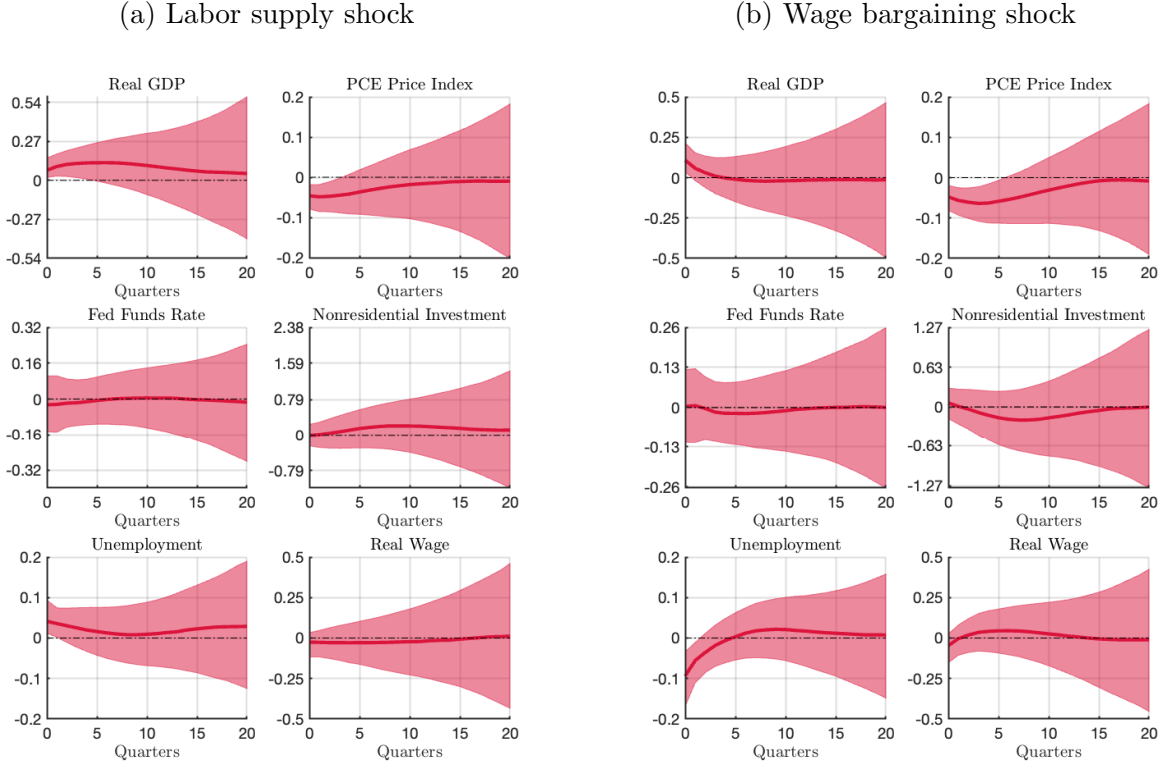


Figure 6: Impulse responses to labor market shocks.

When a negative wage bargaining shock occurs, workers experience a decline in their nominal wage alongside a decrease in the unemployment rate. The real wage remains unaffected upon impact as the lower wages are offset by the assumed decrease in the price level. Subsequently, the price level remains below zero, inducing an increase in the real wage.

Timing

In this section we illustrate the key insight of Figure 1 for this particular model. That is, for a given number of variables, the timing of our algorithm remains computationally feasible as the identified set narrows. In contrast, the time of the accept-reject algorithm increases dramatically and becomes impractical.

Let us begin by comparing the efficiency of the Gibbs sampler algorithm relative to the accept-reject algorithm when replicating the identification scheme in [Chan, Matthes and Yu \(2025\)](#). Figure 7a reports the time (in minutes) per 1,000 effective draws using Algorithm 2 applied to the 35-variable SVAR as a function of the number of shocks identified. We first

approximate the multivariate effective sample size 1 million draws and saving one every 100, and second we divide the time required to obtain such draws by the multivariate effective sample size and scale the resulting number by 1,000. In order to assess the computational time as a function of the size of the identified set, we first obtain draws by identifying only the demand shock, then we add the investment shock, then the financial shock, and so on until we identify the eight shocks in Table 3. As can be seen, the time per 1,000 effective draws remains computationally feasible as the number of sign restrictions increases.

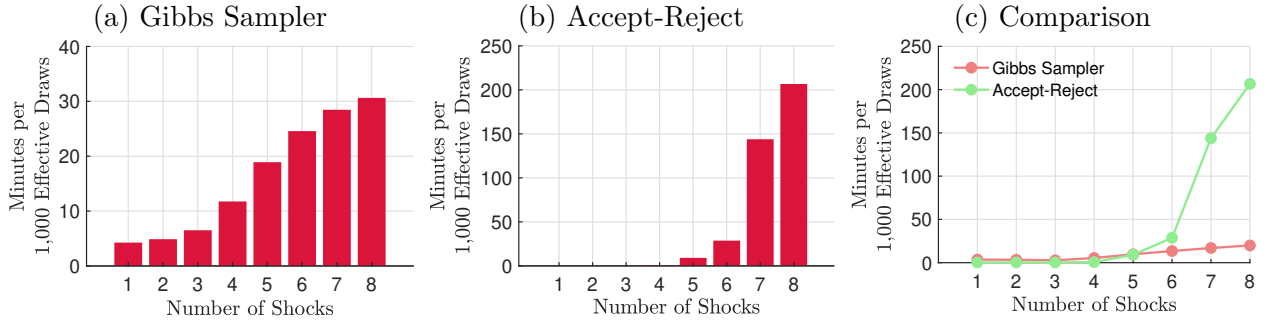


Figure 7: Time Per 1,000 Effective Draws

Figure 7b repeats the same figure but using the efficient accept-reject version of Algorithm 1 proposed by Chan, Matthes, and Yu (2025). We do not need to thin the draws because the draws are independent, hence we produce only 1,000 draws. In this case, the timing increases dramatically as demonstrated in the figure.⁵ Let us highlight that even though both codes (for the Gibbs sampler and the accept-reject sampler) can be further optimized and the time will vary depending on the hardware architecture and the number of variables, once we fix the number of variables the main message emerging from comparing Figures 7a-7b will not be altered: the accept-reject algorithm can deteriorate sharply as the identified set narrows. Figure 7c puts the times together to facilitate the visual comparison.

To emphasize the previous point, let us now consider more shocks to the point that eventually accept-reject can become impractical. To this end, we augment the number of shocks identified in Chan, Matthes, and Yu (2025) by adding the oil price and a consumer sentiment shock described in Table 3. Figures 8a-8d repeat the exercise shown in Figure 7a-7b

⁵If we were using the normal-inverse-Wishart priors the time would deteriorate even further as drawing from the posterior when using asymmetric priors is more efficient.

for the case of 9 and 10 shocks.⁶

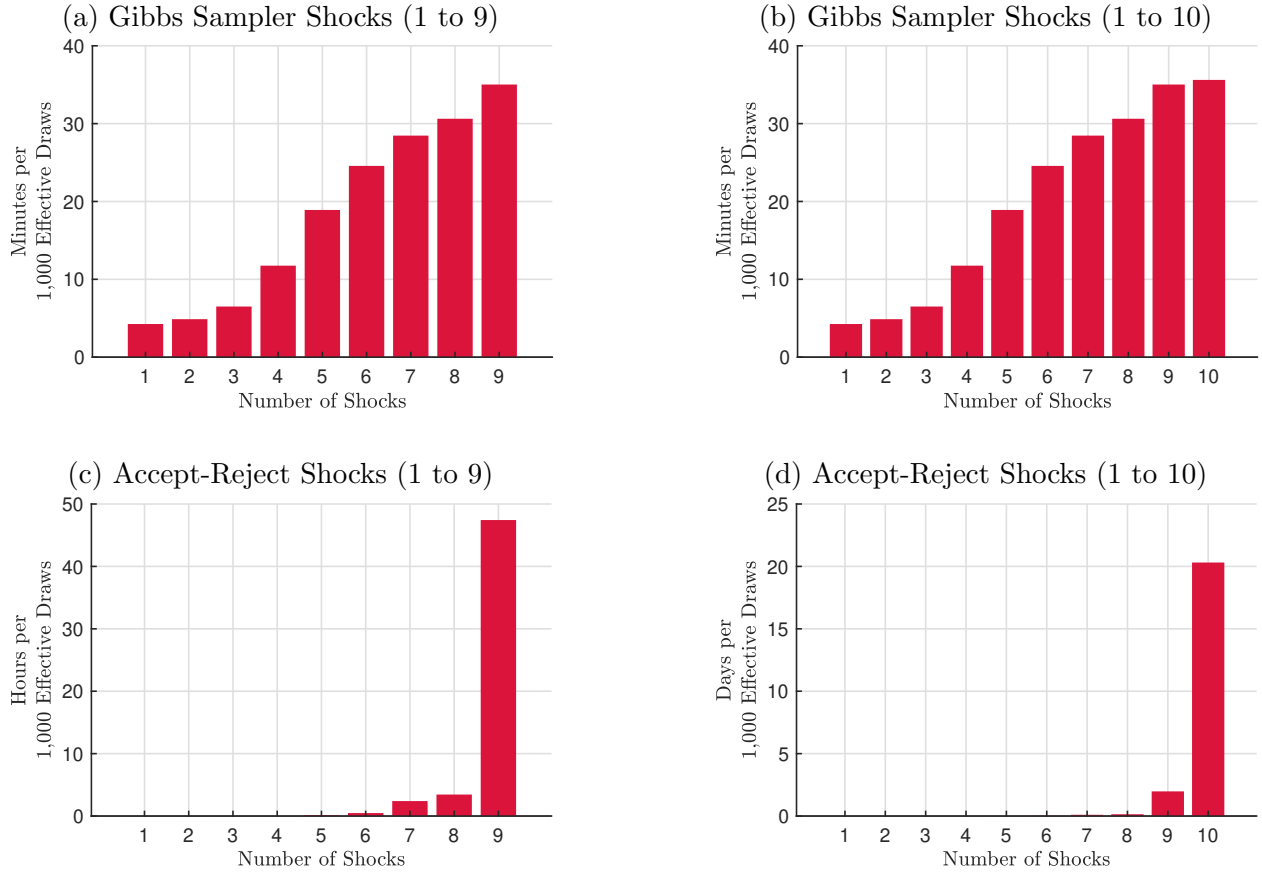


Figure 8: Gibbs Sampler vs. Accept-Reject

Note: The time of the accept-reject algorithm for shocks 9 and 10 is extrapolated based on 10 draws.

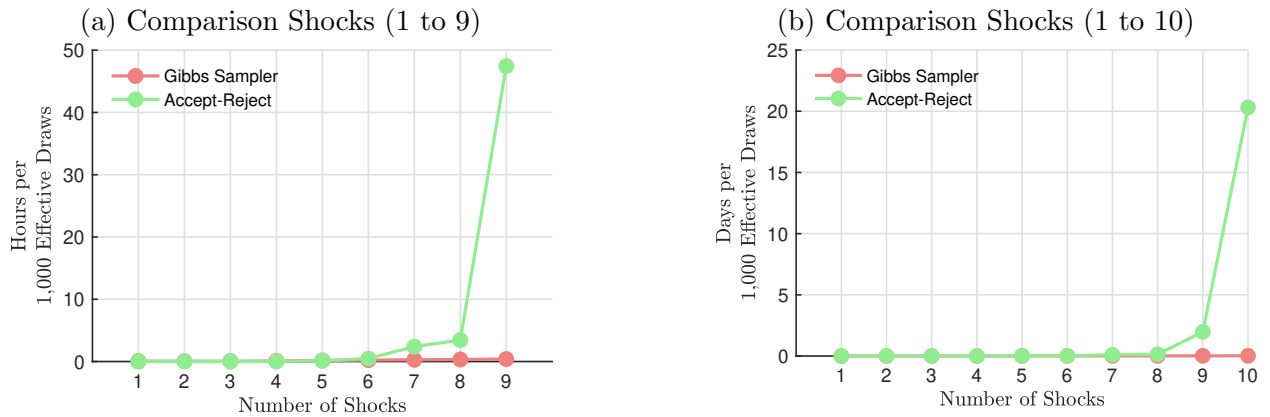


Figure 9: Gibbs Sampler vs. Accept-Reject

Note: The time of the accept-reject algorithm for shocks 9 and 10 is extrapolated based on 10 draws.

⁶The time of the accept-reject algorithm for shocks 9 and 10 is extrapolated based on 10 draws.

As the reader can see, the time does not increase exponentially when we use the ESS approach. In contrast, when we consider 9 shocks under the accept-reject approach the times are now measured in hours and when we consider to 10 shocks the times are measured in days—Figures 8c and 8d provide the times. In order to facilitate the comparison, Figure 9 overlays both timings to make clear that our algorithm handles settings (in terms of number of variables and shocks) that the traditional accept-reject cannot handle.

6 Conclusion

Our Gibbs sampler algorithm opens the door to tackling previously practically infeasible problems using currently available algorithms. In addition, it enables more efficient implementations of current SVAR analysis under both Bayesian and frequentist paradigms. First, from a frequentist perspective, our method directly facilitates the characterization of identified sets in sign-identified SVARs (i.e., [Gafarov, Meier and Montiel Olea, 2018](#)). Second, the efficient characterization of identified sets can improve the implementation of prior-robust Bayesian inference (i.e., [Giacomini and Kitagawa, 2021](#)). Finally, our Gibbs sampler naturally extends to serve as a mutation step within tempered sequential Monte Carlo (SMC), effectively propagating particles through a Markov transition kernel (i.e., [Herbst and Schorfheide, 2014](#)). In this context, a clear and intuitive tempering schedule arises by sequentially introducing restrictions one at a time.

References

- Amir-Ahmadi, P. and T. Drautzburg (2021). Identification and Inference with Ranking Restrictions. *Quantitative Economics* 12(1), 1–39.
- Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2020). Choosing Prior Hyperparameters: With Applications to Time-Varying Parameter Models. *Journal of Business and Economic Statistics* 38(1), 124–136.

- Arias, J., J. F. Rubio-Ramirez, and D. F. Waggoner (2025). Uniform Priors for Impulse Responses. *Econometrica* 93, 695–718.
- Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications. *Econometrica* 86(2), 685–720.
- Baumeister, C. and J. D. Hamilton (2024). Advances in Using Vector Autoregressions to Estimate Structural Magnitudes. *Econometric Theory* 40(3), 472–510.
- Bañbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Caldara, D., M. Cavallo, and M. Iacoviello (2019). Oil Price Elasticities and Oil Price Fluctuations. *Journal of Monetary Economics* 103, 1–20.
- Canova, F. and G. De Nicoló (2002). Monetary Disturbances Matter for Business Fluctuations in the G-7. *Journal of Monetary Economics* 49(6), 1131–1159.
- Carriero, A., G. Kapetanios, and M. Marcellino (2009). Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models. *Journal of Applied Econometrics* 24(5), 763–786.
- Chan, J., C. Matthes, and X. Yu (2025). Large Structural VARs with Multiple Sign and Ranking Restrictions.
- Chan, J. C. C. (2022). Asymmetric Conjugate Priors for Large Bayesian VARs. *Quantitative Economics* 13(3), 1145–1169.
- Crump, R. K., S. Eusepi, D. Giannone, E. Qian, and A. M. Sbordone (2021). A Unified Approach to Measuring u^* . *Brookings Papers on Economic Activity* 2021(1), 1–65.
- Crump, R. K., S. Eusepi, D. Giannone, E. Qian, and A. M. Sbordone (2025). A Large Bayesian VAR of the United States Economy. *International Journal of Central Banking* (Forthcoming).

- Del Negro, M. and F. Schorfheide (2010). Bayesian Macroeconometrics. In *The Oxford Handbook of Bayesian Econometrics*.
- Ellahie, A. and G. Ricco (2017). Government Purchases and the American Business Cycle: Evidence from WWII Procurement. *American Economic Journal: Macroeconomics* 9(4), 1–38.
- Faust, J. (1998). The Robustness of Identified VAR Conclusions about Money. *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.
- Gafarov, B., M. Meier, and J. L. Montiel Olea (2018). Delta-Method Inference for a Class of Set-Identified SVARs. *Journal of Econometrics* 203(2), 316–327.
- Giacomini, R. and T. Kitagawa (2021). Robust Bayesian Inference for Set-Identified Models. *Econometrica* 89(4), 1519–1556.
- Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior Selection for Vector Autoregressions. *Review of Economics and Statistics* 97(2), 436–451.
- Granziera, E., H. R. Moon, and F. Schorfheide (2018). Inference for VARs Identified with Sign Restrictions. *Quantitative Economics* 9(3), 1087–1121.
- Herbst, E. and F. Schorfheide (2014). Sequential Monte Carlo Sampling for DSGE Models. *Journal of Applied Econometrics* 29(7), 1073–1098.
- Kilian, L. and D. P. Murphy (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.
- Kilian, L. and D. P. Murphy (2014). The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. *Journal of Applied Econometrics* 29(3), 454–478.
- Koop, G. (2013). Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics* 28(2), 177–203.
- Leeper, E. M., C. A. Sims, and T. Zha (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity* 1996(2), 1–78.

- Murray, I., R. Adams, and D. MacKay (2010). Elliptical Slice Sampling. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 9*, 541–548.
- Natarovskii, V., D. Rudolf, and B. Sprungk (2021). Geometric Convergence of Elliptical Slice Sampling. *Proceedings of the 38th International Conference on Machine Learning 139*, 7969–7978.
- Neal, R. M. (2003). Slice Sampling. *The Annals of Statistics 31*(3), 705–767.
- Read, M. and D. Zhu (2025). Fast Posterior Sampling in Tightly Identified SVARs Using ‘Soft’ Sign Restrictions. *Federal Reserve Bank of Australia Working Paper* (2025-03).
- Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.
- 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.
- Uhlig, H. (2017). Shocks, Sign Restrictions, and Identification. *Advances in Economics and Econometrics 2*, 95.
- Vats, D., J. M. Flegal, and G. L. Jones (2019). Multivariate Output Analysis for Markov Chain Monte Carlo. *Biometrika 106*(2), 321–337.