FEDERAL RESERVE BANK PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Uniform Priors for Impulse Responses

Jonas E. Arias
Federal Reserve Bank of Philadelphia Research Department

Juan F. Rubio-Ramírez
Emory University, Federal Reserve Bank of Atlanta, and
Federal Reserve Bank of Philadelphia Research Department Visiting Scholar

Daniel F. Waggoner
Emory University and Federal Reserve Bank of Atlanta, Emeritus

WP 22-30

PUBLISHED
September 2022

REVISED
October 2024

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2022.30

Uniform Priors for Impulse Responses

Jonas E. Arias
Federal Reserve Bank of Philadelphia

Juan Rubio-Ramírez*
Emory University
Federal Reserve Bank of Atlanta

Daniel F. Waggoner
Emory University and Federal Reserve Bank of Atlanta, Emeritus

October 6, 2024

Abstract

There has been a call for caution regarding the standard procedure for Bayesian inference in set-identified structural vector autoregressions on the grounds that the common practice of using a uniform prior over the set of orthogonal matrices induces a non-uniform prior for individual impulse responses or other quantities of interest. This paper challenges this call by formally showing that when the focus is on joint inference, the uniform prior over the set of orthogonal matrices is not only sufficient but also necessary for inference based on a uniform joint prior distribution over the identified set for the vector of impulse responses. In addition, we show how to conduct inference based on a uniform joint prior distribution for the vector of impulse responses.

JEL classification: C11, C32.
Keywords: structural vector autoregressions, impulse responses, joint inference, uniform priors.

We thank Jim Hamilton, Lutz Kilian, Mikkel Plagborg-Møller, Mark Watson, and Christian Wolf 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

Structural vector autoregressions (SVARs) identified with sign restrictions are a popular approach to estimating dynamic causal effects in macroeconomics. Many researchers use variants of the method proposed by Uhlig (2005) to conduct Bayesian inference.[1] This conventional method can be used to independently draw from any posterior distribution over the parameterization of interest subject to the identification restrictions. Typically, the parameterization of interest consists of the impulse responses, and the posterior is conjugate.

Within this framework, common practice independently draws from a conjugate uniform-normal-inverse-Wishart posterior distribution over the orthogonal reduced-form parameters and transforms the draws into the objects of interest. A central ingredient of such an approach is the uniform prior over the set of orthogonal matrices with respect to the Haar measure (see Halmos (1950)). The normal-inverse-Wishart part of this prior is viewed as uncontroversial—the Minnesota prior and the "weak" prior defined in Uhlig (2005) are the most popular choices. Some researchers have criticized this conventional approach (see, e.g., Baumeister and Hamilton (2015); Watson (2020)) and suggest caution when using it in applied work.

This paper accomplishes several objectives. First, Baumeister and Hamilton (2015) and Watson (2020) express concern about the fact that the uniform prior over the set of orthogonal matrices induces non-uniform prior distributions over the identified sets of individual impulse responses because the prior and posterior coincide over identified sets.[2] While this fact could be an issue when the number of observations is large enough that reduced-form parameter uncertainty can be disregarded, Inoue and Kilian (2022b) argue that this concern may not be salient when working with tightly identified models based on many sign restrictions and possibly narrative restrictions, as is often the case in applied work. We further ease

[1] See Faust (1998), Uhlig (1998), Canova and De Nicoló (2002), and Rubio-Ramírez, Waggoner, and Zha (2010) for related work and extensions of the approach.

[2] By individual impulse response, we mean the response of a single variable to a single shock at a single horizon.

this concern by showing that the uniform prior over the set of orthogonal matrices induces uniform joint prior and posterior distributions over the identified set for the vector of impulse responses. This result is an "if and only if" statement that holds for any prior distribution for the reduced-form parameters. While uniform joint prior and posterior distributions over the identified set for the vector of impulse responses are not required features, they imply that only the identifying restrictions will set apart observationally equivalent vectors of impulse responses. The vector of impulse responses contains the responses across horizons and shocks and it is the object of interest of several studies arguing that joint distributions are better suited to capture the shape and co-movement of the responses (e.g., Sims and Zha (1999); Fry and Pagan (2011); Inoue and Kilian (2013, 2016, 2019, 2022a,b); Lütkepohl, Staszewska-Bystrova, and Winker (2015a,b, 2018); Kilian and Lütkepohl (2017); Bruder and Wolf (2018); Montiel Olea and Plagborg-Møller (2019), among others).

Second, we show how to construct a uniform joint prior distribution for the vector of impulse responses for models identified with sign restrictions and how to conduct joint posterior inference based on this prior using the conventional approach. In particular, we show that a uniform joint prior distribution for the vector of impulse responses induces a prior for the orthogonal reduced-form parameters such that (1) it is independent between the reduced-form parameters and the orthogonal matrices, (2) the prior for the reduced-form parameters has a particular (model dependent) form, and (3) the prior over the set of orthogonal matrices is uniform. This theoretical result is also an "if and only if" statement. Interestingly, the induced prior distribution for the reduced-form parameters differs from the standard Minnesota prior, and it is similar in spirit to (although also different than) the "weak" prior described in Uhlig (2005). We show that the induced prior for the orthogonal reduced-form parameters defines a uniform-normal-inverse-Wishart posterior distribution over the orthogonal reduced-form parameters. This allows us to use the conventional approach to draw from the joint posterior distribution for the vector of impulse responses implied by a uniform joint prior distribution for the vector of impulse responses. Because of the uniform prior distribution over the set of orthogonal matrices, the conventional approach also induces uniform joint prior and posterior distributions over the identified set for the vector of impulse responses.

To illustrate our theoretical findings, we examine Watson's (2020) empirical example using a uniform joint prior distribution for the vector of impulse responses. Based on the methods in Inoue and Kilian (2022a), we find that the joint credible sets for the vector of impulse responses obtained under this prior are similar but wider than those obtained under the uniform-normal-inverse-Wishart prior distribution for orthogonal reduced-form parameters associated with the standard Minnesota prior. In line with the findings in Inoue and Kilian (2022b), our results suggest that imposing tighter identifying restrictions helps when evaluating joint posteriors. This message gets stronger when considering a uniform joint prior distribution for the vector of impulse responses.

Finally, we generalize our analysis to a broader class of objects of interest.[3] We show how to implement a uniform joint prior distribution for the vector of objects of interest using the conventional approach. For example, imagine a two-variable (price and quantity) stylized model of demand and supply with a uniform joint prior distribution for the objects of interest consisting of some elasticities and the standard deviations of structural shocks. Each particular vector of objects of interest induces a different prior distribution for the orthogonal reduced-form parameters. This induced prior is also model-dependent but need not be uniform over the set of orthogonal matrices conditional on the reduced-form parameters. In the latter case, it is necessary to add an importance sampling step to the conventional method to draw from the induced joint posterior distribution for the vector of objects of interest.

The structure of the paper is as follows. Section 2 describes the conventional method and Section 3 proves that it implies a uniform joint prior distribution over the identified set for the vector of impulse responses. Section 4 shows how to define a uniform joint prior

[3]See Section 3.3 for a formal definition of the class of objects of interest.

distribution for the vector of impulse responses and how to adapt the conventional method to implement it. Section 5 illustrates our approach using the model in Watson (2020). Section 6 concludes.

2 The Conventional Approach

Consider a reduced-form VAR of the form

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

where $\mathbf{y}_t$ is an $n \times 1$ vector of endogenous variables, $\boldsymbol{u}_t$ is an $n \times 1$ vector of reduced-form shocks, $\mathbf{x}_t' = \left[\mathbf{y}_{t-1}' \ \cdots \ \mathbf{y}_{t-p}' \ 1\right]$, $\mathbf{B} = \left[\mathbf{B}_1' \ \cdots \ \mathbf{B}_p' \ \ \mathbf{d}'\right]'$ is an $m \times n$ matrix with $m = np + 1$, $\mathbf{B}_\ell$ is an $n \times n$ matrix of parameters for $1 \leq \ell \leq p$, $\mathbf{d}$ is a $1 \times n$ vector of parameters, p is the lag length, and T is the sample size. The vector $\mathbf{u}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, is Gaussian with mean zero and covariance matrix $\boldsymbol{\Sigma}$. We call $(\mathbf{B}, \boldsymbol{\Sigma})$ the reduced-form parameters.

Let $\mathbf{u}_t = \mathbf{L}_0 \boldsymbol{\varepsilon}_t$ for $1 \leq t \leq T$, where $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_n)$ are structural shocks, $\mathbf{L}_0$ is an $n \times n$ invertible matrix that represents impulse responses at horizon zero, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Given $\mathbf{L}_0$ and $\mathbf{B}$, it is possible to obtain the impulse responses beyond horizon zero recursively, as

$$\mathbf{L}_\ell = \sum_{k=1}^{\min\{\ell,p\}} \mathbf{B}_k' \mathbf{L}_{\ell-k}, \text{ for } \ell > 0. \tag{2}$$

We combine the impulse responses from horizons one through p and the constant term $\mathbf{c} = \mathbf{d}\left(\mathbf{L}_0^{-1}\right)'$ into a single matrix, $\mathbf{L}_+ = \left[\mathbf{L}_1' \ \cdots \ \mathbf{L}_p' \ \ \mathbf{c}'\right]'$, 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 IR parameters. Importantly, when referring to these parameters in vector form we will use the term vector of impulse responses.

The discussion above implicitly defines a mapping from the IR parameters to the reduced-form parameters. In particular, we have that $\mathbf{\Sigma} = \mathbf{L}_0 \mathbf{L}_0'$,

$$\mathbf{B}_\ell = \left(\mathbf{L}_\ell \mathbf{L}_0^{-1}\right)' - \sum_{k=1}^{\ell-1} \left(\mathbf{L}_{\ell-k} \mathbf{L}_0^{-1}\right)' \mathbf{B}_k, \text{ for } 1 \leq \ell \leq p, \text{ and } \mathbf{d} = \mathbf{c}\,\mathbf{L}_0'. \tag{3}$$

In the class of linear Gaussian models under analysis, it is well known that $(\mathbf{L}_0, \mathbf{L}_+)$ and $(\tilde{\mathbf{L}}_0, \tilde{\mathbf{L}}_+)$ are observationally equivalent if and only if $\mathbf{L}_0 = \tilde{\mathbf{L}}_0 \mathbf{Q}$ and $\mathbf{L}_+ = \tilde{\mathbf{L}}_+ \mathbf{Q}$ for some $\mathbf{Q} \in \mathcal{O}(n)$, which is the set of all $n \times n$ orthogonal matrices; see Rubio-Ramírez, Waggoner, and Zha (2010). Hence, the IR parameters are not identified.

This suggests that given any decomposition of the covariance matrix $\mathbf{\Sigma}$ satisfying $h(\mathbf{\Sigma})' h(\mathbf{\Sigma}) = \mathbf{\Sigma}$, we can define a mapping from $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ to $(\mathbf{L}_0, \mathbf{L}_+)$. We will take h to be the upper triangular Cholesky decomposition normalized so that the diagonal is positive. Thus

$$\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \Big(\underbrace{h(\mathbf{\Sigma})'\,\mathbf{Q}}_{\mathbf{L}_0}, \underbrace{\left[\mathbf{L}_1(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})' \;\cdots\; \mathbf{L}_p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})' \quad \mathbf{Q}'(h(\mathbf{\Sigma})^{-1})'\mathbf{d}'\right]'}_{\mathbf{L}_+}\Big), \tag{4}$$

where $\mathbf{L}_\ell(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ for $1 \leq \ell \leq p$ is implicitly defined in Equation (2). The function ϕ is invertible, and both ϕ and its inverse are differentiable. Hence, there exists a diffeomorphism between the IR parameters and the orthogonal reduced-form parameters that we will exploit in the rest of the paper.

2.1 The Priors, the Posteriors, and the Algorithm

The conventional method uses a normal-inverse-Wishart (NIW) distribution prior for $(\mathbf{B}, \mathbf{\Sigma})$. Denote the prior by $NIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$. As shown in Uhlig (1994, 2005), this prior is conjugate and the posterior distribution over the reduced-form parameters is $NIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$, where $\tilde{\nu} = T + \bar{\nu}$, $\tilde{\mathbf{\Omega}} = (\mathbf{X}'\,\mathbf{X} + \bar{\mathbf{\Omega}}^{-1})^{-1}$, $\tilde{\mathbf{\Psi}} = \tilde{\mathbf{\Omega}}(\mathbf{X}'\,\mathbf{Y} + \bar{\mathbf{\Omega}}^{-1}\bar{\mathbf{\Psi}})$, and $\tilde{\mathbf{\Phi}} = \mathbf{Y}'\mathbf{Y} + \bar{\mathbf{\Phi}} + \bar{\mathbf{\Psi}}'\bar{\mathbf{\Omega}}^{-1}\bar{\mathbf{\Psi}} - \tilde{\mathbf{\Psi}}'\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{\Psi}}$, for $\mathbf{Y} = [\mathbf{y}_1 \;\cdots\; \mathbf{y}_T]'$ and $\mathbf{X} = [\mathbf{x}_1 \;\cdots\; \mathbf{x}_T]'$. If we use a uniform prior distribution over the set of orthogonal matrices, then the resulting prior distribution for $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ is uniform-normal-inverse-Wishart (UNIW) and we denote it by $UNIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$. This prior is also conjugate, and the posterior distribution is $UNIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$. Because the UNIW family of distributions is conjugate over $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, it implies a family of distributions over $(\mathbf{L}_0, \mathbf{L}_+)$ that it is conjugate. This is because if the prior and posterior densities have the same functional form over $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, then, because the volume element associated with ϕ will be the same for the prior and posterior densities, the induced prior and posterior densities for $(\mathbf{L}_0, \mathbf{L}_+)$ will also have the same functional form.[4]

There are several routines for making independent draws from any NIW distribution over $(\mathbf{B}, \mathbf{\Sigma})$. Independent draws from the uniform distribution over $\mathcal{O}(n)$ are based on Theorem 3.2 of Stewart (1980), summarized by Proposition 1.

Proposition 1. *Let $\boldsymbol{X}$ be an $n \times n$ random matrix with each element having an independent standard normal distribution. Let $\boldsymbol{X} = \boldsymbol{QR}$ be the QR decomposition of $\boldsymbol{X}$ with the diagonal of $\boldsymbol{R}$ normalized to be positive. The matrix $\mathbf{Q}$ is orthogonal and is drawn from the uniform distribution over $\mathcal{O}(n)$.*

This discussion justifies Algorithm 1 to draw from the conjugate posterior distribution over $(\mathbf{L}_0, \mathbf{L}_+)$ conditional on the sign restrictions. This algorithm can be found in Uhlig (2005) for a single shock and is extended to a set of shocks in Rubio-Ramírez, Waggoner, and Zha (2010).

Algorithm 1. *The following algorithm independently draws from the conjugate posterior distribution over $(\mathbf{L}_0, \mathbf{L}_+)$ conditional on the sign restrictions.*

1. *Draw $(\mathbf{B}, \mathbf{\Sigma})$ independently from $NIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$.*
2. *Draw $\mathbf{Q}$ independently from the uniform distribution over $\mathcal{O}(n)$.*
3. *Keep $(\mathbf{L}_0, \mathbf{L}_+) = \phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ if the sign restrictions are satisfied.*
4. *Return to Step 1 until the required number of draws has been obtained.*

[4] For a formal definition of volume element, see Chapter 5 in Spivak (1965).

Notably, throughout the rest of the paper, all densities will be with respect to the volume measure, even though sometimes we will not explicitly state it. When working with impulse responses or $\mathbf{B}$, the volume measure will be equal to the Lebesgue measure. However, when we are working with symmetric and positive definite matrices or orthogonal matrices, the volume measure will not be Lebesgue. In particular, the volume measure over orthogonal matrices is a Haar measure.

3 Conditional Joint Prior for Impulse Responses

A central ingredient underlying the conventional approach summarized in Section 2 is the uniform prior distribution over the set of orthogonal matrices with respect to the Haar measure. This prior distribution has been criticized by Baumeister and Hamilton (2015) and Watson (2020) because (1) it implies that some marginal prior distributions over the identified sets are non-uniform, and (2) posterior inference is routinely dominated by such non-uniform prior. Several studies such as Wolf (2020) and Giacomini and Kitagawa (2021) have echoed this critique, and as a consequence, there is a growing call for caution for any of the results obtained by the conventional method.

The marginal prior distributions over the identified sets are obtained by replacing Step 1 with a fixed value of the reduced-form parameters and then marginalizing out all but an individual impulse response. We will refer to the prior distributions obtained this way as the conditional prior distributions for individual impulse responses to emphasize that they do condition on the reduced-form parameters. Inoue and Kilian (2022b) draw attention to the fact that the conditional prior distributions for individual impulse responses may give an incomplete picture of the priors embodied in the conventional approach. Fixing the value of the reduced-form parameters eliminates any uncertainty about $(\mathbf{B}, \boldsymbol{\Sigma})$, whereas the conventional approach postulates an NIW distribution prior. In their examples, when uncertainty about the reduced-form parameters is taken into account, the conventional method does not imply that posterior inference is routinely dominated by the prior.

Many questions in SVAR analysis entail examining the shape of impulse responses of multiple variables to several shocks at various horizons. Joint inference on impulse responses, which considers dependencies across these dimensions, offers a more suitable approach for addressing these cases than marginal inference. Consequently, we adopt the perspective of a researcher interested in joint inference and assess the implications of using a uniform prior over the set of orthogonal matrices. In this section, we condition on the reduced-form parameters. Often, we will refer to this prior as conditional joint prior distribution for the vector of impulse responses because it is obtained by conditioning on the reduced-form parameters. We will consider unconditional priors in the next section. Because the posterior reproduces the prior over the identified set, a uniform joint prior distribution over the identified set for the vector of impulse responses ensures the researcher that only the identifying restrictions will set apart observationally equivalent vectors of impulse responses. In this section, we show that the uniform prior distribution over the set of orthogonal matrices with respect to the Haar measure is both a necessary and sufficient condition for having a uniform conditional joint prior distribution for the vector of impulse responses. We will first show an illustrative example and then move to the general results.

3.1 An Illustrative Simple Example

Let us consider a simple example. To reduce the number of parameters, we assume that there are no lags or constant term. In this case, the only impulse response is $\mathbf{L}_0$, and the only reduced-form parameter is $\mathbf{\Sigma}$. The support of the joint prior distribution over the identified set for the vector of impulse responses is of the form

$$\underbrace{\begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \end{bmatrix}}_{\mathbf{L}_0} = \underbrace{\begin{bmatrix} \hat{\ell}_{11} & 0 \\ \hat{\ell}_{21} & \hat{\ell}_{22} \end{bmatrix}}_{\hat{\mathbf{L}}_0} \underbrace{\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ (-1)^i \sin(\theta) & (-1)^{i+1} \cos(\theta) \end{bmatrix}}_{\mathbf{Q}}, \tag{5}$$

where i is either zero or one, $-\pi \le \theta < \pi$, and $\hat{\mathbf{L}}_0\hat{\mathbf{L}}_0' = \mathbf{\Sigma}$ with both $\hat{\ell}_{11}$ and $\hat{\ell}_{22}$ positive. A direct computation shows that for any $\mathbf{L}_0$ given by Equation (5), its norm is $\hat{r} = \sqrt{\hat{\ell}_{11}^2 + \hat{\ell}_{22}^2 + \hat{\ell}_{21}^2}$ and it lies in one of the two two-dimensional subspaces of $\mathbb{R}^4$ with bases

$$\hat{\mathbf{L}}_{\cos}^{i} = \begin{bmatrix} \hat{\ell}_{11} & 0 \\ \hat{\ell}_{21} & (-1)^{i+1}\hat{\ell}_{22} \end{bmatrix} \text{ and } \hat{\mathbf{L}}_{\sin}^{i} = \begin{bmatrix} 0 & \hat{\ell}_{11} \\ (-1)^{i}\hat{\ell}_{22} & \hat{\ell}_{21} \end{bmatrix}, \tag{6}$$

for $i = 0, 1$. This follows from the fact that $\mathbf{L}_0 = \cos(\theta)\hat{\mathbf{L}}_{\cos}^{i} + \sin(\theta)\hat{\mathbf{L}}_{\sin}^{i}$. Also, the vectors $\hat{\mathbf{L}}_{\cos}^{i}$ and $\hat{\mathbf{L}}_{\sin}^{i}$ are perpendicular and length $\hat{r}$. Thus, the set of all $\mathbf{L}_0$ of this form will be two *circles* in $\mathbb{R}^4$ of radius $\hat{r}$.

The joint prior distribution over the identified set for the vector of impulse responses is completely determined by the joint distribution over (θ, i), which can be written as $p(\theta, i) = p(\theta)p(i|\theta)$. Since $\ell_{11} = \hat{\ell}_{11}\cos(\theta)$ and $\ell_{12} = \hat{\ell}_{11}\sin(\theta)$, the conditional prior densities of the individual ℓ_{11} and ℓ_{12} are given by

$$p(\ell_{11}) = \frac{p(\cos^{-1}(\ell_{11}/\hat{\ell}_{11})) + p(-\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))}{\hat{\ell}_{11}\sin(\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))} \text{ and} \tag{7}$$

$$p(\ell_{12}) = \frac{p(\sin^{-1}(\ell_{12}/\hat{\ell}_{11})) + p(\operatorname{sgn}(\ell_{12}/\hat{\ell}_{11})\pi - \sin^{-1}(\ell_{12}/\hat{\ell}_{11}))}{\hat{\ell}_{11}\cos(\sin^{-1}(\ell_{12}/\hat{\ell}_{11}))}, \tag{8}$$

where $\operatorname{sgn}(\cdot)$ is 1 if the argument is positive and -1 otherwise. We provide the derivations of these in Appendix B. We compute and plot the conditional prior densities of the individual ℓ_{11} and ℓ_{12} and the joint prior distribution over the identified set for the vector of impulse responses in two cases. In Case (1), we set a uniform prior distribution over the set of orthogonal matrices with respect to the Haar measure. In this case, the joint prior distribution over the identified set for the vector of impulse responses is uniform, while the conditional densities of the individual ℓ_{11} and ℓ_{12} are not. In Case (2), we choose the prior over the set of orthogonal matrices such that the conditional density of the individual ℓ_{11} is uniform. In this case, neither the conditional densities of the individual ℓ_{12} nor the joint prior distribution over the identified set for the vector of impulse responses are uniform.

Case (1): The conditional joint distribution over $\mathbf{L}_0$ is uniform for every $\boldsymbol{\Sigma}$. In this first case, we set the distribution over $\mathbf{Q}$ to be uniform with respect to the volume measure, which is arc length. The properly scaled density over (θ, i) must be $p(\theta, i) = p(\theta)p(i|\theta) = (1/(2\pi))(1/2)$. By Equations (7) and (8), the conditional marginal densities are $p(\ell_{11}) = \frac{1}{\pi}(\hat{\ell}_{11}^2 - \ell_{11}^2)^{-\frac{1}{2}}$ and $p(\ell_{12}) = \frac{1}{\pi}(\hat{\ell}_{11}^2 - \ell_{12}^2)^{-\frac{1}{2}}$. We provide derivations of these in Appendix B.

Case (2): The conditional distribution of ℓ_{11} is uniform over $[-\hat{\ell}_{11}, \hat{\ell}_{11}]$. If the conditional distribution of ℓ_{11} is uniform, then $p(\ell_{11}) = 1/(2\hat{\ell}_{11})$ and by Equation (7), the distribution of θ must satisfy $p(\theta) + p(-\theta) = \sin(\theta)/2$ for $0 \leq \theta < \pi$. Is there a choice of $p(\theta)$ so that the conditional distribution of ℓ_{12} is uniform? Appendix B shows that there is no choice of $p(\theta)$ such that the conditional distribution of ℓ_{11} and ℓ_{12} are both uniform. This illustrates a point already made by Baumeister and Hamilton (2015): One cannot have uniform distributions over the identified sets of all of the individual impulse responses. We choose $p(\theta) = |\sin(\theta)/4|$ and $p(i|\theta) = 1/2$, which implies that the conditional distribution of ℓ_{11} is uniform and probably does the least violence to the conditional distribution of ℓ_{12}. In this case $p(\ell_{12}) = |\ell_{12}|/(2\hat{\ell}_{11}(\hat{\ell}_{11}^2 - \ell_{12}^2)^{\frac{1}{2}})$, as will be shown in Appendix B.

Figure 1 shows the joint distribution. The support of the distribution of $\mathbf{L}_0$, conditional on $\boldsymbol{\Sigma}$, consists of two circles in $\mathbb{R}^4$ of radius $\hat{r}$. We plot the conditional joint density over one of the two circles. In Case (1), the conditional joint distribution is uniform. In Case (2), this is not the case; the density goes to zero at certain points.

Figure 2 plots the conditional densities of ℓ_{11} and ℓ_{12} for the two cases. The dotted lines in Figure 2 are the conditional densities in Case (1), and the solid lines correspond to Case (2). For Case (2), the conditional distribution of ℓ_{11} is uniform by construction, but the conditional distribution of ℓ_{12} is farther from uniform than it is in Case (1). Figure 2 illustrates the dangers of analyzing marginal densities. Therefore, Case (1) shows that a uniform prior for

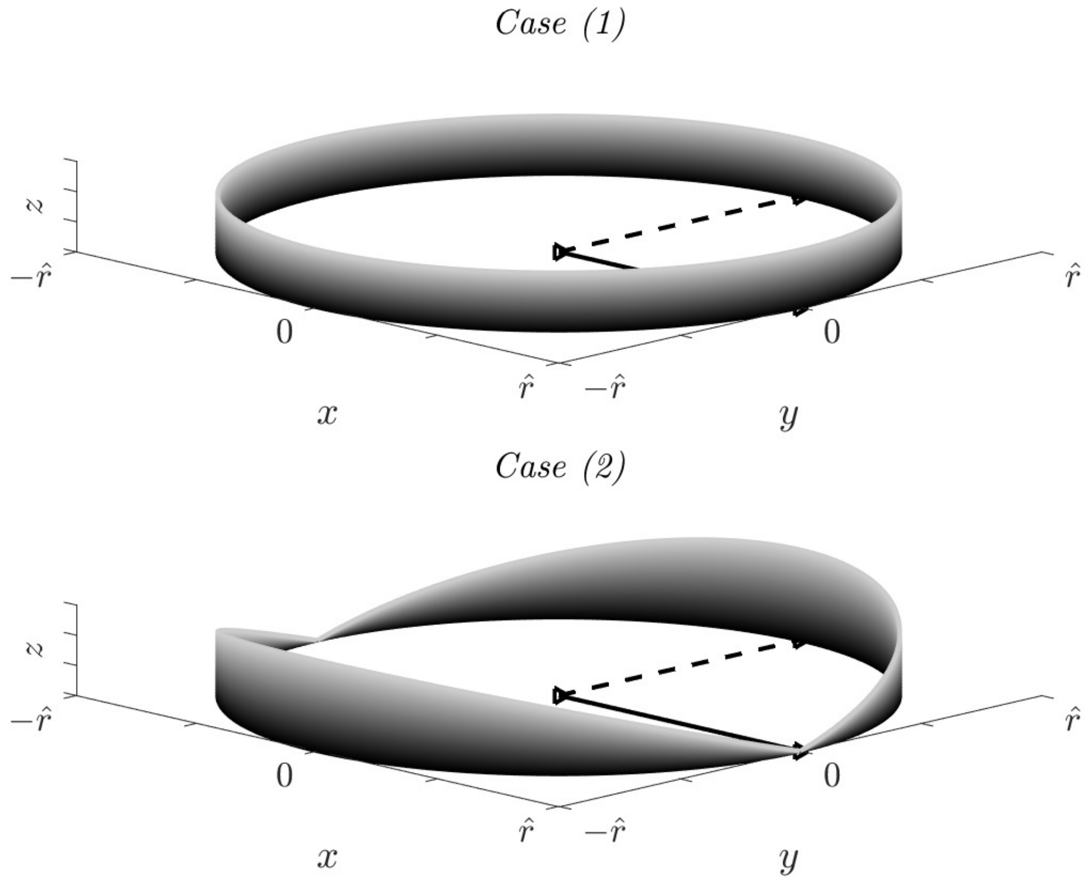

Figure 1: Conditional joint density for Cases (1) and (2). The solid vector is $\hat{\mathbf{L}}^i_{\cos} \in \mathbb{R}^4$, the dotted vector is $\hat{\mathbf{L}}^i_{\sin} \in \mathbb{R}^4$, and $z = p(\mathbf{L}_0)$, with $\mathbf{L}_0 = (x\hat{\mathbf{L}}^i_{\cos} + y\hat{\mathbf{L}}^i_{\sin})/\hat{r}$.

Q implies a uniform joint prior distribution over the identified set for the vector of impulse responses, although a researcher who analyzes conditional prior distributions for individual impulse responses may conclude otherwise. Case (2) implies that one can choose priors for **Q** such that the conditional density of ℓ_{11} is uniform. This prior for **Q** is not uniform and will imply non-uniform conditional densities of ℓ_{12} and non-uniform joint prior distribution over the identified set for the vector of impulse responses.

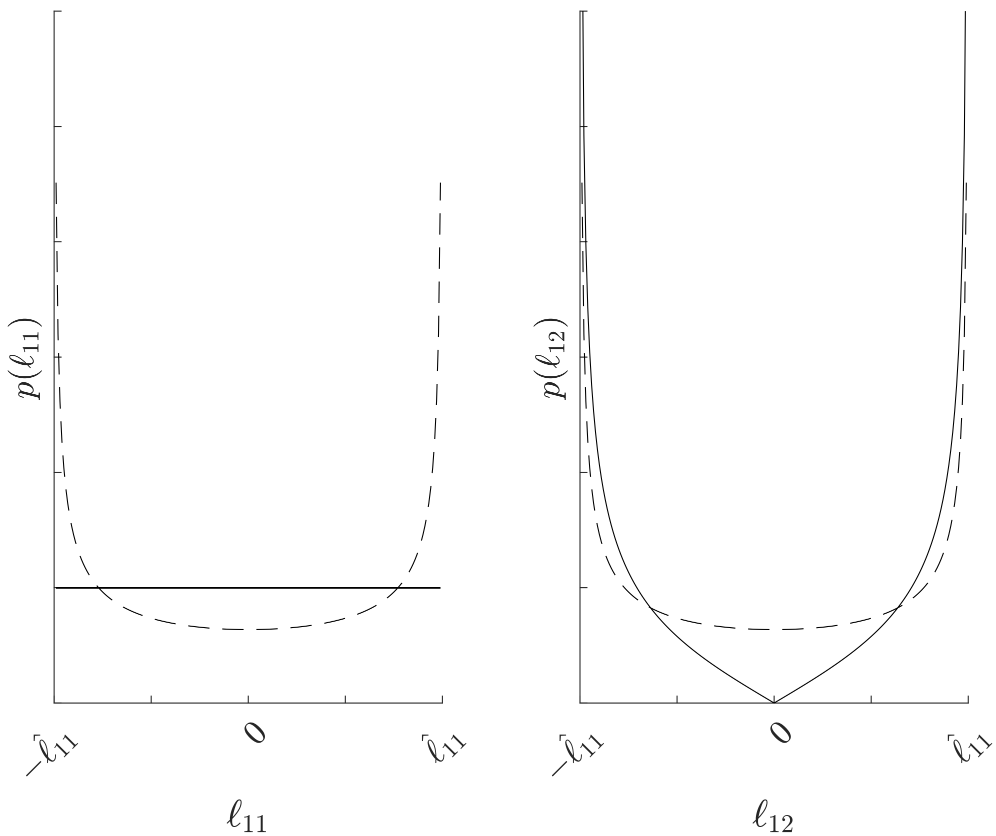

Figure 2: The dotted lines are the conditional densities of ℓ_{11} and ℓ_{12} for Case (1). The solid lines are the conditional densities of ℓ_{11} and ℓ_{12} for Case (2).

3.2 General Results for Impulse Responses

Are there distributions over the IR parameters such that the conditional joint prior distribution for the vector of impulse responses is uniform? The answer is yes, and the results to follow give the conditions required for this to be the case. Interestingly, the conventional method implies a uniform joint prior distribution over the identified set for the vector of impulse responses.

Before stating the proposition, we need a precise understanding of what it means to condition on the reduced-form parameters. Given the reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$, the support of the joint distribution of the IR parameters conditional on $(\mathbf{B}, \mathbf{\Sigma})$ is

$$\mathcal{P}(\mathbf{B}, \mathbf{\Sigma}) = \{(\mathbf{L}_0, \mathbf{L}_+) = \phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) \mid \text{for every } \mathbf{Q} \in \mathcal{O}(n)\},$$

which is a smooth manifold because $\mathcal{O}(n)$ is a smooth manifold and the invertible function ϕ is continuously differentiable. The manifold structure induces a natural measure over $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$, which is called the volume measure.[5] For example, the volume measure over one-dimensional manifolds is the arc length, and the volume measure over two-dimensional manifolds is the surface area. If $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is a density over the IR parameters, then the density conditional on $(\mathbf{B}, \mathbf{\Sigma})$ with respect to the volume measure over $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$ will be proportional to $\pi(\mathbf{L}_0, \mathbf{L}_+)$. The volume measure is the only measure, up to a scale factor, that has this property. In this sense, the volume measure is the natural one. Thus, conditional on $(\mathbf{B}, \mathbf{\Sigma})$, the density with respect to the volume measure over $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$ will be uniform if and only if $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is constant over $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$.

The volume and Haar measures over $\mathcal{O}(n)$ are related. A Haar measure is any measure over $\mathcal{O}(n)$ that is invariant under right multiplication by an orthogonal matrix and is unique up to a scale factor. The volume measure over $\mathcal{O}(n)$ has this property and thus is a Haar measure. With these ingredients, we can now show the following proposition.

Proposition 2. *For every density over the IR parameters with respect to the Lebesgue measure, the density with respect to the volume measure over* $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$*, conditional on* $(\mathbf{B}, \mathbf{\Sigma})$*, is uniform for every* $(\mathbf{B}, \mathbf{\Sigma})$ *if and only if the induced distributions over the orthogonal reduced-form parameters* $(\mathbf{B}, \mathbf{\Sigma})$ *and* $\mathbf{Q}$ *are independent and the distribution of* $\mathbf{Q}$ *is uniform with respect to the Haar measure.*

Proof. See Appendix A. □

Thus, for every density over the IR parameters with respect to the Lebesgue measure, the density with respect to the volume measure over $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$ is constant over observationally equivalent vectors of impulse responses if and only if the induced distributions over the orthogonal reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$ and $\mathbf{Q}$ are independent and the distribution of $\mathbf{Q}$ is uniform with respect to the Haar measure. Proposition 2 essentially follows from

[5] See Munkres (1991), Chapter 5, for details of how the volume measure is defined over manifolds.

the fact that the volume element for the mapping ϕ does not depend on $\mathbf{Q}$.[6] A similar result will hold for any parameterization such that the volume element of the mapping to the orthogonal reduced-form parameters does not depend on $\mathbf{Q}$, for example, the standard structural parameterization. All volume elements in this paper will be computed using Theorem 21.3 in Munkres (1991). One could claim the same in terms of observationally equivalence. The proof in terms of observationally equivalence is also simple. Two impulse responses are observationally equivalent if and only if there exists a value of the reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$ such that both of the impulse responses lie in the support of the distribution conditional on $(\mathbf{B}, \mathbf{\Sigma})$.

Because they are "if and only if" statements, Proposition 2 brings to the fore the virtue of joint distributions over the IR parameters that induce a distribution over the orthogonal reduced-form parameters such that the distribution over the set of orthogonal matrices is uniform.[7] Consequently, to have a uniform joint prior distribution over the identified set for the vector of impulse responses one *must* use a prior distribution over the set of orthogonal matrices that is uniform. Any other choice of prior over the set of orthogonal matrices *will* imply a non-uniform joint prior distribution over the identified set for the vector of impulse responses. This is true for any prior distribution over the reduced-form parameters; hence, researchers can choose any prior distribution over the reduced-form parameters that respects their beliefs about the data.

The results in this section are relevant for the robust methodology developed by Giacomini and Kitagawa (2021). First, only a uniform prior over the set of orthogonal matrices induces a uniform prior over observationally equivalent vectors of impulse responses, and hence, only in this case can researchers claim that the identification problem is only resolved utilizing sign restrictions, preserving the virtues that made inference based on sign restrictions a practical tool in empirical macroeconomics. Second, while the analysis in Giacomini and Kitagawa

[6]An analytical expression for this volume element will be obtained in Proposition 4.

[7]If a distribution over the orthogonal reduced-form parameters is such that the distribution over the set of orthogonal matrices is uniform for all reduced-form parameters, then the reduced-form parameters and the orthogonal matrices must be independent.

(2021) could potentially be extended to the case of joint inference, such an extension is challenging and, hence, our propositions offer useful insights to researchers concerned with the role of the prior when conducting joint posterior inference.

We have shown that the conventional method does imply a uniform joint prior distribution over the identified set for the vector of impulse responses. Later, we will eliminate the conditionality on the reduced-form parameters and show that it is possible to have a uniform joint prior distribution for the vector of impulse responses and that it can be implemented by the conventional method.

3.3 Extension to Objects of Interest

In empirical work, the object of interest does not always need to be the vector of impulse responses. We now extend the results above to general objects of interest. Denote the vector of objects of interest by $\mathbf{\Upsilon}$ and the transformation from $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ to $\mathbf{\Upsilon}$ by ϕ_o. In our class of objects of interest, we assume that ϕ_o is a diffeomorphism and $\mathbf{\Upsilon}$ is an open subset of $\mathbb{R}^{n^2+nm}$, and we use the Lebesgue measure over $\mathbf{\Upsilon}$. To illustrate the type of objects of interest that our approach can accommodate, let us reproduce the example described in Section 1. We define our vector of objects of interest as $\mathbf{\Upsilon} = (\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4) \in \mathbb{R}^4$, where $\upsilon_1 = \ell_{22}/\ell_{12}$, $\upsilon_2 = \ell_{21}/\ell_{11}$, $\upsilon_3 = (\ell_{12}\ell_{21} - \ell_{11}\ell_{22})/\ell_{12}$, and $\upsilon_4 = (\ell_{11}\ell_{22} - \ell_{12}\ell_{21})/\ell_{11}$, with ℓ_{ij} denoting the entry (i, j) of the matrix of contemporaneous impulse responses $\mathbf{L}_0$. Consequently, υ_1 and υ_2 are some elasticities, and υ_3 and υ_4 are some other parameters of interest, such as standard deviations of structural shocks. Clearly, in this example there is a diffeomorphism between $\mathbf{L}_0$ and $\mathbf{\Upsilon}$, therefore, there is a diffeomorphism between $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ and $\mathbf{\Upsilon}$.

Let us consider a joint prior distribution over the identified set for the vector of objects of interest. Often, we will refer to this prior as the conditional joint prior distribution for the vector of objects of interest.[8] To characterize the prior for the orthogonal reduced-form

[8]The joint prior distribution over the identified set for the vector of objects of interest (or equivalently the conditional joint prior distribution for the vector of objects of interest) is obtained by conditioning on the reduced-form parameters.

parameters induced by a uniform joint prior distribution over the identified set for the vector of objects of interest, we first notice that the support of the joint distribution of the vector of objects of interest conditional on $(\mathbf{B}, \mathbf{\Sigma})$ is the smooth manifold.

$$\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma}) = \{\mathbf{\Upsilon} = \phi_o(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) \mid \text{for every } \ \mathbf{Q} \in \mathcal{O}(n)\},$$

where, as in the case of $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$, the smooth manifold $\mathcal{O}(n)$ induces the volume measure over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$. If $\pi(\mathbf{\Upsilon})$ is a density over the objects of interest, then the density conditional on $(\mathbf{B}, \mathbf{\Sigma})$ with respect to the volume measure over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$ will be proportional to $\pi(\mathbf{\Upsilon})$. Thus, conditional on $(\mathbf{B}, \mathbf{\Sigma})$, the density with respect to the volume measure over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$ will be uniform if and only if $\pi(\mathbf{\Upsilon})$ is constant over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$.

Proposition 3. *For every density over the objects of interest with respect to the Lebesgue measure, the density with respect to the volume measure over* $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$*, conditional on* $(\mathbf{B}, \mathbf{\Sigma})$*, is uniform for every* $(\mathbf{B}, \mathbf{\Sigma})$ *if and only if the induced distribution over the orthogonal reduced-form parameters is such that* $\pi(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$ *is proportional to* $v_{\phi_o}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$*, where* v_{ϕ_o} *is the volume element induced by* ϕ_o*.*

Proof. See Appendix A. □

Proposition 3 is an extension of Proposition 2 for general objects of interest. Proposition 2 implies that the induced prior over the set of orthogonal matrices is uniform for impulse responses because the volume element does not depend on $\mathbf{Q}$. In the case of general objects of interest, Proposition 3 states that the volume element $v_{\phi_o}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ may depend on $\mathbf{Q}$ and that in those cases a uniform joint prior distribution over the identified set for the vector of objects of interest will not induce a uniform prior over the set of orthogonal matrices.

4 Uniform Joint Prior for Impulse Responses

In this section, we show how to use the conventional method to conduct posterior inference based on a uniform joint prior distribution for the vector of impulse responses conditional on the sign restrictions. To do so, we analytically derive the prior distribution over the orthogonal reduced-form parameters induced by a uniform prior distribution for the IR parameters. This step is essential because the orthogonal reduced-form parameters are convenient for obtaining independent and identically distributed draws. Then, we derive a closed-form expression for the posterior over the orthogonal reduced-form parameters induced by a uniform prior distribution for the IR parameters. This posterior has an NIW and will allow us to use the conventional method to draw from it. We illustrate it using the empirical example in Watson (2020).

4.1 Prior for the Orthogonal Reduced-Form Parameters

Suppose $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is any density over the IR parameters. In that case, the induced density over the orthogonal reduced-form parameters will be $\pi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \pi(\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, where v_ϕ is the volume element induced by ϕ. The volume element can be calculated analytically using Proposition 4 below.

Proposition 4. *The volume element is* $v_\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$.

Proof. See Appendix A. □

The reader should notice that the volume element does not depend on h or $\mathbf{Q}$. Using Proposition 4, if $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is any density over the IR parameters, then the induced density over the orthogonal reduced-form parameters will be

$$\pi(\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_\phi((\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = \pi(\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}. \quad (9)$$

This last expression justifies the following proposition

Proposition 5. *The joint prior distribution for the vector of impulse responses is uniform if and only if the induced prior distributions over the orthogonal reduced-form parameters* $(\mathbf{B}, \mathbf{\Sigma})$ *and* $\mathbf{Q}$ *are independent, the distribution of* $\mathbf{Q}$ *is uniform with respect to the Haar measure, and the distribution over the reduced-form parameters has density proportional to* $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$.

Proof. The first two claims follow from Proposition 2 and the last from Equation (9). □

Proposition 5 shows that if one defines a uniform prior distribution for the IR parameters, then one is defining a prior for the orthogonal reduced-form parameters with three features: (1) it is independent between $(\mathbf{B}, \mathbf{\Sigma})$ and $\mathbf{Q}$, (2) the prior for the reduced-form parameters has a density that is proportional to $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$, and (3) the prior for $\mathbf{Q}$ is uniform with respect to the Haar measure. Because of this last feature, Proposition 2 shows that a uniform prior distribution for the IR parameters implies a uniform joint prior distribution over the identified set for the vector of impulse responses. Importantly, if the joint prior distribution for the vector of impulse responses is uniform, then the prior over the set defined by any one-to-one and onto linear transformation of the IR parameters will be uniform, and the marginal prior over any subset of the vector of impulse responses will also be uniform. At this stage, it is important to highlight that the induced prior for the reduced-form parameters is similar in spirit to (although also different than) the "weak" prior described in Uhlig (2005).

4.2 Posterior over the Orthogonal Reduced-Form Parameters

The following proposition from DeJong (1992) shows that a prior for the reduced-form parameters proportional to $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$ implies an NIW posterior.

Proposition 6. *Let* $a > 2n + 2 + m - T$. *If the reduced-form prior density is proportional to* $|\det(\mathbf{\Sigma})|^{-\frac{a}{2}}$, *then the NIW posterior density over the reduced-form parameters is defined by* $NIW_{\left(\hat{\nu}(a), \hat{\mathbf{S}}, \hat{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}\right)}(\mathbf{B}, \mathbf{\Sigma})$, *where* $\hat{\nu}(a) = T + a - m - n - 1$, $\hat{\mathbf{S}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$, *and* $\hat{\mathbf{B}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\mathbf{Y}$.

With Proposition 6 in hand, we have the following corollary characterizing the posterior over the orthogonal reduced-form parameters induced by a uniform prior distribution for the IR parameters.

Corollary 1. *If the prior density over the orthogonal reduced-form parameters is proportional to* $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$*, the posterior density over the orthogonal reduced-form parameters is* $UNIW_{\left(\hat{\nu}(-(m-3)),\hat{\mathbf{S}},\hat{\mathbf{B}},(\mathbf{X}'\mathbf{X})^{-1}\right)}(\mathbf{B},\mathbf{\Sigma})$.

Corollary 1 implies that if one wants to conduct inference based on a uniform prior distribution for the IR parameters, then one must have a particular (model-dependent) posterior over the reduced-form parameters. Specifically, the marginal posterior of $\mathbf{\Sigma}$ is inverse-Wishart with parameters $\hat{\nu}(-(m-3))$ and $\hat{\mathbf{S}}$, and the posterior of $\mathbf{B}$, conditional on $\mathbf{\Sigma}$, is normal with mean $\hat{\mathbf{B}}$ and variance $\mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}$.

4.3 The Conventional Method

The preceding discussion justifies the use of the conventional method for independently drawing from the posterior distribution for the IR parameters conditional on the sign restrictions implied by the uniform prior distribution for the IR parameters. Specifically, one can combine Algorithm 1 with the posterior over the orthogonal reduced-form parameters, as detailed in Corollary 1. To independently draw from the conjugate posterior distribution over $(\mathbf{L}_0, \mathbf{L}_+)$, conditional on the sign restrictions dictated by the uniform prior distribution for the IR parameters, one may refer to Algorithm 1, where Step 1 involves independently drawing from $NIW\left(\hat{\nu}(-(m-3)), \hat{\mathbf{S}}, \hat{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}\right)$. We regard our approach as a complement to the work of Plagborg-Møller (2019). While his approach does not facilitate independent draws, it offers the advantage of not necessitating invertibility.

Should one always impose the uniform prior distribution for the IR parameters? The answer clearly is no. It implies a lack of persistence, and one might strongly believe a priori that macroeconomic time series are reasonably persistent as described in the Minnesota prior or its variants. In this case, Proposition 2 tells us that the uniform distribution over the orthogonal matrices implies a uniform conditional joint prior distribution for the vector of impulse responses. The uniform joint prior distribution for the vector of impulse responses could be appealing to researchers concerned with the robustness of their conclusions. It amounts to the "weak" NIW prior for the reduced-form parameters, and it will get easily overthrown by any persistence in the data.

4.4 Extension to Objects of Interest

We show how to use the conventional method to conduct posterior inference based on a uniform joint prior distribution for a general vector of objects of interest conditional on the sign restrictions. If $\pi(\boldsymbol{\Upsilon})$ is any density over the vector of objects of interest, the induced density over the orthogonal reduced-form parameters is $\pi(\phi_o(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))v_{\phi_o}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. This justifies the following proposition

Proposition 7. *A joint prior distribution for the vector of objects of interest is uniform if and only if the equivalent prior density over the orthogonal reduced-form parameters is proportional to* $v_{\phi_o}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$.

Proof. Since $\pi(\phi_o(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})) \propto 1$, we have $\pi(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) \propto v_{\phi_o}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. □

Proposition 7 is a generalization of Proposition 5 for general vectors of objects of interest, where it is important to highlight that it may not be possible to compute the volume element analytically. In general, it is the case that the volume element $v_{\phi_o}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ depends on $\mathbf{Q}$ and, hence, the induced prior over the set of orthogonal matrices may not be uniform. An immediate implication of Proposition 7 is that a uniform joint prior distribution for the vector of objects of interest implies uniform joint prior and posterior distributions over the identified set for the vector of objects of interest.

We now show how to use the conventional methods to independently draw from the posterior distribution for the objects of interest parameters conditional on the sign restrictions for inference based on a uniform prior distribution for the objects of interest parameters. The algorithm below is a simple adaptation of Algorithm 1 that incorporates an importance sampling step. In order to justify the weights in the importance sampling step, note that the likelihood is proportional to $NIW_{(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}})}(\mathbf{B},\mathbf{\Sigma})$, where $\hat{\nu}=T-m-n-1$, $\hat{\mathbf{\Omega}}=(\mathbf{X}'\mathbf{X})^{-1}$, $\hat{\mathbf{\Psi}}=\hat{\mathbf{\Omega}}\mathbf{X}'\mathbf{Y}$, and $\hat{\mathbf{\Phi}}=\mathbf{Y}'\mathbf{Y}-\hat{\mathbf{\Psi}}'\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{\Psi}}$. If the prior of the objects of interest is uniform, then the posterior density will also be proportional to $NIW_{(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}})}(\mathbf{B},\mathbf{\Sigma})$.

Algorithm 2. *The following algorithm independently draws from the posterior distribution for the objects of interest parameters conditional on the sign restrictions implied by a uniform prior distribution for the objects of interest parameters.*

1. *Draw* $(\mathbf{B},\mathbf{\Sigma})$ *independently from the* $NIW\,(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})$ *distribution.*
2. *Draw* $\mathbf{Q}$ *independently from the uniform distribution over* $\mathcal{O}(n)$*.*
3. *If* $\mathbf{\Upsilon}=\phi_o(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ *satisfies the sign restrictions, then set its importance weight to*

$$\frac{NIW_{(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}})}(\mathbf{B},\mathbf{\Sigma})v_{\phi_o}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})}{NIW_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{B},\mathbf{\Sigma})}.$$

   Otherwise, set its importance weight to zero.
4. *Return to Step 1 until the required number of draws has been obtained.*

The choice of $(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})$ matters. An obvious choice would be $(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})=\left(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}}\right)$, which would simplify the importance weight. More generally, one could choose $(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})$ to maximize the effective sample size of the importance sampler. It is also important to highlight that one could use Algorithm 2 to work with any joint posterior distribution for the vector of objects of interest provided that Step 3 is modified accordingly.

5 An Application

We use the empirical application in Watson (2020) to illustrate how to conduct inference based on a uniform joint prior distribution for the vector of impulse responses. We will contrast the results with those obtained using the Minnesota prior for the reduced-form parameters.

5.1 Data, Model, Identification Restrictions, and Prior

Watson's (2020) SVAR analysis relies on quarterly data for the U.S. economy over the period 1984Q1:2007Q4. The variables included in the model are $\mathbf{y}_t' = (\Delta(y_t - n_t), n_t, \Delta p_t, i_t^L)$, where y_t denotes the logarithm of real output per hour for all workers in the nonfarm business sector (Bureau of Labor Statistics, U.S. (2019b)), n_t the logarithm of hours worked per capita (Bureau of Labor Statistics, U.S. (2019a,c)), p_t the logarithm of the price level (Bureau of Economic Analysis, U.S. (2019)), and i_t^L the 10-year Treasury bond rate (Board of Governors of the Federal Reserve System (2019)). The SVAR is a constant parameter variant of Debortoli, Galí, and Gambetti (2020) featuring 4 lags and an intercept. It is assumed that fluctuations in $\mathbf{y}_t'$ are driven by technology, demand, supply, and monetary policy shocks, which are identified with sign and zero restrictions.

The identifying restrictions are as follows. The technology shock is the only structural shock that can affect labor productivity in the long run. Four quarters after a demand shock, the responses of output, the price level, and the 10-year Treasury bond rate are negative. Four quarters after a monetary policy shock, the responses of output and the price level are negative, while the impulse response of the 10-year Treasury bond rate is positive. Four quarters after a supply shock, the response of output is negative, while the response of inflation is positive. We also impose stability of the VAR. The zero restrictions on the long-run impulse responses have a particular structure that can be exploited to use Algorithm 1.[9]

The Minnesota prior is as follows. We set $\bar{\nu} = n + 2$, which is the minimum value $\bar{\nu}$ can

[9]Given the reduced-form parameters, uniformly drawing a four-dimensional orthogonal matrix conditional on the zero restrictions is equivalent to uniformly drawing a three-dimensional orthogonal matrix using Proposition 1 and then mapping it to a four-dimensional orthogonal using a Householder matrix that depends only on the reduced-form parameters. The space of $\mathbf{u}_t$'s that do not have permanent effects on labor productivity is three-dimensional. See Appendix C.

take that guarantees the existence of a prior mean for $\mathbf{\Sigma}$. The matrix $\bar{\mathbf{\Phi}}$ is diagonal, with $\bar{\mathbf{\Phi}} = \text{diag}(\phi_1, \phi_2, \phi_3, \phi_4)$. The values for $\bar{\mathbf{\Psi}}$ and $\bar{\mathbf{\Omega}}$ are chosen to have a flat density over the constant term ($\text{Var}(\mathbf{d} \mid \mathbf{\Sigma}) = 10^7\mathbf{\Sigma}$) and the following first and second moments over the slope parameters

$$\mathbb{E}\left((\mathbf{B}_\ell)_{ij} \mid \mathbf{\Sigma}\right) = \begin{cases} 1 & \text{if } i = j = 2 \text{ and } \ell = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Cov}\left((\mathbf{B}_\ell)_{ij}, (\mathbf{B}_r)_{hm} \mid \mathbf{\Sigma}\right) = \begin{cases} \lambda^2 \frac{1}{\ell^2} \frac{\mathbf{\Sigma}_{jm}(\bar{\nu}-n-1)}{\phi_i} & \text{if } i = h \text{ and } \ell = r \\ & \text{for all } i, j, h, m, \ell, r = 1, \ldots 4 \\ 0 & \text{otherwise.} \end{cases}$$

We will treat λ and $\bar{\mathbf{\Phi}}$ as hyperparameters. We follow Giannone, Lenza, and Primiceri (2015) in choosing the values for these parameters that maximize the marginal likelihood. This yields $\lambda = 0.3953$, and $\bar{\mathbf{\Phi}} = \text{diag}(2.0419, 0.5241, 0.0586, 0.2103)$.

For completeness, we will begin the analysis comparing the posterior distributions of individual impulse responses implied by the uniform prior distribution for the IR parameters with the posterior distributions of individual impulse responses implied by the prior distribution for the IR parameters induced by the described Minnesota prior. Figure 3 shows the equal-tailed 68 percent point-wise posterior probability bands of individual impulse responses implied by each of the priors. The figure shows how the uniform joint prior distribution for the vector of impulse responses implies more posterior uncertainty.[10]

Next, we compare marginal and joint inference when using the uniform prior distribution for the IR parameters. Figure 4 compares the Bayes estimator of the joint posterior distribution for the vector of impulse responses (dashed lines) and its 68 percent credible set (solid light gray lines) under the additively separable absolute loss function following Inoue and Kilian

[10]The results of this comparison are based on 10,000 draws from the posterior distribution conditional on the identifying restrictions.

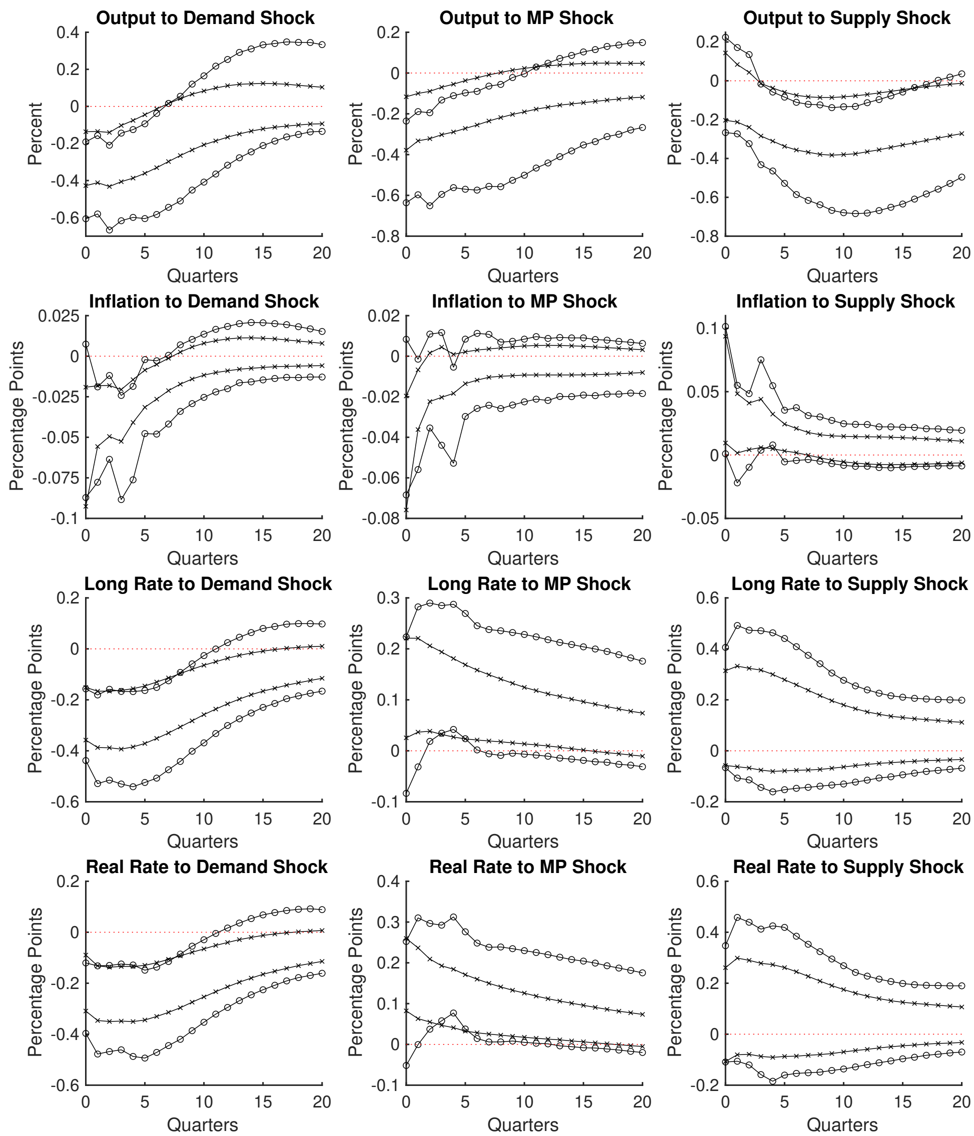

Figure 3: The solid lines with circle (∘) markers depict the equal-tailed 68 percent marginal posterior probability bands of individual impulse responses implied by the uniform joint prior distribution for the vector of impulse responses. The solid lines with cross (×) markers depict the equal-tailed 68 percent marginal posterior probability bands of individual impulse responses implied by the Minnesota prior.

(2022a) with the commonly used equal-tailed 68 percent point-wise posterior probability bands (solid lines with circle (○) markers).[11,12] In contrast to point-wise probability bands, the joint credible set for the Bayes estimator restricts all of its members to satisfy the dependence structure of the impulse responses. As a consequence, as shown in the figure, the joint credible sets are wider than the conventional point-wise probability bands. While most of the 68 point-wise probability bands for individual impulse responses do not contain zero, the 68 percent joint credible set contains zero at all except the restricted horizons. Hence, when conducting joint inference, it becomes clear that this particular model does not seem to be tightly identified by the restrictions. These conclusions are robust to using the sup-t Bayesian joint credible sets proposed by Montiel Olea and Plagborg-Møller (2019).

We conclude this section by comparing the joint posterior distribution for the vector of impulse responses implied by the uniform prior and the Minnesota prior. Figure 5 shows the Bayes estimator of the joint posterior distribution for the vector of impulse responses and its 68 percent credible set under the additively separable absolute loss function when using a uniform joint prior distribution for the vector of impulse responses (dashed lines for the estimator and solid light gray lines for the credible set) and when using the joint prior distribution for the vector of impulse responses induced by the Minnesota prior (dashed-dotted lines for the estimator and solid dark gray lines for the credible set). Focusing on the Bayes estimators, the Minnesota prior and the uniform prior for impulse responses imply broadly similar estimates with some exceptions such as the impact response of output to a supply shock. The 68 percent credible sets are much narrower when using the Minnesota prior. Still, a visual inspection reveals that in both cases, there is substantial joint uncertainty about the macroeconomic consequences of the shocks under study. A similar picture emerges when using the sup-t Bayesian joint credible sets. As mentioned above, this is clearly in line with the conclusions in Inoue and Kilian (2022a).

[11]As remarked by Inoue and Kilian (2022a), other loss functions, such as a quadratic loss, could be used.

[12]In the case of joint inference, the Bayes estimator and the credible set are based on 10,000 draws. Increasing the number of draws results in a more accurate depiction of the joint posterior distribution.

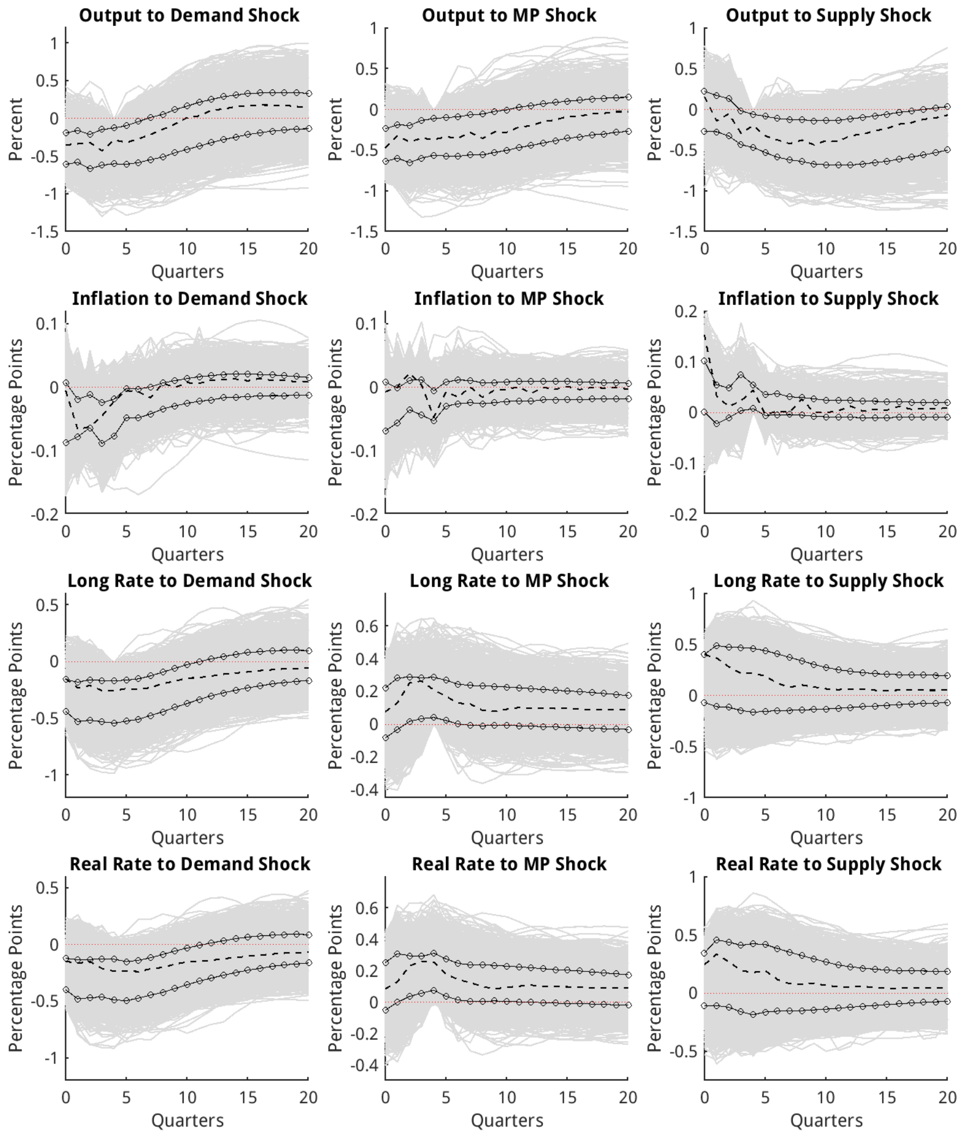

Figure 4: Bayes estimator of the joint posterior distribution for the vector of impulse responses (dashed lines) and its 68 percent credible set (solid light gray lines) under the additively separable absolute loss function. The solid lines with circle (∘) markers depict the equal-tailed 68 percent unconditional prior distributions for individual impulse responses. Both posteriors are implied by the uniform joint prior distribution for the vector of impulse responses.

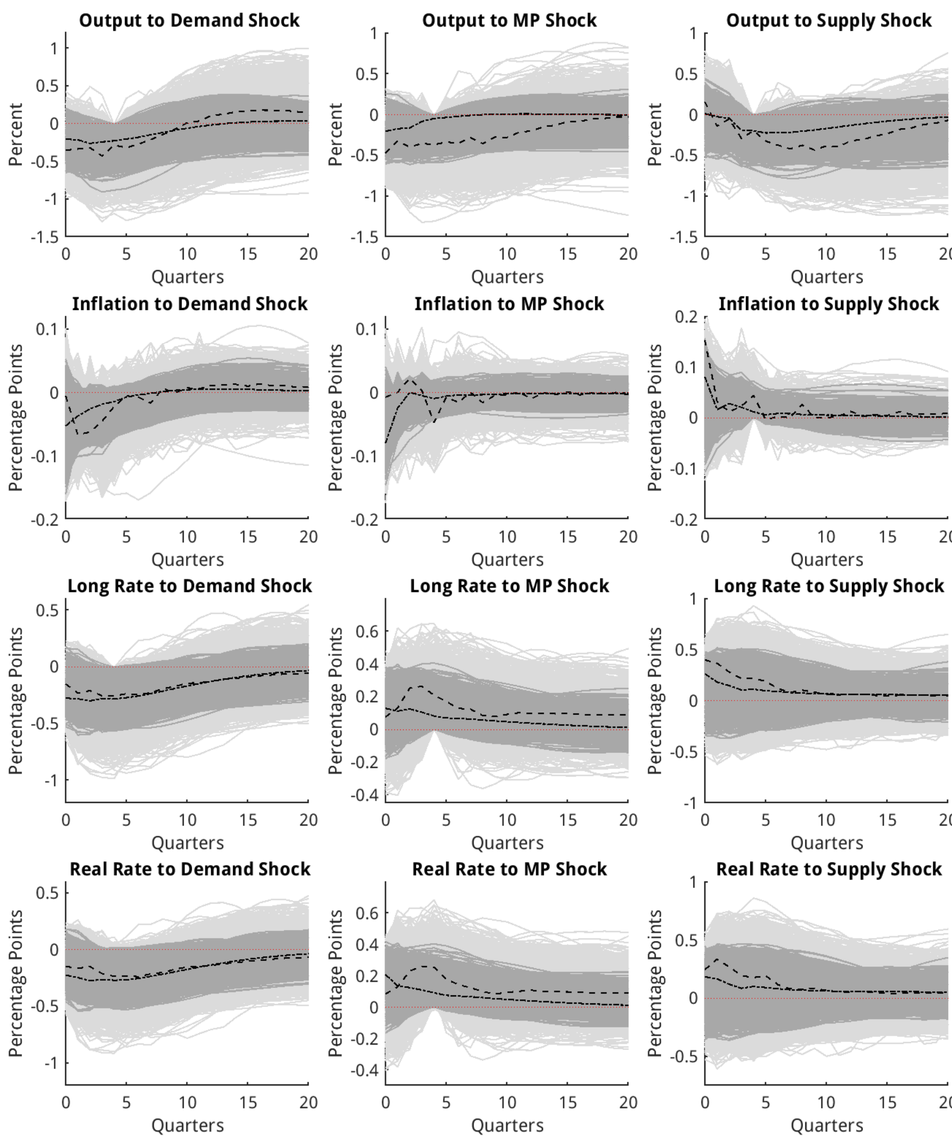

Figure 5: Bayes estimator of joint posterior impulse responses (dashed black lines) and its 68 percent credible set under the additively separable absolute loss function under uniform joint prior distribution for the vector of impulse responses (solid light gray lines) and under the Minnesota prior (dashed-dotted lines and solid dark gray lines for the 68 percent credible set).

6 Conclusion

Our paper demonstrates that there is nothing fundamentally wrong with the conventional method for Bayesian inference in SVARs identified with sign restrictions. We show that the uniform prior over the set of orthogonal matrices is not only sufficient but also necessary to have uniform joint prior and posterior distributions over the identified set for the vector of impulse responses. The key is to consider joint distributions instead of marginal distributions. The most popular choice of prior when using the conventional method induces a uniform joint prior distribution over the identified set for the vector of impulse responses, and straightforward variants of the approach can be used to conduct joint inference using either a uniform joint prior distribution for the vector of impulse responses or a joint prior distribution for the vector of objects of interest within a general class of objects of interest.

Our paper can also be viewed as complementing Giacomini and Kitagawa (2021) for researchers whose goal is to perform joint posterior inference without favoring some vector of impulse responses over others a priori. This is because even though their prior robust numerical methodology is attractive, it does not consider the case of joint inference, and such an extension is challenging.

This paper has focused on SVARs identified with sign restrictions. Nevertheless, the conventional method can also be used to independently draw from the posterior distribution for the IR parameters implied by a uniform prior distribution over such parameterization in SVARs identified with sign and zero restrictions. The same applies when the objective is to draw from the posterior distribution for the objects of interest parameters implied by a uniform prior distribution over such parameterization conditional on sign and zero restrictions. As described in Arias, Rubio-Ramírez, and Waggoner (2018), in both cases, an importance sampling step could be needed depending on the nature of the parameterization of interest and the zero restrictions in use.

A Proofs of Propositions 2, 3, and 4

Proof of Proposition 2. If π is any density of the impulse responses with respect to the Lebesgue measure, then the induced density over orthogonal reduced-form parameters with respect to volume measure is

$$p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \frac{\pi(\phi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}{2^{\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{-(m-3)}{2}}}.$$

So, the density π is constant over the set $\mathcal{P}(\mathbf{B}, \mathbf{\Sigma})$ iff $p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ does not depend on $\mathbf{Q}$. Since $p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = p(\mathbf{B}, \mathbf{\Sigma})p(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$, $p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ does not depend on $\mathbf{Q}$ iff $p(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$ is constant. If $p(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$ is constant, then the induced distributions of $(\mathbf{B}, \mathbf{\Sigma})$ and $\mathbf{Q}$ are independent, and the distribution of $\mathbf{Q}$ must be uniform with respect to the Haar measure. □

Proof of Proposition 3. Suppose $\pi(\mathbf{\Upsilon})$ is any density over the objects of interest parameterization with respect to the Lebesgue measure, then the induced density over the orthogonal reduced-form parameters with respect to the volume measure will be $\pi(\mathbf{B}, \mathbf{\Sigma})\pi(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma}) = \pi(\phi_o(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{\phi_o}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$. If $\pi(\mathbf{\Upsilon})$ is constant over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$, then $\pi(\phi_o(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))$ will not depend on $\mathbf{Q}$ and $\pi(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$ is proportional to $v_{\phi_o}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, though the proportionality constant, which is equal to $\pi(\phi_o(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))/\pi(\mathbf{B}, \mathbf{\Sigma})$, could depend on $\mathbf{B}$ and $\mathbf{\Sigma}$. If $\pi(\mathbf{Q} \mid \mathbf{B}, \mathbf{\Sigma})$ is proportional to $v_{\phi_o}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, then $\pi(\phi_o(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))$ cannot depend on $\mathbf{Q}$ and so is constant over $\mathcal{P}_o(\mathbf{B}, \mathbf{\Sigma})$. □

Proof of Proposition 4. Let $\mathbf{A}_0 = \left(\mathbf{L}_0^{-1}\right)'$ and $\mathbf{A}_+ = \mathbf{B}\,\mathbf{A}_0$. Multiplying Equation (1) on the right by $\mathbf{A}_0$ gives $\mathbf{y}_t'\,\mathbf{A}_0 = \mathbf{x}_t'\,\mathbf{A}_+ + \boldsymbol{\varepsilon}_t'$ for $1 \leq t \leq T$, which is often called the structural form and $(\mathbf{A}_0, \mathbf{A}_+)$ the structural parameters. For $1 \leq \ell \leq p$, let $\mathbf{A}_\ell = \mathbf{B}_\ell\, A_0$. Multiplying Equation (3) on the right by $\mathbf{A}_0$ gives $\mathbf{A}_\ell = \mathbf{A}_0\,\mathbf{L}_\ell'\,\mathbf{A}_0 - \sum_{k=1}^{\ell-1}\left(\mathbf{L}_{\ell-k}\mathbf{L}_0^{-1}\right)'\mathbf{A}_k$. Since $\mathbf{A}_+ = \begin{bmatrix}\mathbf{A}_1' & \cdots & \mathbf{A}_p' & \mathbf{c}'\end{bmatrix}'$, this recursively defines a mapping from the IR parameters to the structural parameters, which we denote by f. It follows from Proposition 1 of Arias, Rubio-Ramírez, and Waggoner (2018) that the volume element of $f \circ \phi$ is $v_{f\circ\phi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-\frac{2n+m+1}{2}}$, which implies that the volume element of ϕ is $v_\phi(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}) = \frac{2^{-\frac{n(n+1)}{2}}|\det(\mathbf{\Sigma})|^{-\frac{2n+m+1}{2}}}{v_f(\phi(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}$. Because $\mathbf{A}_k$ does not depend on $\mathbf{L}_j$ for $j > k$, the derivative of f is a block lower triangular $(n^2(p+1)+n) \times (n^2(p+1)+n)$ matrix

$$\begin{bmatrix} -\mathbf{K}_{n,n}(\mathbf{L}_0' \otimes \mathbf{L}_0)^{-1} & 0 & \cdots & 0 & 0 \\ \times & (\mathbf{L}_0 \otimes \mathbf{L}_0')^{-1}\mathbf{K}_{n,n} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \times & \times & \cdots & (\mathbf{L}_0 \otimes \mathbf{L}_0')^{-1}\mathbf{K}_{n,n} & 0 \\ 0 & 0 & \cdots & 0 & \boldsymbol{I}_n \end{bmatrix}$$

where $\mathbf{K}_{n,n}$ is the commutation matrix, which is the unique $n^2 \times n^2$ matrix such that $\text{vec}(\mathbf{X}') = \mathbf{K}_{n,n}\,\text{vec}(\mathbf{X})$ for every $n \times n$ matrix $\mathbf{X}$. The volume element of f is the absolute value of the determinant of the above matrix, which is $|\det(\mathbf{L}_0)|^{-2n(p+1)}$. Since $\det(\mathbf{L}_0) = \det(\mathbf{\Sigma})^{\frac{1}{2}}$, the volume element of ϕ is $v_\phi(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}) = 2^{-\frac{n(n+1)}{2}}|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$. □

B Proofs of Claims from Section 3.1

B.1 Derivation of Equation (7)

The function that maps $(\theta, i) \in [-\pi, \pi) \times \{0, 1\}$ to $\ell_{11} = \hat{\ell}_{11}\cos(\theta) \in [-\hat{\ell}_{11}, \hat{\ell}_{11}]$ is not one-to-one over its entire domain, but is one-to-one over each of the four subdomains of the form $S_{+,i} = [0, \pi) \times \{i\}$ or $S_{-,i} = [-\pi, 0) \times \{i\}$. Let $\tilde{\ell}_{11} = \ell_{11}/\hat{\ell}_{11}$. We follow the convention that $\cos^{-1}(\cdot) \in [0, \pi]$. Over $S_{+,i}$, the inverse of the above mapping is $(\theta, i) = (\cos^{-1}(\tilde{\ell}_{11}), i) \in S_{+,i}$, and over $S_{-,i}$, the inverse of the above mapping is $(\theta, i) = (-\cos^{-1}(\tilde{\ell}_{11}), i) \in S_{-,i}$. Since the derivative of $\cos(\theta)$ is $-\sin(\theta)$, by the usual change of variable theorem, the density over $\ell_{11} \in [-\hat{\ell}_{11}, \hat{\ell}_{11}]$ induced by the density $p(\theta)p(i|\theta)$ over $[-\pi, \pi) \times \{0, 1\}$ is

$$p(\ell_{11}) = \frac{p(\tilde{\theta})p(0|\tilde{\theta})}{|\hat{\ell}_{11}\sin(\tilde{\theta})|} + \frac{p(\tilde{\theta})p(1|\tilde{\theta})}{|\hat{\ell}_{11}\sin(\tilde{\theta})|} + \frac{p(\hat{\theta})p(0|\hat{\theta})}{|\hat{\ell}_{11}\sin(\hat{\theta})|} + \frac{p(\hat{\theta})p(1|\hat{\theta})}{|\hat{\ell}_{11}\sin(\hat{\theta})|} = \frac{p(\tilde{\theta})}{|\hat{\ell}_{11}\sin(\tilde{\theta})|} + \frac{p(\hat{\theta})}{|\hat{\ell}_{11}\sin(\hat{\theta})|},$$

where $\tilde{\theta} = \cos^{-1}(\tilde{\ell}_{11})$ and $\hat{\theta} = -\tilde{\theta}$. Since $\sin(\hat{\theta}) = -\sin(\tilde{\theta})$, $\sin(\tilde{\theta}) \geq 0$, and $\hat{\ell}_{11} > 0$,

$$p(\ell_{11}) = \frac{p(\tilde{\theta}) + p(\hat{\theta})}{\hat{\ell}_{11}\sin(\tilde{\theta})} = \frac{p(\tilde{\theta}) + p(\hat{\theta})}{(\hat{\ell}_{11}^2 - \ell_{11}^2)^{\frac{1}{2}}}, \tag{10}$$

where the last equality follows from the fact that $\sin(\tilde{\theta}) = (1 - \cos^2(\tilde{\theta}))^{\frac{1}{2}}$ and will be of use in Appendix B.4. The first equality is Equation (7).

B.2 Derivation of Equation (8)

The function that maps $(\theta, i) \in [-\pi, \pi) \times \{0, 1\}$ to $\ell_{12} = \hat{\ell}_{11}\sin(\theta) \in [-\hat{\ell}_{11}, \hat{\ell}_{11}]$ is not one-to-one over its entire domain, but is one-to-one over each of the four subdomains of the form $S_{c,i} = [-\pi/2, \pi/2) \times \{i\}$ or $S_{d,i} = ([-\pi, -\pi/2) \cup [\pi/2, \pi)) \times \{i\}$. Let $\tilde{\ell}_{12} = \ell_{12}/\hat{\ell}_{11}$. We follow the convention that $\sin^{-1}(\cdot) \in [-\pi/2, \pi/2]$. Over $S_{c,i}$, the inverse of the above mapping is $(\theta, i) = (\sin^{-1}(\tilde{\ell}_{12}), i) \in S_{c,i}$, and over $S_{d,i}$, the inverse of the above mapping is $(\theta, i) = (\mathrm{sgn}(\tilde{\ell}_{12})\pi - \sin^{-1}(\tilde{\ell}_{12}), i) \in S_{d,i}$. Since the derivative of $\sin(\theta)$ is $\cos(\theta)$, by the usual change of variable theorem, the density over $\ell_{12} \in [-\hat{\ell}_{11}, \hat{\ell}_{11}]$ induced by the density $p(\theta)p(i|\theta)$ over $[-\pi, \pi) \times \{0, 1\}$ is

$$p(\ell_{12}) = \frac{p(\tilde{\theta})p(0|\tilde{\theta})}{|\hat{\ell}_{11}\cos(\tilde{\theta})|} + \frac{p(\tilde{\theta})p(1|\tilde{\theta})}{|\hat{\ell}_{11}\cos(\tilde{\theta})|} + \frac{p(\hat{\theta})p(0|\hat{\theta})}{|\hat{\ell}_{11}\cos(\hat{\theta})|} + \frac{p(\hat{\theta})p(1|\hat{\theta})}{|\hat{\ell}_{11}\cos(\hat{\theta})|} = \frac{p(\tilde{\theta})}{|\hat{\ell}_{11}\cos(\tilde{\theta})|} + \frac{p(\hat{\theta})}{|\hat{\ell}_{11}\cos(\hat{\theta})|},$$

where $\tilde{\theta} = \sin^{-1}(\tilde{\ell}_{12})$ and $\hat{\theta} = \mathrm{sgn}(\tilde{\ell}_{12})\pi - \tilde{\theta}$. Since $\cos(\tilde{\theta}) = -\cos(\hat{\theta})$, $\cos(\tilde{\theta}) \geq 0$, and $\hat{\ell}_{11} \geq 0$, we have

$$p(\ell_{12}) = \frac{p(\tilde{\theta}) + p(\hat{\theta})}{\hat{\ell}_{11}\cos(\tilde{\theta})} = \frac{p(\tilde{\theta}) + p(\hat{\theta})}{(\hat{\ell}_{11}^2 - \ell_{12}^2)^{\frac{1}{2}}}, \tag{11}$$

where the last equality follows from the fact that $\cos(\tilde{\theta}) = (1 - \sin^2(\tilde{\theta}))^{\frac{1}{2}}$ and will be of use in Appendix B.4. The first equality is Equation (8).

B.3 Proof that the Distributions over ℓ_{11} and ℓ_{12} Cannot Both Be Uniform

If the conditional distribution of ℓ_{11} is uniform, then $p(\ell_{11}) = 1/(2\hat{\ell}_{11})$ and the distribution of θ must satisfy

$$p(\theta) + p(-\theta) = \sin(\theta)/2, \text{ for } 0 \leq \theta < \pi. \tag{12}$$

If the conditional distribution of ℓ_{12} is uniform, then $p(\ell_{12}) = 1/(2\hat{\ell}_{11})$ and, because $\text{sgn}(\ell_{12}/\hat{\ell}_{11}) = \text{sgn}(\sin^{-1}(\ell_{12}/\hat{\ell}_{11}))$, the distribution of θ must satisfy

$$p(\theta) + p(\text{sgn}(\theta)\pi - \theta) = \cos(\theta)/2 \text{ for } -\pi/2 \leq \theta \leq \pi/2. \tag{13}$$

So, for $0 \leq \theta \leq \pi/2$, it must be the case that $\cos(\theta)/2 = p(\theta) + p(\pi - \theta)$, thus

$$\cos(\theta)/2 = \sin(\theta)/2 - p(-\theta) + \sin(\pi - \theta)/2 - p(-\pi + \theta) = \sin(\theta) - \cos(\theta)/2.$$

The first equality follows by substitution using Equation (13). The second equality follows by two substitutions using Equation (12). The last equality follows by substitution using Equation (13) and from the fact that $\sin(\theta) = \sin(\pi - \theta)$. This would imply that $\cos(\theta) = \sin(\theta)$, which is not true.

B.4 The Density of ℓ_{11} and ℓ_{12} in Cases (1) and (2)

In Case (1), it had to be the case that $p(\theta) = 1/(2\pi)$ and $p(i|\theta) = 1/2$. Equation (10) gives $p(\ell_{11}) = 1/(\pi(\hat{\ell}_{11}^2 - \ell_{11}^2)^{\frac{1}{2}})$. Equation (11) gives $p(\ell_{12}) = 1/(\pi(\hat{\ell}_{11}^2 - \ell_{12}^2)^{\frac{1}{2}})$.

In Case (2), we chose $p(\theta) = |\sin(\theta)/4|$ and $p(i|\theta) = 1/2$. Equation (10) gives

$$p(\ell_{11}) = \frac{|\sin(\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))| + |\sin(-\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))|}{4\hat{\ell}_{11}\sin(\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))} = \frac{1}{2\hat{\ell}_{11}},$$

because $\sin(-\cos^{-1}(\ell_{11}/\hat{\ell}_{11})) = -\sin(\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))$. Equation (11) gives

$$p(\ell_{12}) = \frac{|\sin(\sin^{-1}(\ell_{12}/\hat{\ell}_{11}))| + |\sin(\operatorname{sgn}(\ell_{12}/\hat{\ell}_{11})\pi - \sin^{-1}(\ell_{12}/\hat{\ell}_{11}))|}{4(\hat{\ell}_{11}^2 - \ell_{12}^2)^{\frac{1}{2}}} = \frac{|\ell_{12}|}{2\hat{\ell}_{11}(\hat{\ell}_{11}^2 - \ell_{12}^2)^{\frac{1}{2}}}$$

because $\sin(\operatorname{sgn}(\ell_{12}/\hat{\ell}_{11})\pi - \sin^{-1}(\ell_{12}/\hat{\ell}_{11})) = \sin(\sin^{-1}(\ell_{12}/\hat{\ell}_{11})) = \ell_{12}/\hat{\ell}_{11}$.

C Posterior Simulation of Watson (2020)

The model in Watson (2020) has three zero restrictions on the long-run impulse response of labor productivity growth. The long-run impulse response is given by

$$\mathbf{L}_\infty = \left(\mathbf{A}_0' - \sum_{i=1}^{p} \mathbf{A}_i'\right)^{-1} = \left(\mathbf{I}_n - \sum_{i=1}^{p} \mathbf{B}_i'\right)^{-1} (\mathbf{A}_0^{-1})' = \left(\mathbf{I}_n - \sum_{i=1}^{p} \mathbf{B}_i'\right)^{-1} h(\mathbf{\Sigma})'\,\mathbf{Q},$$

where $\mathbf{B}_i = \mathbf{A}_i\,\mathbf{A}_0^{-1}$. If labor productivity is the first variable and the technology shock is ordered last, then the first three elements in the first row of $\mathbf{L}_\infty$ must be zero. Given a non-zero n-vector $\mathbf{x}$, the Householder matrix $\mathbf{H}_n(\mathbf{x})$ is given by $\mathbf{H}_n(\mathbf{x}) = \mathbf{I}_n - 2\,\frac{\mathbf{x}\mathbf{x}'}{\mathbf{x}'\mathbf{x}}$. Householder matrices are reflection matrices, and hence orthogonal. If $\mathbf{x}$ and $\mathbf{y}$ are two distinct unit vectors, then $\mathbf{x}'\mathbf{H}_n(\mathbf{x}-\mathbf{y}) = \mathbf{y}$. Let $\mathbf{x}(\mathbf{B},\mathbf{\Sigma})'$ be the first row of $\left(\mathbf{I}_n - \sum_{i=1}^{p}\mathbf{B}_i'\right)^{-1} h(\mathbf{\Sigma})'$, normalized to be of unit length, and let $\mathbf{e}_4$ be the last column of $\mathbf{I}_4$. It is easy to see that $\left(\mathbf{I}_n - \sum_{i=1}^{p}\mathbf{B}_i'\right)^{-1} h(\mathbf{\Sigma})'\mathbf{H}_n(\mathbf{x}(\mathbf{B},\mathbf{\Sigma}) - \mathbf{e}_4)$ will satisfy the zero restrictions, as long as $\mathbf{x}(\mathbf{B},\mathbf{\Sigma}) \neq \mathbf{e}_4$. Furthermore, if $\mathbf{L}_\infty = \left(\mathbf{I}_n - \sum_{i=1}^{p}\mathbf{B}_i'\right)^{-1} h(\mathbf{\Sigma})'\,\mathbf{Q}$ satisfies the zero restrictions, then $\mathbf{Q}$ must be of the form $\mathbf{H}_n(\mathbf{x}(\mathbf{B},\mathbf{\Sigma}) - \mathbf{e}_4)\mathbf{P}$, where

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & \pm 1 \end{bmatrix} \tag{14}$$

and $\mathbf{P}_3 \in \mathcal{O}(3)$. Thus, given the reduced-form parameters $(\mathbf{B},\mathbf{\Sigma})$, a $\mathbf{Q}$ can be obtained by (1) drawing $\mathbf{P}_3$ using Proposition 1, (2) drawing ± 1 uniformly, (3) forming $\mathbf{P}$, and (4) finally multiplying by the Householder matrix $\mathbf{H}_n(\mathbf{x}(\mathbf{B}, \mathbf{\Sigma}) - \mathbf{e}_4)$, we obtain a uniform draw from $\mathcal{O}(4)$ conditional on the zero restrictions. In addition, it can be shown that the mapping from $\mathbf{P}_3$ and ± 1 to the IR parameters conditional on the zero restrictions does not depend on $\mathbf{P}_3$ or ± 1. This implies that the ratio of volume elements associated with the target and the proposals does not depend on $\mathbf{Q}$. Thus, Algorithm 1 can be used in this case, provided that a simple re-weighing step is implemented. Notice that Proposition 2 directly applies to the IR parameters identified with sign restrictions. It can be shown that they also apply to the model in Watson (2020) with other IR parameters defined as $(\mathbf{L}_0, \mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \mathbf{L}_\infty, \mathbf{c})$. The mapping from these IR parameters to the orthogonal reduced-form parameters is one-to-one and onto, although we do have to restrict the parameters so that $\mathbf{L}_\infty$ is well defined. Using these IR parameters, the zero restrictions define a lower dimensional linear subspace where the volume measure is Lebesgue.

References

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, 685–720.

Baumeister, C. and J. D. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

Board of Governors of the Federal Reserve System, U.S. (2019): "Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity and Investment Basis [GS10]," Vintage 2019-04-01.

Bruder, S. and M. Wolf (2018): "Balanced Bootstrap Joint Confidence Bands for Structural Impulse Response Functions," *Journal of Time Series Analysis*, 39, 641–664.

Bureau of Economic Analysis, U.S. (2019): "Gross Domestic Product: Implicit Price Deflator [GDPDEF]," Vintage 2019-03-28.

Bureau of Labor Statistics, U.S. (2019a): "Nonfarm Business Sector: Hours Worked for All Workers [HOANBS]," Vintage 2019-03-07.

——— (2019b): "Nonfarm Business Sector: Labor Productivity (Output per Hour) for All Workers [OPHNFB]," Vintage 2019-03-07.

——— (2019c): "Population Level [CNP16OV]," Vintage 2019-04-05.

Canova, F. and G. De Nicoló (2002): "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Debortoli, D., J. Galí, and L. Gambetti (2020): "On the Empirical (Ir)Relevance of the Zero Lower Bound Constraint," *NBER Macroeconomics Annual*, 34, 141–170.

DeJong, D. N. (1992): "Co-Integration and Trend-Stationarity in Macroeconomic Time Series: Evidence from the Likelihood Function," *Journal of Econometrics*, 52, 347–370.

Faust, J. (1998): "The Robustness of Identified VAR Conclusions About Money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

Fry, R. and A. Pagan (2011): "Sign Restrictions in Structural Vector Autoregressions: A Critical Review," *Journal of Economic Literature*, 49, 938–960.

Giacomini, R. and T. Kitagawa (2021): "Robust Bayesian Inference for Set-identified Models," *Econometrica*, 89, 1519–1556.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97, 436–451.

Halmos, P. R. (1950): *Measure Theory*, Springer.

Inoue, A. and L. Kilian (2013): "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 177, 1–13.

——— (2016): "Joint Confidence Sets for Structural Impulse Responses," *Journal of Econometrics*, 192, 421–432.

——— (2019): "Corrigendum to "Inference on Impulse Response Functions in Structural VAR Models" [Journal of Econometrics 177 (2013) 1–13]," *Journal of Econometrics*, 209, 139–143.

——— (2022a): "Joint Bayesian Inference About Impulse Responses in VAR Models," *Journal of Econometrics*, 231, 457–476.

——— (2022b): "The Role of the Prior in Estimating VAR Models with Sign Restrictions," *FRB of Dallas Working Paper No. 2030*.

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015a): "Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions," *International Journal of Forecasting*, 31, 782–798.

——— (2015b): "Confidence Bands for Impulse Responses: Bonferroni vs. Wald," *Oxford Bulletin of Economics and Statistics*, 77, 800–821.

——— (2018): "Calculating Joint Confidence Bands for Impulse Response Functions Using Highest Density Regions," *Empirical Economics*, 55, 1389–1411.

Montiel Olea, J. L. and M. Plagborg-Møller (2019): "Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs," *Journal of Applied Econometrics*, 34, 1–17.

Munkres, J. R. (1991): *Analysis on Manifolds*, Westview Press, Advanced Book Classics.

Plagborg-Møller, M. (2019): "Bayesian Inference on Structural Impulse Response Functions," *Quantitative Economics*, 10, 145–184.

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, 665–696.

Sims, C. A. and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67, 1113–1155.

Spivak, M. (1965): *Calculus on Manifolds*, Benjamin/Cummings.

Stewart, G. (1980): "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators," *SIAM Journal on Numerical Analysis*, 17, 403–409.

Uhlig, H. (1994): "What Macroeconomists Should Know About Unit Roots: A Bayesian Perspective," *Econometric Theory*, 10, 645–671.

——— (1998): "The Robustness of Identified VAR conclusions about Money: A Comment," *Carnegie-Rochester Conference Series on Public Policy*, 49, 245–263.

——— (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Watson, M. W. (2020): "Comment," *NBER Macroeconomics Annual*, 34, 182–193.

Wolf, C. K. (2020): "SVAR (Mis)identification and the Real Effects of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*, 12, 1–32.