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

Federal Reserve Bank of Atlanta

Emory University

February 22, 2023

Abstract

There has been a call for caution when using the conventional method for Bayesian inference in set-identified structural vector autoregressions on the grounds that the uniform prior over the set of orthogonal matrices could be nonuniform for key objects of interest. This paper challenges this call. Although the prior distributions of individual impulse responses induced by the conventional method may be nonuniform, they typically do not drive the posteriors if one does not condition on the reduced-form parameters. Importantly, 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. We also propose variants of the conventional method to conduct inference based on a uniform joint prior distribution for the vector of impulse responses. We generalize our results to vectors of objects of interest beyond impulse responses.

JEL classification: C11; C32.

Keywords: Structural vector autoregressions, priors, posteriors, impulse responses, joint inference.

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 for estimating dynamic causal effects in macroeconomics. Many researchers use variants of the methods proposed by Uhlig (2005) and Rubio-Ramírez, Waggoner, and Zha (2010) to conduct Bayesian inference.[1] These conventional methods can be used to independently draw from any posterior distribution over the parameterization of interest subject to the identifying restrictions. Typically, the parameterization of interest consists of the impulse responses and the posterior is conjugate.

When working within this typical framework, the conventional methods boil down to independently drawing from a conjugate uniform-normal-inverse-Wishart posterior distribution over the orthogonal reduced-form parameterization and transforming the draws into the parameterization of interest. A central ingredient underlying such an approach is the uniform prior distribution over the set of orthogonal matrices with respect to the Haar measure. 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 strongly caution against using it in applied work.

Both Baumeister and Hamilton (2015) and Watson (2020) adopt a similar line of reasoning to support their critique. First, they abstract from uncertainty about the reduced-form parameters by fixing them. Second, they draw orthogonal matrices from the uniform distribution and argue that the prior distributions over the identified sets of individual impulse responses may be nonuniform.[2] Because the prior and posterior coincide over the identified sets, posterior distributions over the identified sets of individual impulse responses may also be nonuniform. As a consequence, they conclude that posterior inference could be governed by the prior over the set of orthogonal matrices.[3] Baumeister and Hamilton (2015) claim that "users of these methods can in some cases end up performing hundreds of thousands of calculations, ostensibly analyzing the data, but in fact doing nothing more

[1] See also Faust (1998) and Canova and De Nicoló (2002).

[2] For each value of the reduced-form parameters, the identified set is all impulse responses, for a particular variable, shock, and horizon, associated with the given reduced-form parameters that also satisfy the sign restrictions.

[3] See also Wolf (2020).

than generating draws from a prior distribution that they never even acknowledged assuming" (pages 1964-1965). Similarly, Watson (2020) writes that "'good' priors lead to good inference and conversely for bad priors. Sorting out the good from the bad requires careful presentation and justification for the prior actually used, a point forcefully and convincingly made in theory and practice in Baumeister and Hamilton (2015, 2019). In this regard, the kinds of flat (Haar) priors made on the rotation matrix [...] seem counterproductive" (page 192). Baumeister and Hamilton (2015) go a step further and argue that the search for uniform priors over any objects of interest is ill-fated: "Because the objects of interest in structural VARs are highly nonlinear functions of the underlying parameters, the quest for 'noninformative' priors for structural VARs is destined to fail" (page 1979).

This paper accomplishes three objectives. First, we illustrate the consequences of fixing the reduced-form parameters using the empirical example in Watson (2020). We begin by demonstrating that the conditional prior distributions of individual impulse responses tend to differ greatly from the unconditional prior distributions of individual impulse responses embodied in the conventional approach.[4] It is important to clarify that it is not the case that either Baumeister and Hamilton (2015) or Watson (2020) explicitly equalize conditional to unconditional priors in finite samples. However, Baumeister and Hamilton (2015) claim that "applied researchers often ignore posterior uncertainty about" reduced-form parameters (page 1975). We are not aware of any applied work that ignores posterior uncertainty about reduced-form parameters, hence, we think that the differences between conditional and unconditional priors need to be clarified. Then, we show that if one employs the standard priors used in the literature that allow for reduced-form parameter uncertainty, generally the posterior distributions of individual impulse responses induced by the conventional method are not an artifact of the priors. We illustrate by example that posterior medians and probability intervals tend to be quite different from the corresponding statistics based on the prior. Second, we demonstrate that, although the prior and posterior distributions over the identified sets of *individual* impulse responses implicit in the conventional method

[4]The conditional prior (posterior) distributions of individual impulse responses are equivalent to the prior (posterior) distributions over the identified sets of individual impulse responses and they are obtained after conditioning on the reduced-form parameters and then marginalizing out all but an individual impulse response. The unconditional prior (posterior) distributions of individual impulse responses are equivalent to the prior (posterior) distributions of individual impulse responses and they are obtained after marginalizing out all but an individual impulse response without conditioning on the reduced-form parameters. By individual impulse response we mean the response of a single variable to a single shock at a single horizon.

may be nonuniform, the conventional method induces uniform joint prior and posterior distributions over the identified set for the *vector* of impulse responses. There is a growing literature making the case that only joint distributions capture the shape and co-movement of the responses, which is generally the ultimate interest of empirical studies (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). Thus, it is essential that we take a joint approach rather than the more traditional marginal one employed by Baumeister and Hamilton (2015) and Watson (2020). Third, we show that Baumeister and Hamilton's (2015) conjecture that "the quest for 'noninformative' priors for structural VARs is destined to fail" is not true. 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. We generalize the construction and implementation of a uniform joint prior over a broader class of objects of interest.

First, using the empirical example in Watson (2020), we provide evidence that, although it is true that the conditional prior distributions of individual impulse responses may be nonuniform, the influence of the prior on posterior inference is not as severe as suggested by Baumeister and Hamilton (2015) and Watson (2020). The key is not to condition on the reduced-form parameters, and to recognize the reduced-form parameter uncertainty embedded in the conventional approach. We show that the difference between the conditional and unconditional prior distributions of individual impulse responses tends to be striking. In our example, the probability intervals based on the unconditional and conditional priors are disjoint for many combinations of variables, shocks, and horizons. More importantly, using the standard Minnesota prior distribution over the reduced-form parameters, we show that the unconditional posterior distributions of individual impulse responses typically are not an artifact of the prior. The probability intervals of the unconditional prior and posterior distributions of several individual impulse responses are also disjoint for many combinations. Other empirical studies, Inoue and Kilian (2022b) and Shin and Zhong (2020), also find that the unconditional posterior distributions of individual impulse responses do not simply reproduce the prior. In addition, using the multiple-prior Bayesian approach described in Giacomini and Kitagawa (2021), we confirm that, for this empirical example, posterior

inference about the mean of the impulse responses is robust to the uniform prior over the set of orthogonal matrices embedded in the conventional method.

Second, Baumeister and Hamilton (2015) and Watson (2020) express concern about the fact that the conventional approach induces nonuniform prior distributions over the identified sets of individual impulse responses because the prior and posterior coincide over identified sets. While this fact could be an issue in the hypothetical case when the number of observations is large enough that reduced-form parameter uncertainty can be disregarded, Inoue and Kilian (2022b) demonstrate that this concern may be ignored in tightly identified models. To further ease this concern, we show that the conventional method induces uniform joint prior and posterior distributions over the identified set for the vector of impulse responses. This is an "if and only if" theoretical result that holds for any prior distribution over the reduced-form parameters, as long as the prior distribution over the orthogonal matrices is uniform. Any other choice of prior over the set of orthogonal matrices will imply nonuniform joint prior and posterior distributions over the identified set for the vector of impulse responses. While having uniform joint prior and posterior distributions over the identified set for the vector of impulse responses is not a required feature, it is a desirable one. By construction, the likelihood is uniform over the identified sets, and as a result, having uniform joint prior and posterior distributions over the identified set for the vector of impulse responses assures the researcher that only the identifying restrictions will set apart observationally equivalent vectors of impulse responses.

Third, we show that Baumeister and Hamilton's (2015) conjecture that "the quest for 'noninformative' priors for structural VARs is destined to fail" is not true. As it is common, we begin by considering the vector of impulse responses as the vector of objects of interest. Subsequently, we generalize our findings to a more general class of objects of interest. In particular, we show that a uniform joint prior distribution for the vector of impulse responses induces a particular (model dependent) prior distribution over the reduced-form parameters and a uniform prior distribution over the orthogonal matrices. This theoretical result is also an "if and only if" statement. Any other choice of prior over the set of orthogonal matrices will imply a nonuniform joint prior distribution for the vector of impulse responses. Interestingly, the prior distribution over the reduced-form parameters required for this result differs from the standard Minnesota prior. It is similar in spirit to (although also different than) the "weak" prior described in Uhlig (2005). We show that the induced prior over the

orthogonal reduced-form parameterization defines a uniform-normal-inverse-Wishart posterior distribution over the orthogonal reduced-form parameterization. This allows us to adapt 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. Obviously, because of the uniform prior distribution over the 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 slightly wider than those obtained under the uniform-normal-inverse-Wishart prior distribution over orthogonal reduced-form parameterization 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 results to a broader class of objects of interest.[5] Specifically, for any objects of interest within this class, 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 over the impact impulse responses of price and quantity to demand and supply shocks, the short-term prices elasticities of demand and supply, and some lag structural coefficients. In this case, the vector of objects of interest consists of the coefficients associated with the two impact impulse responses, the two short-term price elasticities, and the lag parameters. Each particular vector of objects of interest induces a particular prior distribution over the orthogonal reduced-form parameterization. This induced prior is also model dependent but need not be uniform over the set of orthogonal matrices. In the latter case, it is necessary to add an importance sampling step to draw from the induced joint posterior distribution for the vector of objects of interest. Using a simplified version of the labor market model described in Baumeister and Hamilton (2015), we compare the joint credible sets for the vector of objects of interest to those induced by the conventional uniform-

[5]See Section 6 for a formal definition of the class of objects of interest.

normal-inverse-Wishart prior over orthogonal reduced-form parameterization associated with a standard Minnesota prior.[6] Although the posterior credible sets are similar regardless of the priors under analysis, our results reinforce the earlier conclusion that imposing tighter identifying restrictions could help to reduce joint posterior uncertainty.

The structure of the paper is as follows. Section 2 describes the conventional method. Section 3 provides evidence that the importance of the uniform prior distribution over the set of orthogonal matrices for posterior impulse responses inference emphasized by Baumeister and Hamilton (2015) and Watson (2020) is overstated. Section 4 proves that the conventional approach implies a uniform joint prior distribution over the identified set for the vector of impulse responses. Section 5 shows how to define a uniform joint prior distribution for the vector of impulse responses and how to adapt the conventional method to implement it. Section 6 generalizes this result to other vectors of objects of interest. Finally, Section 7 concludes.

2 The Conventional Approach

Let the SVAR be represented by:

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

where, for $1 \leq t \leq T$, $\mathbf{y}_t$ is an $n \times 1$ vector of endogenous variables, $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ vector of structural shocks, $\mathbf{x}_t' = \left[\mathbf{y}_{t-1}' \cdots \mathbf{y}_{t-p}' \; 1\right]$, $\mathbf{A}_+' = \left[\mathbf{A}_1' \cdots \mathbf{A}_p' \;\; \mathbf{c}'\right]$, $\mathbf{A}_\ell$ 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. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions, is Gaussian with mean zero and covariance matrix $\boldsymbol{I}_n$, the $n \times n$ identity matrix. We call $\mathbf{A}_0$ and $\mathbf{A}_+$ the structural parameters and we refer to $(\mathbf{A}_0, \mathbf{A}_+)$ as the structural parameterization. In the class of linear Gaussian models under analysis, $(\mathbf{A}_0, \mathbf{A}_+)$ 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. Hence, the structural parameters are not identified.

The reduced-form representation of Equation (1) is $\mathbf{y}_t' = \mathbf{x}_t' \mathbf{B} + \mathbf{u}_t'$, for $1 \leq t \leq T$,

[6] See Section 5 in Baumeister and Hamilton (2015) for the description of the full model.

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. Given any decomposition of the covariance matrix $\boldsymbol{\Sigma}$ satisfying $h(\boldsymbol{\Sigma})'h(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}$, we can define a mapping from $(\mathbf{A}_0, \mathbf{A}_+)$ to $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ by $f_h(\mathbf{A}_0, \mathbf{A}_+) = \left(\mathbf{A}_+\,\mathbf{A}_0^{-1}, (\mathbf{A}_0\,\mathbf{A}_0')^{-1}, h\left((\mathbf{A}_0\,\mathbf{A}_0')^{-1}\right)\mathbf{A}_0\right)$. We will take h to be the upper triangular Cholesky decomposition, normalized so that the diagonal is positive, though any differentiable decomposition would do. The function f_h is invertible, with its inverse defined by $f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = (h(\boldsymbol{\Sigma})^{-1}\mathbf{Q}, \mathbf{B}\,h(\boldsymbol{\Sigma})^{-1}\mathbf{Q})$. This makes clear how the structural parameters depend on the reduced-form parameters and orthogonal matrices. We call $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ the orthogonal reduced-form parameterization. Notice that Equation (1) can alternatively be written as $\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \boldsymbol{\varepsilon}_t'\mathbf{Q}'h(\boldsymbol{\Sigma})$ for $1 \leq t \leq T$. As we explain in Section 2.1, the orthogonal reduced-form parameterization is convenient to obtain independent draws from the well-known uniform-normal-inverse-Wishart distribution. Then, these independent draws are mapped into structural parameters or impulse responses, which are in turn used to independently sample from any distribution over the parameterization of interest such as the normal-generalized-normal distribution over the structural parameters used in Arias, Rubio-Ramírez, and Waggoner (2018).

To solve the identification problem, one often imposes sign and/or zero restrictions on either the structural parameters or some function of the structural parameters. The theory and simulation techniques that we develop apply to sign and zero restrictions on any differentiable function $\boldsymbol{F}(\mathbf{A}_0, \mathbf{A}_+)$ from the structural parameters to the space of $r \times n$ matrices that satisfies the condition $\boldsymbol{F}(\mathbf{A}_0\,\mathbf{Q}, \mathbf{A}_+\,\mathbf{Q}) = \boldsymbol{F}(\mathbf{A}_0, \mathbf{A}_+)\,\mathbf{Q}$, for every $\mathbf{Q} \in \mathcal{O}(n)$, which, for example, is true for impulse responses. For the sake of clarity, the exposition of the algorithms and the theory developed in this paper will consider only sign restrictions, but they can be easily extended to zero restrictions, as we highlight in the conclusion. If only sign restrictions are imposed, the function $\boldsymbol{F}$ needs only to be continuous. The sign restrictions are imposed using $s_j \times r$ full row rank matrices $\boldsymbol{S}_j$, where $0 \leq s_j$, for $1 \leq j \leq n$. Each matrix $\boldsymbol{S}_j$ will define the sign restrictions on the j^{th} structural shock for $1 \leq j \leq n$. In particular, we assume that $\boldsymbol{S}_j\boldsymbol{F}(\mathbf{A}_0, \mathbf{A}_+)\boldsymbol{e}_j > \mathbf{0}$ for $1 \leq j \leq n$, where $\boldsymbol{e}_j$ is the j^{th} column of $\boldsymbol{I}_n$.

Importantly, 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 structural parameters, 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.

2.1 The Prior and Posterior Distributions

The conventional methods use a normal-inverse-Wishart distribution prior over the reduced-form parameters.[7] If such prior is $NIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$, then the posterior distribution over the reduced-form parameters is $NIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$, where $\tilde{\nu} = T + \bar{\nu}$, $\tilde{\mathbf{\Omega}} = (\mathbf{X}'\mathbf{X} + \bar{\mathbf{\Omega}}^{-1})^{-1}$, $\tilde{\mathbf{\Psi}} = \tilde{\mathbf{\Omega}}(\mathbf{X}'\mathbf{Y} + \bar{\mathbf{\Omega}}^{-1}\bar{\mathbf{\Psi}})$, $\tilde{\mathbf{\Phi}} = \mathbf{Y}'\mathbf{Y} + \bar{\mathbf{\Phi}} + \bar{\mathbf{\Psi}}'\bar{\mathbf{\Omega}}^{-1}\bar{\mathbf{\Psi}} - \tilde{\mathbf{\Psi}}'\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{\Psi}}$, 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 over the orthogonal reduced-form parameterization is uniform-normal-inverse-Wishart. We denote it by $UNIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$. This prior is conjugate and the implied posterior distribution over the orthogonal reduced-form parameterization is also uniform-normal-inverse-Wishart, denoted by $UNIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$.

Arias, Rubio-Ramírez, and Waggoner (2018) show that the density over the structural parameterization induced by a uniform-normal-inverse-Wishart density over the orthogonal reduced-form parameterization is proportional to a normal-generalized-normal density:

$$NGN_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{A}_0, \mathbf{A}_+) \propto \underbrace{|\det(\mathbf{A}_0)|^{\nu-n} e^{-\frac{1}{2}\operatorname{vec}(\mathbf{A}_0)'(\boldsymbol{I}_n \otimes \mathbf{\Phi})\operatorname{vec}(\mathbf{A}_0)}}_{\text{generalized-normal}} \underbrace{e^{-\frac{1}{2}\operatorname{vec}(\mathbf{A}_+ - \mathbf{\Psi}\mathbf{A}_0)'(\boldsymbol{I}_n \otimes \mathbf{\Omega})^{-1}\operatorname{vec}(\mathbf{A}_+ - \mathbf{\Psi}\mathbf{A}_0)}}_{\text{conditionally normal}}.$$

Thus, if we independently draw $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ from $UNIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$ distribution and then transform the draws using f_h^{-1}, we are in fact independently drawing $(\mathbf{A}_0, \mathbf{A}_+)$ from a $NGN(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$ distribution over the structural parameterization.

There are several routines available for making independent draws from any normal-inverse-Wishart distribution over the reduced-form parameters. Independent draws from the uniform distribution over the set of orthogonal matrices are normally based on Theorem 3.2

[7] 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 $\mathbf{\Phi}$, an $m \times n$ matrix $\mathbf{\Psi}$, and an $m \times m$ symmetric and positive definite matrix $\mathbf{\Omega}$. We denote this distribution by $NIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$ and its density by $NIW_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma})$. Furthermore,

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

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 $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}$ be the QR decomposition of $\boldsymbol{X}$ with the diagonal of $\boldsymbol{R}$ normalized to be positive. The matrix $\mathbf{Q}$ is orthogonal and drawn from the uniform distribution over $\mathcal{O}(n)$.*

2.2 The Algorithm

The preceding discussion justifies using Algorithm 1 to draw from the normal-generalized-normal posterior distribution over the structural parameterization conditional on the sign restrictions.

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

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

2.3 The Impulse Response Parameterization

In some cases, the researcher will be interested in using a conjugate prior distribution over the impulse response parameterization. Henceforward, we will refer to this representation as the IR parameterization, defined as $(\mathbf{L}_0, \cdots, \mathbf{L}_p, \mathbf{c})$, where p is the number of lags, the element in row i and column j of the $n \times n$ matrix $\mathbf{L}_k$ is the response of the i^{th} variable to the j^{th} structural shock at horizon k, and $\mathbf{c}$ is the $1 \times n$ constant term from the structural parameterization. The matrices $\mathbf{L}_k$ are functions of the structural parameterization and they are recursively defined by $\mathbf{L}_0 = \left(\mathbf{A}_0^{-1}\right)'$ and $\mathbf{L}_k = \sum_{\ell=1}^{k} \left(\mathbf{A}_\ell \mathbf{A}_0^{-1}\right)' \mathbf{L}_{k-\ell}$, for $1 \leq k \leq p$. The matrices $\mathbf{A}_k$ are also functions of the IR parameterization and they are recursively defined by $\mathbf{A}_0 = \left(\mathbf{L}_0^{-1}\right)'$ and $\mathbf{A}_k = \left(\mathbf{L}_k \mathbf{L}_0^{-1}\right)' \mathbf{A}_0 - \sum_{\ell=1}^{k-1} \left(\mathbf{L}_{k-\ell} \mathbf{L}_0^{-1}\right)' \mathbf{A}_\ell$, for $1 \leq k \leq p$. Let $\mathbf{L}_+' = \begin{bmatrix} \mathbf{L}_1' & \cdots & \mathbf{L}_p' & \mathbf{c}' \end{bmatrix}$. Then, the IR parameterization can be represented by $(\mathbf{L}_0, \mathbf{L}_+)$. Importantly, when referring to

this parameterization in vector form we will use the term vector of impulse responses. We will denote the mapping from the IR parameterization to the structural parameterization by f_{ir} and the mapping from the IR parameterization to the orthogonal reduced-form parameterization by $\phi_h = f_h \circ f_{ir}$.

The advantage of the IR parameterization is that statements about the marginal distribution of individual impulse responses, which are complicated non-linear functions of the structural parameters, directly translate into statements about the marginal distribution for individual parameters in the IR parameterization, at least through horizon p. Similarly, statements about the joint distribution for the vector of impulse responses through horizon p directly translate into statements about the joint distribution over the IR parameterization.

Within the typical framework described in the introduction, Algorithm 1 can be used to draw from a conjugate posterior distribution over the IR parameterization conditional on the sign restrictions. We will just need to modify Step 3 to consider the following transformation from the orthogonal reduced-form representation to the IR representation $(\mathbf{L}_0, \mathbf{L}_+) = \phi_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.

3 Re-Examining Watson (2020)

Baumeister and Hamilton (2015) and Watson (2020) report the prior distributions over the identified sets of selected individual impulse responses induced by Algorithm 1. These distributions are obtained by replacing Step 1 with a fixed value of the reduced-form parameters. For example, on page 1972, Baumeister and Hamilton (2015) write "this algorithm can be viewed as generating draws from a Bayesian prior distribution for $\mathbf{H}$ [$\mathbf{L}_0$ in our notation] conditional on $\mathbf{\Omega}[\mathbf{\Sigma}]$," and document the properties of such an approach. We will refer to these prior distributions as the conditional prior distributions of individual impulse responses to emphasize that they do condition on the reduced-form parameters. Both papers claim that the conditional prior distributions of individual impulse responses may be nonuniform and, hence, call into question the results obtained by the conventional methods. In this section, we analyze these two claims in the context of the empirical application in Watson (2020).

3.1 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 capita in the nonfarm business sector, n_t the logarithm of hours worked per capita, p_t the logarithm of the price level, and i_t^L the 10-year Treasury bond rate.[8] 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 the sign and zero restrictions described in Table 1.[9]

Variable \Shock	Technology	Demand	Monetary Policy	Supply
Restrictions on 4-quarter ahead IRs				
Output		–	–	–
Price Level		–	–	
Inflation				+
10-Year Treasury Bond Rate		–	+	
Restrictions on long-run IRs				
Labor Productivity Growth		0	0	0

Table 1: Identification Restrictions

We parameterize reduced-form parameter uncertainty as a standard Minnesota prior. We set $\bar{\nu} = n + 2$, which is the minimum value $\bar{\nu}$ can take that guarantees the existence of a prior mean for $\boldsymbol{\Sigma}$. The matrix $\bar{\boldsymbol{\Phi}}$ is diagonal, with $\bar{\boldsymbol{\Phi}} = \text{diag}(\phi_1, \phi_2, \phi_3, \phi_4)$. The values for $\bar{\boldsymbol{\Psi}}$ and $\bar{\boldsymbol{\Omega}}$ are chosen to have a flat density over the constant term ($\text{Var}(\mathbf{d} \mid \boldsymbol{\Sigma}) = 10^7 \boldsymbol{\Sigma}$) and the following first and second moments over the slope parameters:

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

[8] We kindly obtained this data from Mark Watson and it can be replicated with the following mnemonics from the FRED database: real output per hour of all persons in the non-farm business sector (OPHNFB), hours of all persons in the nonfarm business sector (HOANBS), civilian noninstitutionalized population (CNP16OV), GDP deuniformor (GDPDEF), and the 10-Year Treasury Constant Maturity Rate (GS10).

[9] We also impose stability of the VAR throughout our paper. This is done by discarding the unstable draws in Step 1 of Algorithm 1, which is a type of sign restriction.

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

where $\mathbf{B}_\ell = \mathbf{A}_\ell\mathbf{A}_0^{-1}$ and $\mathbf{d} = \mathbf{c}\mathbf{A}_0^{-1}$. We will treat λ and $\bar{\boldsymbol{\Phi}}$ as hyperparameters. We follow Giannone, Lenza, and Primiceri (2015) in choosing the values for these parameters that maximize the marginal likelihood. This yields $\lambda = 0.3453$, and $\bar{\boldsymbol{\Phi}} = \text{diag}(2.5217, 0.3497, 0.0478, 0.1724)$.

Following the conventional method described in Section 2, we assume a uniform prior distribution over the set of orthogonal matrices. The resulting uniform-normal-inverse-Wishart prior over the orthogonal reduced-form parameterization induces a conjugate prior distribution over the IR parameterization. The zero restrictions on the long-run impulse responses have a particular structure that can be exploited to draw from the conjugate posterior distribution over the IR parameterization conditional on the sign and zero restrictions using Algorithm 1 with Step 2 slightly modified. In Appendix A we show that, 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.

3.2 Nonuniform Priors

In order to compute conditional prior distributions of individual impulse responses, Watson (2020) sets the reduced-form parameters equal to $\hat{\mathbf{B}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\mathbf{Y}$ and $\hat{\boldsymbol{\Sigma}} = \frac{\hat{\mathbf{S}}}{T+n}$, where $\hat{\mathbf{S}} = (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})$. Ignoring reduced-form parameter uncertainty is either justified on the grounds that, in the hypothetical case, when the number of observations is large enough the reduced-form parameters are pinned down (as in Watson, 2020) or by arguing that "applied researchers often ignore posterior uncertainty about $\boldsymbol{\Omega}$ [$\boldsymbol{\Sigma}$] and simply condition on the average residual variance" (as claimed by Baumeister and Hamilton, 2015). However, we are not aware of any applied work on set-identified Bayesian SVARs where reduced-form parameter uncertainty is either ignored or negligible.

The light and dark red areas in Figure 1, which reproduce Figure 6 in Watson (2020), show the equal-tailed 68 and 95 percent conditional prior probability intervals of individual

impulse responses. Based on this figure, Watson (2020) states (on page 189) that the "prior on $\mathbf{R}$ [$\mathbf{Q}$] together with the authors' equality and sign restrictions imply quite informative priors on the impulse responses" over the identified sets.[10] Figure 1 in Baumeister and Hamilton (2015) makes similar arguments using a more general framework. While this claim is correct, the conditional prior distributions of individual impulse responses are not the prior distributions of individual impulse responses induced by Algorithm 1. We will refer to the prior distributions of individual impulse responses as the unconditional prior distributions of individual impulse responses to emphasize that they do not condition on the reduced-form parameters.

As made clear by Inoue and Kilian (2022b), it is the interaction of the prior over the set of orthogonal matrices with the priors over the reduced-form parameters that determines the shape of the prior distributions of individual impulse responses implicit in the conventional approach. Thus, it is incorrect to view the algorithm "as generating draws from a Bayesian prior distribution for $\mathbf{L}_0$ conditional on $\mathbf{\Sigma}$." Figure 1 shows that there are substantial differences between the conditional and the unconditional prior distributions in the application under consideration. The dark and light gray areas show the equal-tailed 68 and 95 percent unconditional prior distributions of individual impulse responses, respectively. Noticeably, the probability intervals based on the unconditional and conditional priors are disjoint for several variable, shock, and horizon triplets. The response of the long interest rate to a demand shock four quarters after the shock provides an example: While the 68 percent probability interval based on the unconditional prior is $[-0.04, 0.00]$, the same interval for the conditional prior is $[-0.29, -0.09]$. The same is true for the response of the real rate to a supply shock eight quarters after the shock: The 68 percent probability interval based on the unconditional prior is $[-0.02, 0.01]$, while the same interval for the conditional prior is $[0.12, 0.23]$. When the probability intervals are not disjoint, the figure shows that in general the distributions either have very different medians or convey very different degrees of uncertainty. Thus, Figure 1 shows that Baumeister and Hamilton (2015) and Watson (2020) mischaracterize the unconditional prior distributions of individual impulse responses induced by Algorithm 1.

To make even more clear that it is not the case that the conditional prior distributions of individual impulse responses tell us anything about the unconditional prior distributions

[10]The figures reported in this paper address some errata in the original code used in Watson (2020). The figures reported here reflect these corrections.

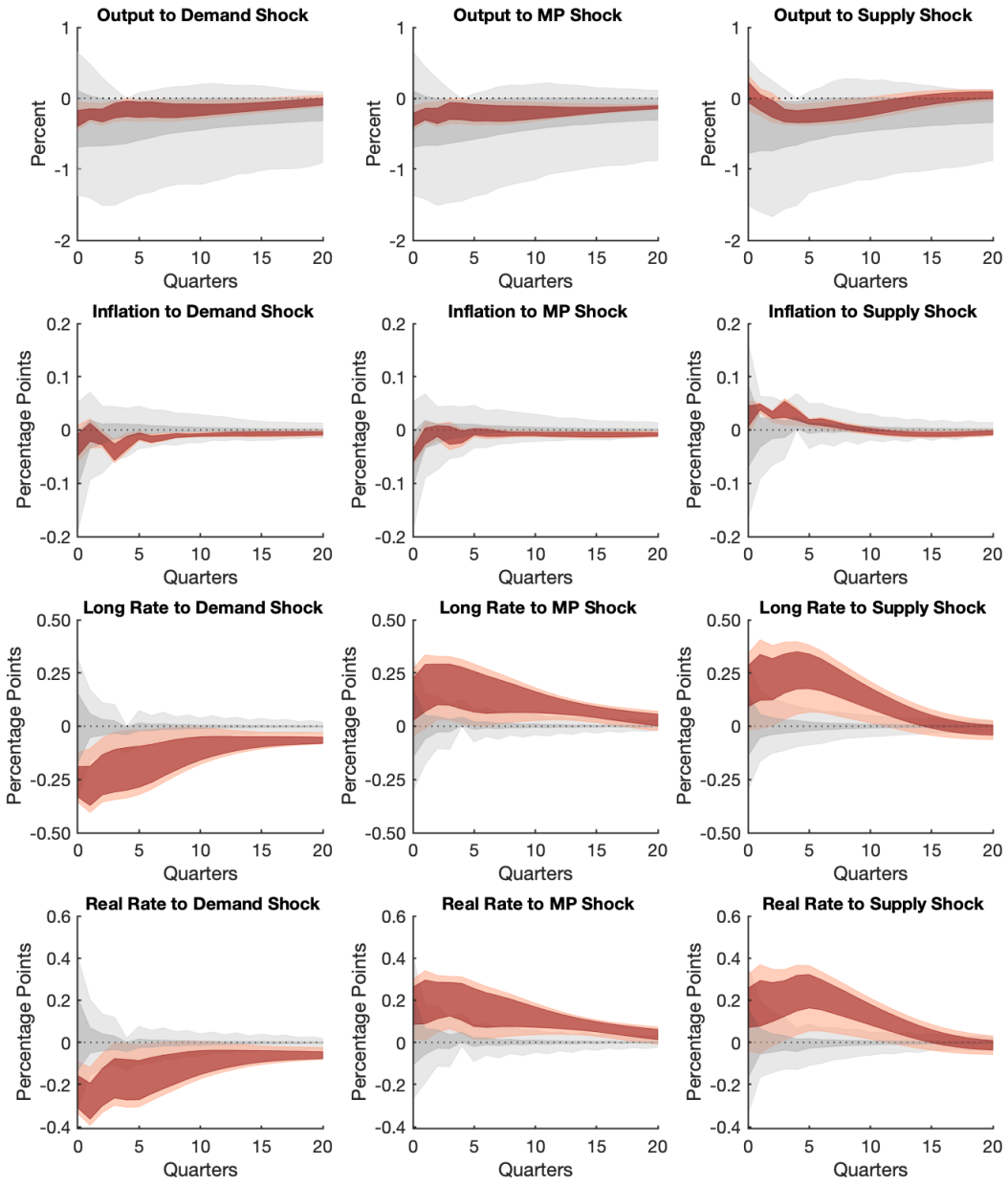


Figure 1: The dark (light) red areas show the equal-tailed 68 (95) percent conditional prior distributions of individual impulse responses. The dark (light) gray areas show the equal-tailed 68 (95) percent unconditional prior distributions of individual impulse responses.

of individual impulse responses, Figure 2 compares the conditional and unconditional prior distributions for the response of inflation to a demand shock four quarters after the shock. A user of the conditional prior distribution would conclude that the prior has negative support

while about 50 percent of the unconditional prior probability mass is above zero.

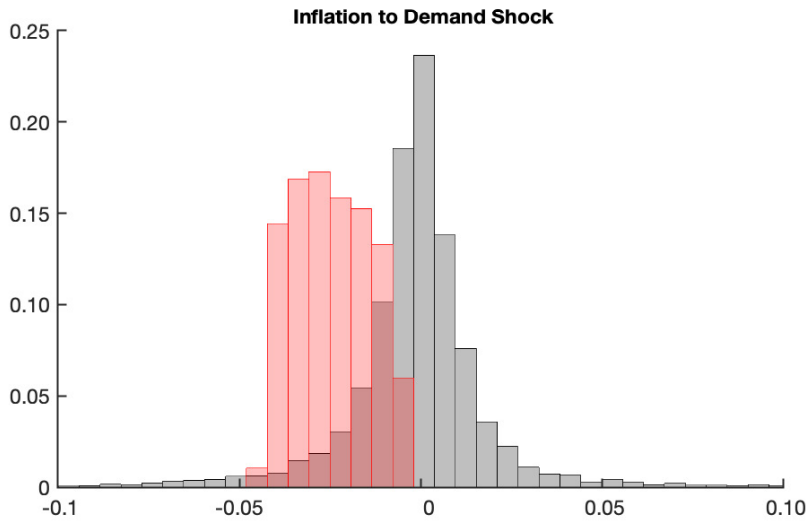


Figure 2: The red histograms show the conditional prior distribution for the response of inflation to a demand shock four quarters after the shock. The gray histograms show the unconditional prior distribution for the response of inflation to a demand shock four quarters after the shock.

If one wants to assess the role of the prior in driving the posterior distributions of individual impulse responses obtained using Algorithm 1, then one has to compare the unconditional prior and posterior distributions of individual impulse responses. Figure 3 shows that there are important differences between the unconditional posterior distributions of individual impulse responses in the application under consideration. The dark and light gray (green) areas show the equal-tailed 68 and 95 percent unconditional prior (posterior) distributions of individual impulse responses. As was the case when comparing the unconditional and conditional priors, the probability intervals based on the unconditional prior and posterior distributions of individual impulse responses are disjoint for several variable, shock, and horizon triplets. The response of the real rate to a monetary policy shock four quarters after the shock provides an example: While the 68 percent probability interval based on the unconditional prior is $[0.00, 0.05]$, the 68 percent probability interval based on the posterior is $[0.06, 0.32]$. The response of inflation to a supply shock one quarter after the shock provides another example: While the 68 percent probability interval based on the unconditional prior is $[-0.03, 0.02]$, the 68 percent probability interval based on the posterior is $[0.02, 0.06]$, which indicates that the data support the view that positive supply shocks lead to increases in the

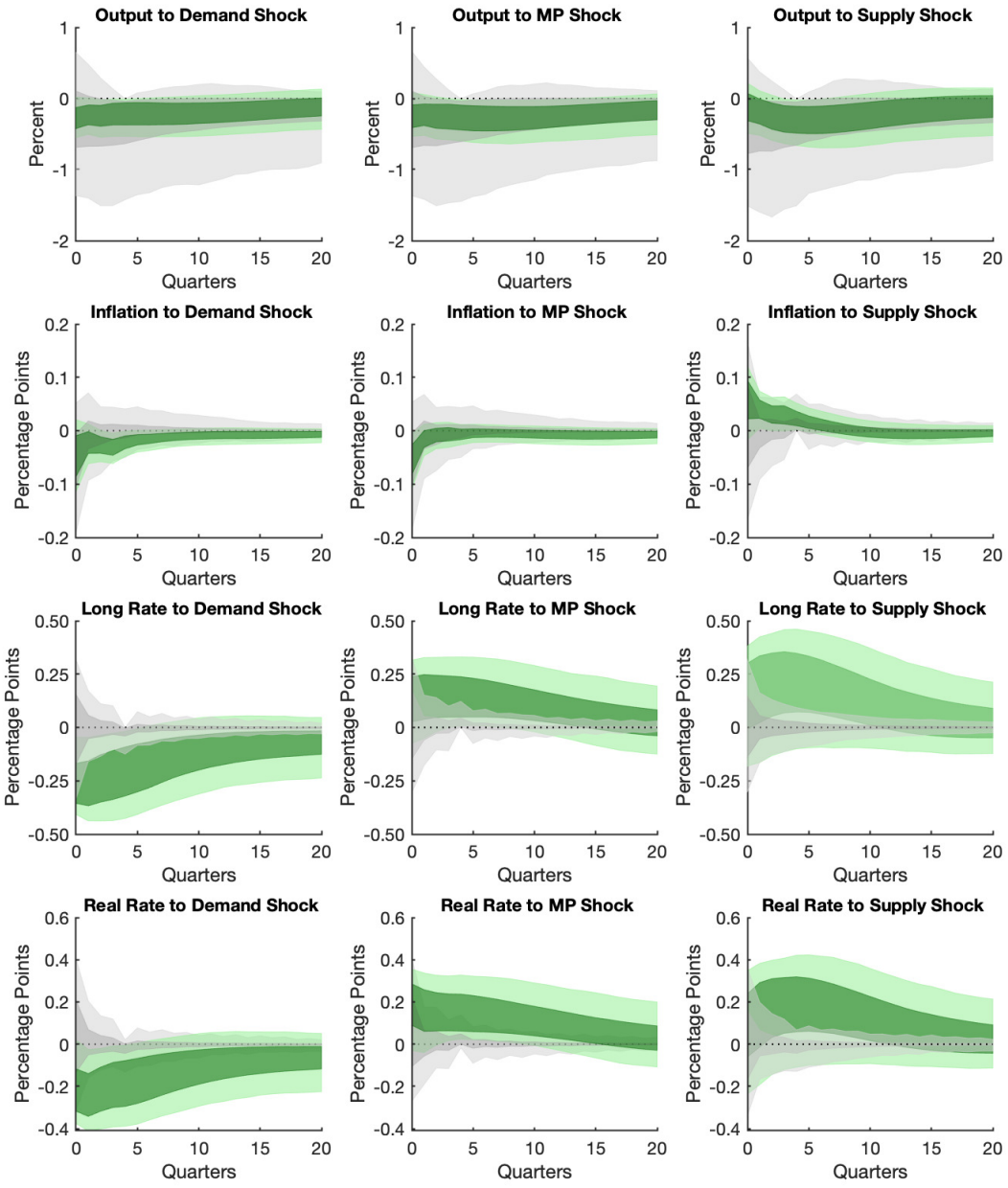


Figure 3: The dark (light) gray areas show the equal-tailed 68 (95) percent prior distributions of individual impulse responses. The dark (light) green areas show the equal-tailed 68 (95) percent posterior distributions of individual impulse responses.

inflation rate. As before, if the probability intervals are not disjoint, the figure shows that in general the prior and posterior distributions either have very different medians or convey very different degrees of uncertainty.

To reinforce that the unconditional prior distributions of individual impulse responses do not drive the unconditional posterior distributions of individual impulse responses, Figure 4 compares the unconditional prior and posterior distributions for the response of inflation to a demand shock four quarters after the shock. While most of the probability mass of the unconditional posterior distribution is below zero, about 50 percent of the unconditional prior probability mass is above zero.

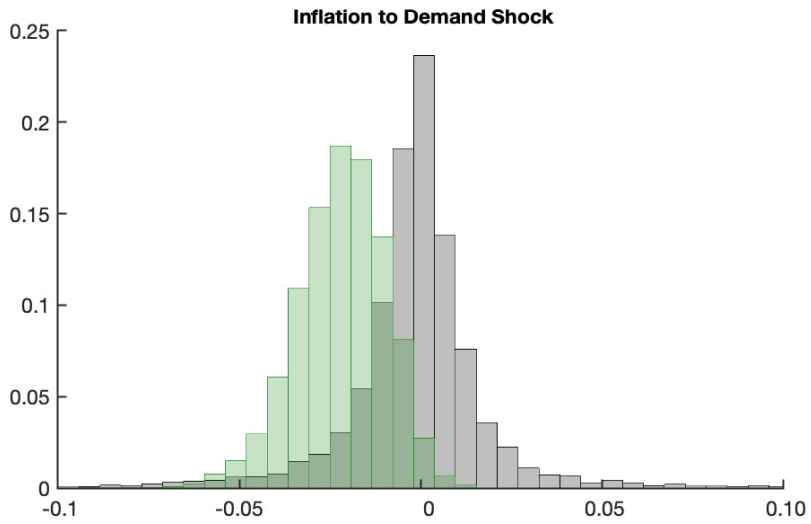


Figure 4: The gray histograms show the unconditional prior distribution for the response of inflation to a demand shock four quarters after the shock. The green histograms show the unconditional posterior distribution for the response of inflation to a demand shock four quarters after the shock.

3.3 Multiple-Prior Bayesian Approach

We think that the results above convincingly show that the unconditional posterior distributions of individual impulse responses are not an artifact of the unconditional prior distributions of individual impulse responses. But it is the case that the prior over the set of orthogonal matrices is not updated by the data. In addition, there is a concern that the prior over the impulse responses implied by the conventional method could induce researchers to report bands that would be too narrow from a frequentist perspective: Baumeister and Hamilton (2022) "The confidence bands used by practitioners are much too narrow from the perspective of a frequentist who is unpersuaded by the implicit prior information that

underlies the popular methods" (page 3). To address these issues, Giacomini and Kitagawa (2021) adopt a multiple-prior Bayesian approach for inference in set-identified models like the ones described in this paper. This novel and interesting robust Bayesian approach removes the need to specify the prior for the orthogonal matrices given the reduced-form parameters. Among other features, this method allows the researcher to analyze the sensitivity of the posterior distributions of individual impulse responses to the choice of the uniform prior over orthogonal matrices by reporting the set of posterior means for a given prior over the reduced-form parameters as well as robust credible regions. Figure 5 illustrates Giacomini and Kitagawa's (2021) procedure in the context of our application. The dashed black lines are the lower and upper bound posterior means. The magenta dashed lines show the smallest robust credible regions with credibility 68 percent.

Although the 68 percent robust credible regions are clearly wider than the equal-tailed 68 percent posterior probability bands (shown by the dark green areas), the lower and upper bounds for the posterior means (depicted by the dashed black lines) show that the identifying restrictions convey a reasonable amount of information. To simplify the exposition, let us classify the impulse responses implied by the conventional method into two groups: one where the equal-tailed 68 percent posterior probability bands and the set of posterior means simultaneously have either positive or negative support at some horizon, and another where there is no horizon for which the latter occurs. Under this classification, the conventional methods provide "convincing" evidence about all the impulse responses except for the ones of the long (nominal) and real rate to a supply shock, at least at some horizon.

Even though there are also impulse responses—such as the impact impulse responses of the real rate—for which the lower and upper bounds for the posterior means contain zero while the conventional methods provide "convincing" evidence, our general assessment of Figure 5 is that the gap between the conclusions obtained under the conventional and the robust Bayesian approach is not large.

3.4 Summary

To summarize, Figures 1 through 5 reveal four features of the conventional methods. First, as suggested by Baumeister and Hamilton (2015) and Watson (2020), Algorithm 1 may generate nonuniform prior distributions over the identified sets of individual impulse responses, but we

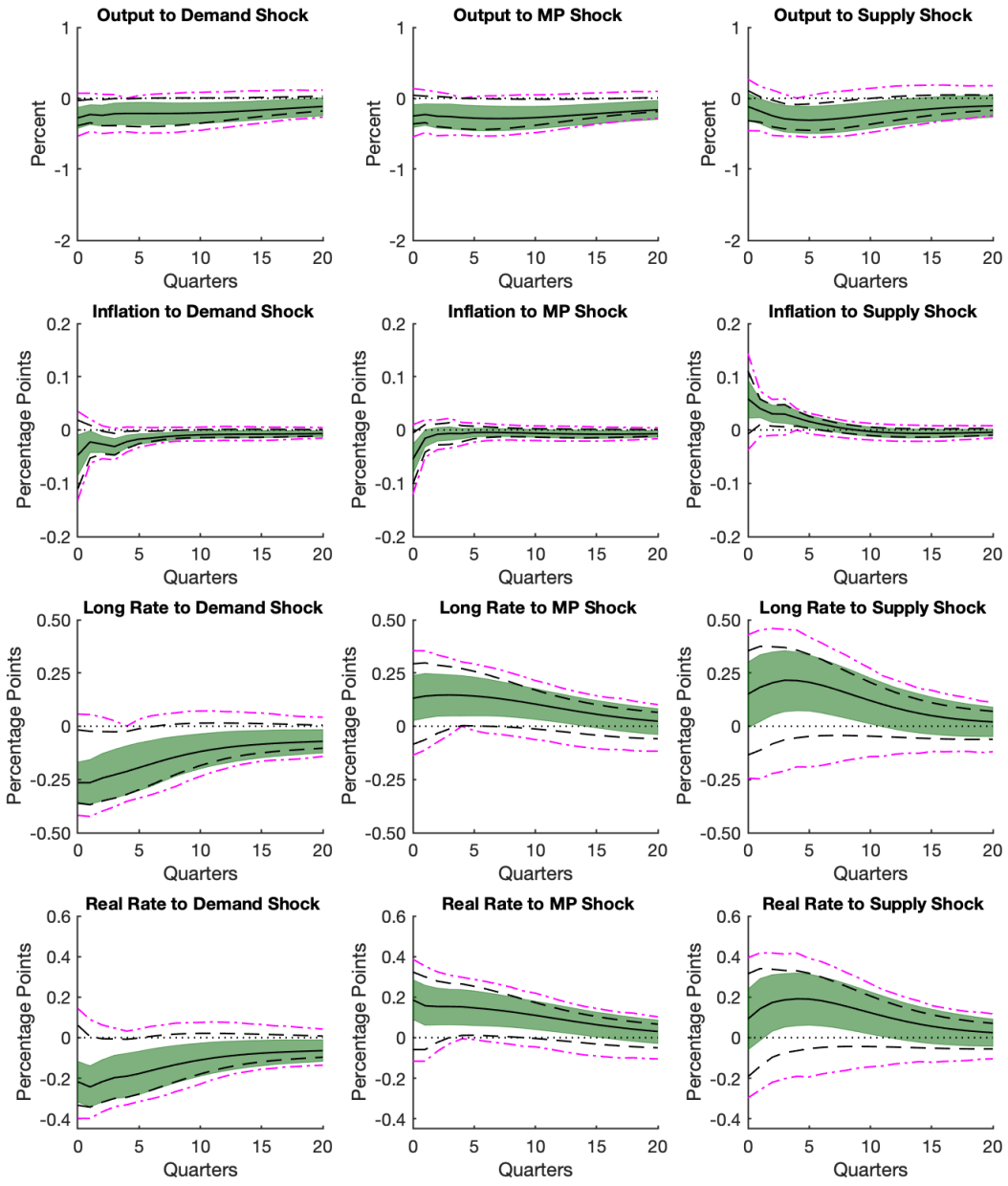


Figure 5: The solid black lines (dark green areas) show the posterior means (equal-tailed 68 percent posterior probability bands) of individual impulse responses. The dashed black lines (dash-dotted magenta lines) are the lower and upper bound posterior means (smallest robust credible region with credibility 68 percent) implied by Giacomini and Kitagawa's (2021) procedure.

are not aware of a single Bayesian application of the conventional approach where estimation uncertainty is negligible. Additionally, even if a researcher were to deliberately ignore reduced-

form parameter uncertainty, Inoue and Kilian (2022b) show that the nonuniformity of the prior distributions over the identified sets of individual impulse responses is not a concern in tightly identified models. In the next section, we further strengthen the rationale for the conventional approach by showing that the undesirability of the approach reflected in the nonuniformity of the prior distributions over the identified sets of individual impulse responses is a consequence of focusing on marginal distributions instead of joint ones. We demonstrate that the joint prior distribution over the identified set for the vector of impulse responses induced by the conventional methods is always uniform. This is an important result because the ultimate interest of empirical studies is in simultaneously assessing a range of different impulse responses to capture their shape and co-movement. Thus, it is essential to analyze joint distributions as opposed to the marginal distributions emphasized by Baumeister and Hamilton (2015) and Watson (2020) and still common in applied work.

Second, we showed by example that the unconditional prior distributions of individual impulse responses induced by Algorithm 1 are noticeably different from the conditional prior distributions. Hence, it is wrong to use conditional prior distributions of individual impulse responses to "characterize the prior distributions that are implicit" in the conventional approach as claimed on page 1972 of Baumeister and Hamilton (2015).

Third, while our findings and those of Shin and Zhong (2020) and Inoue and Kilian (2022b) show that results are not bound to be an artifact of the conventional prior specification, Giacomini and Kitagawa (2021) show examples where the conclusions from a prior robust approach differ from those of the standard approach and hence the role of the unconditional prior distributions of individual impulse responses in posterior inference needs to be determined on a case by case basis.

Fourth, the multiple-prior Bayesian approach of Giacomini and Kitagawa (2021) shows that although the mean impulse responses depend on the choice of prior over the set of orthogonal matrices, for this particular example the choice of a uniform prior does not seem to drive the main conclusions.

4 Conditional Joint Prior for Impulse Responses

Because the posterior reproduces the prior over the identified set, we understand why a researcher would want a uniform joint prior distribution over the identified set for the vector

of impulse responses. Oftentimes, we will refer to this prior as the conditional joint prior distribution for the vector of impulse responses.[11] Although the issue is less relevant when the identified sets are narrow (see Inoue and Kilian, 2022b), in this section we answer the following question: Are there distributions over the IR parameterization such that the conditional joint prior distribution for the vector of impulse responses is uniform? Or equivalently, are there distributions over the IR parameterization such that the joint prior distribution over the identified set for the vector of impulse responses is uniform? The answer is yes, and Propositions 2 and 3 give the conditions required for this to be the case. Interestingly, the conventional methods described in Uhlig (2005) and Rubio-Ramírez, Waggoner, and Zha (2010) imply 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. If f_p denotes the projection from the orthogonal reduced-form parameterization onto the reduced-form parameters, then $\phi = f_p \circ \phi_h$ is the mapping from the IR parameterization to the reduced-form parameters, and ϕ does not depend on h. Given the reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$, the set $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ will be the submanifold that is the support of the joint distribution of the IR parameterization conditional on $(\mathbf{B}, \mathbf{\Sigma})$. The submanifold structure induces a natural measure over $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$, which is called the volume measure.[12] If $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is a density over the IR parameterization, then the density conditional on $(\mathbf{B}, \mathbf{\Sigma})$ with respect to the volume measure over $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ 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 $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ will be uniform if and only if $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is constant over $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$.

Proposition 2. *For every density over the IR parameterization with respect to Lebesgue measure, the density with respect to the volume measure over $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$, 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-*

[11] The joint prior distribution over the identified set for the vector of impulse responses (or equivalently the conditional joint prior distribution for the vector of impulse responses) is obtained conditioning on the reduced-form parameters.

[12] See the discussion of Theorem 2 in Arias, Rubio-Ramírez, and Waggoner (2018) for an outline of how the volume measure is defined over submanifolds. To see that $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ is a submanifold, let $(\mathbf{L}_0, \mathbf{L}_+)$ be any IR parameters such that $\phi(\mathbf{L}_0, \mathbf{L}_+) = (\mathbf{B}, \mathbf{\Sigma})$. Since $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma}) = \{(\mathbf{L}_0 \mathbf{Q}, \mathbf{L}_+ \mathbf{Q}) | \mathbf{Q} \in \mathcal{O}(n)\}$ and $\mathcal{O}(n)$ is a smooth manifold, so is $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$.

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 B. □

Proposition 2 was stated in terms of conditional distributions. There is an equivalent formulation in terms of observationally equivalent parameters.

Proposition 3. *For every density over the IR parameterization with respect to Lebesgue measure, the density with respect to the volume measure over* $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ *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.*

Proof. Follows from Proposition 2 and the fact that two impulse responses are observationally equivalent if and only if there exists a reduced-form parameter $(\mathbf{B}, \mathbf{\Sigma})$ 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, Propositions 2 and 3 bring to the fore the virtue of joint distributions over the IR parameterization that induce a distribution over the orthogonal reduced-form parameterization such that the distribution over the set of orthogonal matrices is uniform.[13] 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 nonuniform 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. The results in this section are relevant for the robust methodology developed by Giacomini and Kitagawa (2021). First, one should acknowledge that 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 researchers can claim that the identification problem is only resolved by means of sign and zero restrictions, preserving the virtues that made inference based on sign identified SVARs a useful tool in empirical macroeconomics. Second, while the analysis in Giacomini and Kitagawa (2021) could potentially be extended

[13] If a distribution over the orthogonal reduced-form parameterization 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.

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.

4.1 The Case $n = 2$

If $n = 2$, then it is possible to analytically illustrate Proposition 2. This example is also useful for illustrating why focusing on marginal distributions can be misleading about the nonuniformity of the priors. To further reduce the number of parameters, we assume 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:

$$\underbrace{\begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \end{bmatrix}}_{\mathbf{L}_0} = \underbrace{\begin{bmatrix} \hat{\ell}_{11} & 0 \\ \hat{\ell}_{21} & \hat{\ell}_{22} \end{bmatrix}}_{\hat{\mathbf{L}}_0} \underbrace{\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ (-1)^i \sin(\theta) & (-1)^{i+1} \cos(\theta) \end{bmatrix}}_{\mathbf{Q}}, \tag{2}$$

where i is either zero or one, $-\pi \leq \theta \leq \pi$, and $\hat{\mathbf{L}}_0 = h(\mathbf{\Sigma})'$.

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:

$$\begin{aligned} p(\ell_{11}) &= \frac{p(\cos^{-1}(\ell_{11}/\hat{\ell}_{11})) + p(-\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))}{\hat{\ell}_{11}\sin(\cos^{-1}(\ell_{11}/\hat{\ell}_{11}))}, \\ p(\ell_{12}) &= \frac{p(\sin^{-1}(\ell_{12}/\hat{\ell}_{11})) + p(\mathrm{sgn}(\ell_{12})\pi - \sin^{-1}(\ell_{12}/\hat{\ell}_{11}))}{\hat{\ell}_{11}\cos(\sin^{-1}(\ell_{12}/\hat{\ell}_{11}))}, \end{aligned} \tag{3}$$

where $\mathrm{sgn}(\ell_{12})$ is one if ℓ_{12} is positive and minus one otherwise. 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 Cases (1) and (2) described below. To economize language, in the rest of the section we minimize the usage of the word *prior*.

Case (1): 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 the distribution of θ must satisfy $p(\theta) + p(-\theta) = \sin(\theta)/2$ for $0 \leq \theta \leq \pi$. If $p(\theta)$ is *any* non-negative function defined over $[-\pi, \pi]$ that integrates to one and satisfies $p(\theta) + p(-\theta) = \sin(\theta)/2$, for $0 \leq \theta \leq \pi$, then the conditional distribution of ℓ_{11} will be uniform.

Is there a choice of $p(\theta)$ so that the conditional distribution of ℓ_{21} will be uniform? If the conditional distribution of ℓ_{12} is uniform, then $p(\ell_{12}) = 1/(2\hat{\ell}_{11})$ and the distribution of θ must satisfy $p(\theta) + p(\text{sgn}(\theta)\pi - \theta) = \cos(\theta)/2$ for $-\pi/2 \leq \theta \leq \pi/2$. Since $p(-\theta) = \sin(\theta)/2 - p(\theta)$ for $0 \leq \theta \leq \pi$, it must be the case that $\cos(\theta)/2 = \cos(-\theta)/2 = p(-\theta) + p(-\pi + \theta) = \sin(\theta)/2 - p(\theta) + \sin(\pi - \theta)/2 - p(\pi - \theta) = \sin(\theta) - \cos(\theta)/2$ for $0 \leq \theta \leq \pi/2$, which is not true. So, there is no choice of $p(\theta)$ such that the conditional distribution of ℓ_{11} and ℓ_{12} are both uniform.

Case (2): The conditional joint distribution over $\mathbf{L}_0$ is uniform for every $\boldsymbol{\Sigma}$. By Proposition 2, the induced distribution over $\mathbf{Q}$ will be uniform with respect to the volume measure, which is arc length in this case. So, the properly scaled density over (θ, i) is $p(\theta, i) = p(\theta)p(i|\theta) = (1/(2\pi))(1/2)$. By Equation (3), the conditional marginal density of ℓ_{11} is $p(\ell_{11}) = \frac{1}{\pi}(\hat{\ell}_{11}^2 - \ell_{11}^2)^{-\frac{1}{2}}$. The density of ℓ_{12} will have the same form.

The two cases are illustrated in Figures 6 and 7. Figure 6 plots the conditional densities of ℓ_{11} and ℓ_{12}, while Figure 7 does the same for the conditional joint distribution over $\mathbf{L}_0$ for both cases. For simplicity, we chose $\boldsymbol{\Sigma}$ to be the identity, so that $\hat{\mathbf{L}}_0$ is also the identity. For Case (1), we chose $p(\theta) = |\sin(\theta)/4|$ 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 the density of ℓ_{12} is $p(\ell_{12}) = |\ell_{12}|(1 - \ell_{12}^2)^{-\frac{1}{2}}/2$.

The solid lines in Figure 6 are the conditional densities in Case (1) and the dotted lines in Figure 6 correspond to Case (2). For Case (1), the conditional distribution of ℓ_{11} is uniform by construction, but the conditional distribution of ℓ_{12} is farther from uniform than in Case (2), where neither conditional distributions are uniform. Figure 6 illustrates the dangers of analyzing marginal densities. While Case (2) implies a uniform joint prior distribution over the identified set for the vector of impulse responses, a researcher analyzing the conditional prior distributions of individual impulse responses may conclude otherwise.

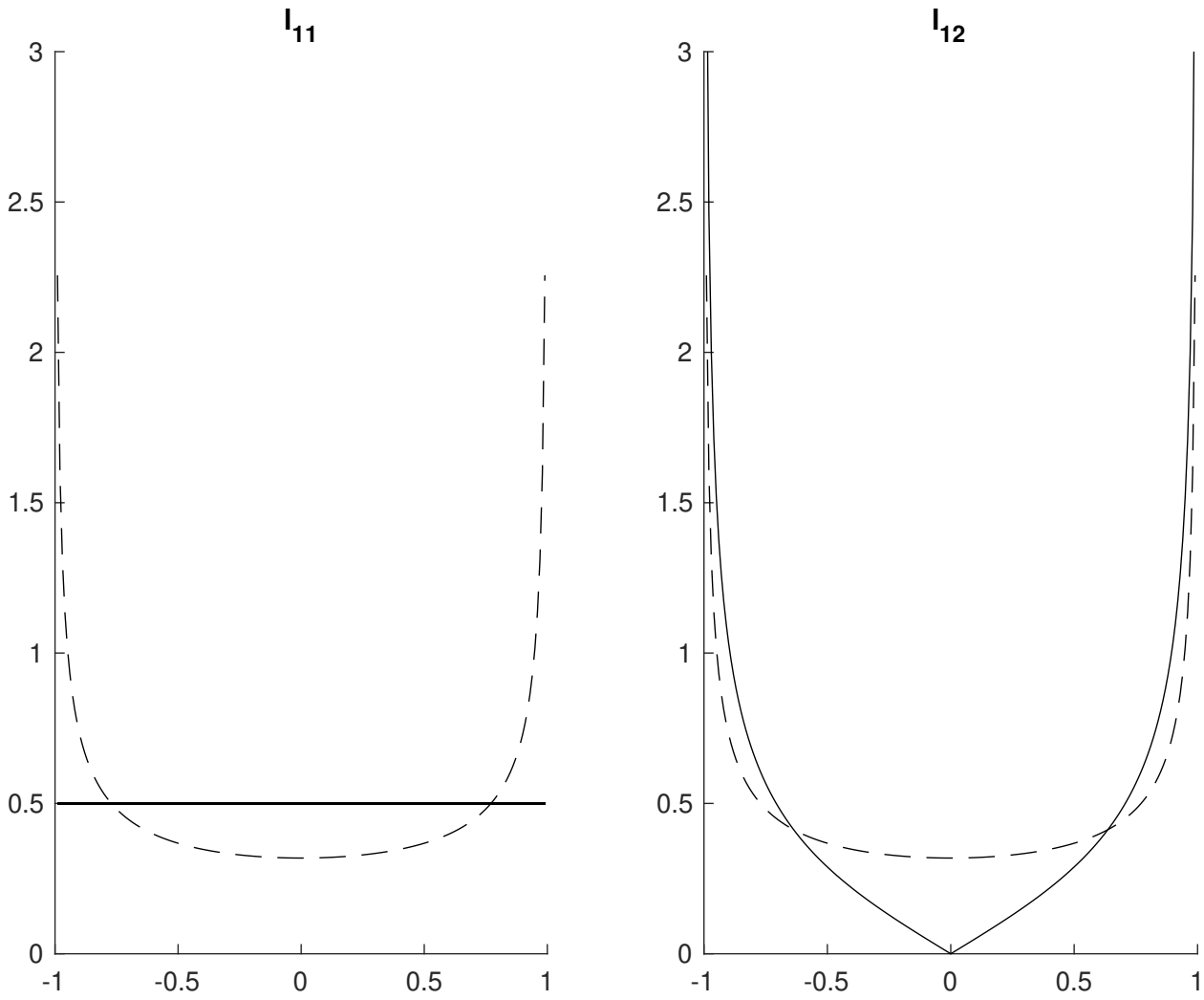


Figure 6: Case (1): The solid lines are the conditional densities of ℓ_{11} and ℓ_{12} with the conditional density of ℓ_{11} forced to be uniform. Case (2): The dotted lines are the conditional densities of ℓ_{11} and ℓ_{12} implied by the uniform conditional joint distribution over $\mathbf{L}_0$.

Case (1) illustrates a point already made by Baumeister and Hamilton (2015): One cannot have uniform prior distributions over the identified sets of individual impulse responses.

Let us now depict the joint distributions. From Equation (2), it is easy to see that the support of the distribution of $\mathbf{L}_0$, conditional on $\mathbf{\Sigma}$, are two ellipses in $\mathbb{R}^4$. If $\hat{\ell}_{21} = 0$, then this reduces to the simpler case of two circles in $\mathbb{R}^4$, both of radius $\sqrt{\hat{\ell}_{11}^2 + \hat{\ell}_{22}^2}$. Since, in the plots we take $\mathbf{\Sigma} = \mathbf{I}_2$, we are in the simpler case. In Figure 7, we plot the conditional joint density over one of the two circles for both Cases (1) and (2).[14] In Case (2), the conditional joint distribution was uniform by construction. In Case (1), this is not true and the density

[14]Each point on the circle corresponds to a tuple $(\ell_{1,1}, \ell_{1,2}, \ell_{2,1}, \ell_{2,2})$ in the support of $\mathbf{L}_0$ conditional on $\mathbf{\Sigma}$ and lies in a plane in $\mathbb{R}^4$.

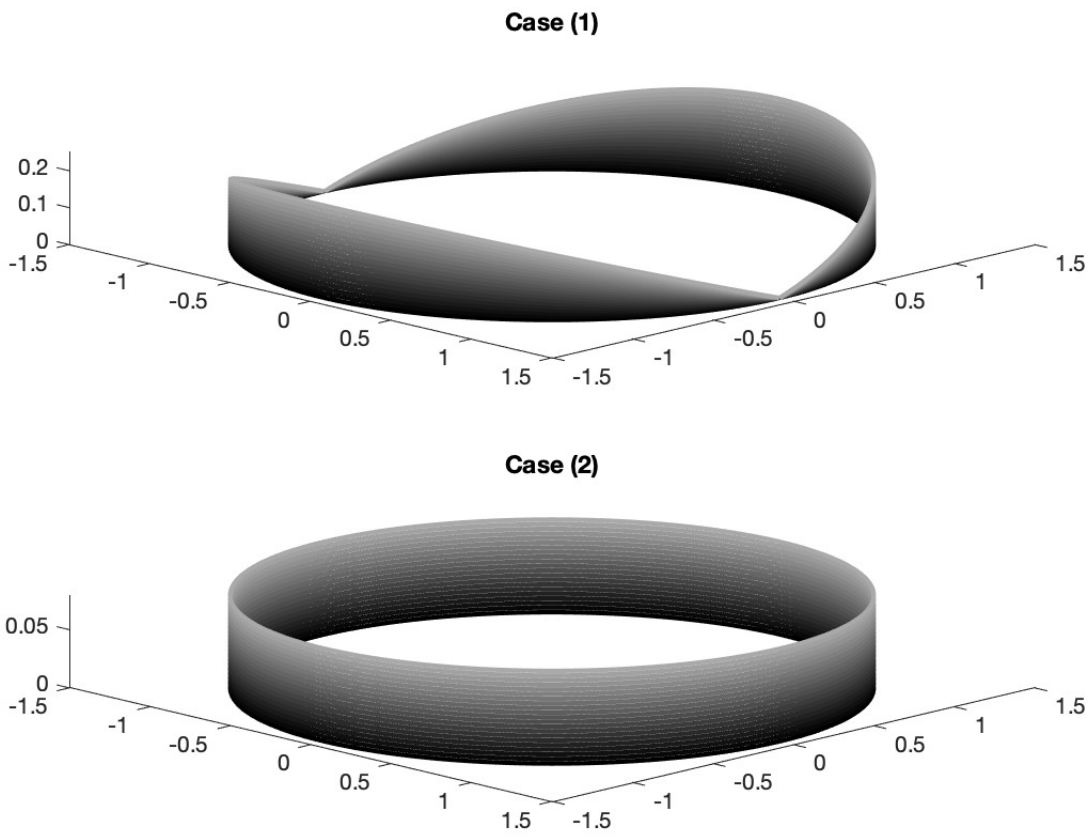


Figure 7: Case (1): Conditional joint density over one of the two circles, with the conditional marginal density of ℓ_{11} forced to be uniform. Case (2): Uniform conditional joint density over one of the two circles implied by a uniform conditional joint prior distribution for the vector of impulse responses.

goes to zero at certain points. This is mirrored in Figure 6 where the conditional marginal distribution of ℓ_{12} also goes to zero in Case (1).

In this section, we have shown that the conventional methods do imply a uniform joint prior distribution over the identified set for the vector of impulse responses. As mentioned, the key is to focus on joint distributions instead of marginals. In the next sections, we demonstrate that the conjecture in Baumeister and Hamilton (2015)—"Because the objects of interest in structural VARs are highly nonlinear functions of the underlying parameters, the quest for 'noninformative' priors for structural VARs is destined to fail" (page 1979)—is not true. We will begin by showing that it is possible to have a uniform joint prior distribution for the vector of impulse responses and that such a type of prior can be implemented by the methods

in Uhlig (2005), Rubio-Ramírez, Waggoner, and Zha (2010), and Arias, Rubio-Ramírez, and Waggoner (2018). Then, we extend the results to a general class of objects of interest.

5 Uniform Joint Prior for Impulse Responses

In this section, we show how to use the conventional methods to conduct posterior inference based on a uniform prior distribution over the IR parameterization conditional on the sign restrictions.[15] To do so, we analytically derive the prior distribution over the orthogonal reduced-form parameterization induced by a uniform prior distribution over the IR parameterization. This step is important because, as highlighted in Section 2, the orthogonal reduced-form parameterization is convenient for obtaining independent and identically distributed draws. Then, we derive a closed form expression for the posterior over the orthogonal reduced-form parameterization induced by a uniform joint prior distribution for the vector of impulse responses. To conclude the section, we propose an algorithm to use the conventional methods with a uniform prior distribution over the IR parameterization and we illustrate it using the empirical example in Watson (2020).

5.1 Prior over the Orthogonal Reduced-Form Parameterization

If $\pi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ is any density over the orthogonal reduced-form parameterization, then by Theorem 1 in Appendix C, the induced density over the IR parameterization will be $\pi(\phi_h(\mathbf{L}_0, \mathbf{L}_+))v_{\phi_h}(\mathbf{L}_0, \mathbf{L}_+)$, where $v_{\phi_h}(\mathbf{L}_0, \mathbf{L}_+)$ is the volume element induced by ϕ_h. It is easy to verify that the hypotheses of Theorem 1 in Appendix C are satisfied and so the theorem is applicable. The volume element can be computed analytically using Proposition 4, described below.

Proposition 4. *The volume element of* ϕ_h *at* $(\mathbf{L}_0, \mathbf{L}_+)$ *is* $v_{\phi_h}(\mathbf{L}_0, \mathbf{L}_+) = 2^{\frac{n(n+1)}{2}} |\det(\mathbf{L}_0)|^{-(m-3)}$.

Proof. See Appendix B. □

The reader should notice that the volume element does not depend on the choice of h and that it is immediate that the volume element can be written in terms of the orthogonal reduced-form parameterization as shown in the following corollary of Proposition 4:

[15] Oftentimes, we will refer to this prior as a uniform joint prior distribution for the vector of impulse responses.

Corollary 1. *The volume element of* ϕ_h^{-1} *evaluated at* $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ *is* $v_{\phi_h^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$.

Using Corollary 1, we have that if $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is any density over the IR parameterization, then the induced density over the orthogonal reduced-form parameterization will be:

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

This last expression justifies the following proposition:

Proposition 5. *A joint prior distribution for the vector of impulse responses is uniform if and only if the equivalent prior density over the orthogonal reduced-form parameterization is proportional to* $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$.

Proof. If $\pi(\mathbf{L}_0, \mathbf{L}_+)$ is any density over the IR parameterization, then the induced density over the orthogonal reduced-form parameterization will be:

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

Since a uniform joint prior distribution for the vector of impulse responses implies that $\pi(\phi_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) \propto 1$, $\pi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) \propto |\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$. □

Proposition 5 shows that if one defines a uniform prior distribution over the IR parameterization, then one is irremediably defining a prior over the orthogonal reduced-form parameterization whose density is proportional to $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$. Importantly, Proposition 5 applies to those cases in which the impulse responses of interest are linear transformations of the IR parameterization. It is also the case that a prior over the orthogonal reduced-form parameterization proportional to $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$ induces a uniform joint prior for any subset of the vector of impulse responses. This is because if the prior distribution over the IR parameterization is uniform, the joint prior over the subset of the vector of impulse responses is also uniform. For example, in Section 3, the vector of objects of interest are the impulse responses of output, inflation, the long rate, and the real rate to a demand shock, a monetary policy shock, and a supply shock, respectively. These objects are a linear transformation of a subset of the vector of impulse responses.

The following corollary of Proposition 5 shows three additional straightforward consequences of working with a uniform prior distribution over the IR parameterization. First,

the induced posterior over the orthogonal reduced-form parameterization is such that $(\mathbf{B}, \mathbf{\Sigma})$ and $\mathbf{Q}$ are independent. Second, the induced prior density over the reduced-form parameters takes a particular form. Third, the induced prior distribution of $\mathbf{Q}$ is uniform.

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

Hence, because of the results in Section 4, a uniform prior distribution over the IR parameterization implies a uniform joint prior and posterior distributions over the identified set for the vector of impulse responses.

5.2 Posterior over the Orthogonal Reduced-Form Parameterization

The following proposition due to DeJong (1992) shows that a prior over the reduced-form parameters proportional to $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$ implies a normal-inverse-Wishart posterior over the reduced-form parameters.

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 normal-inverse-Wishart 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 $\hat{\nu}(a) = T + a - m - n - 1$.

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

Corollary 3. *If the prior density over the orthogonal reduced-form parameterization is proportional to* $|\det(\mathbf{\Sigma})|^{\frac{m-3}{2}}$, *the posterior density over the orthogonal reduced-form parameterization is:*

$$UNIW_{\left(\hat{\nu}(-(m-3)), \hat{\mathbf{S}}, \hat{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}\right)}(\mathbf{B}, \mathbf{\Sigma}).$$

Corollary 3 implies that if one wants to conduct inference based on a uniform prior distribution over the IR parameterization, then one must have a particular (model dependent)

uniform normal-inverse-Wishart posterior over the orthogonal reduced-form parameterization. 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 $\mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}$.

5.3 The Algorithm

The above discussion justifies Algorithm 2 to independently draw from the posterior distribution over the IR parameterization conditional on the sign restrictions implied by a uniform prior distribution over the IR parameterization.

Algorithm 2. *The following algorithm independently draws from the joint posterior distribution for the vector of impulse responses conditional on the sign restrictions implied by a uniform joint prior distribution for the vector of impulse responses.*

1. *Draw* $(\mathbf{B}, \mathbf{\Sigma})$ *independently from the NIW* $\left(\hat{\nu}(-(m-3)), \hat{\mathbf{S}}, \hat{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}\right)$ *distribution.*
2. *Draw* $\mathbf{Q}$ *independently from the uniform distribution over* $\mathcal{O}(n)$ *using Proposition 1.*
3. *Keep* $(\mathbf{L}_0, \mathbf{L}_+) = \phi_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ *if the sign restrictions are satisfied.*
4. *Return to Step 1 until the required number of draws has been obtained.*

We see this algorithm as complementing Plagborg-Møller (2019). While his approach cannot produce independent draws, it does not require invertibility.

5.4 An Application

We illustrate how to conduct inference based on a uniform joint prior distribution for the vector of impulse responses using the model described in Section 3. For completeness, we will begin the analysis comparing the unconditional posterior distributions of individual impulse responses implied by the uniform prior distribution over the IR parameterization with the unconditional posterior distributions of individual impulse responses implied by the prior distribution over the IR parameterization induced by the more commonly used Minnesota prior discussed in Section 3. Figure 8 shows the equal-tailed 68 percent conditional posterior probability intervals of individual impulse responses implied by each of the priors. As expected, the figure shows how the uniform joint prior distribution for the vector of

impulse responses implies more posterior uncertainty. In some cases, e.g., the responses of the real rate, the uncertainty (measured as the width of the intervals) increases noticeably.

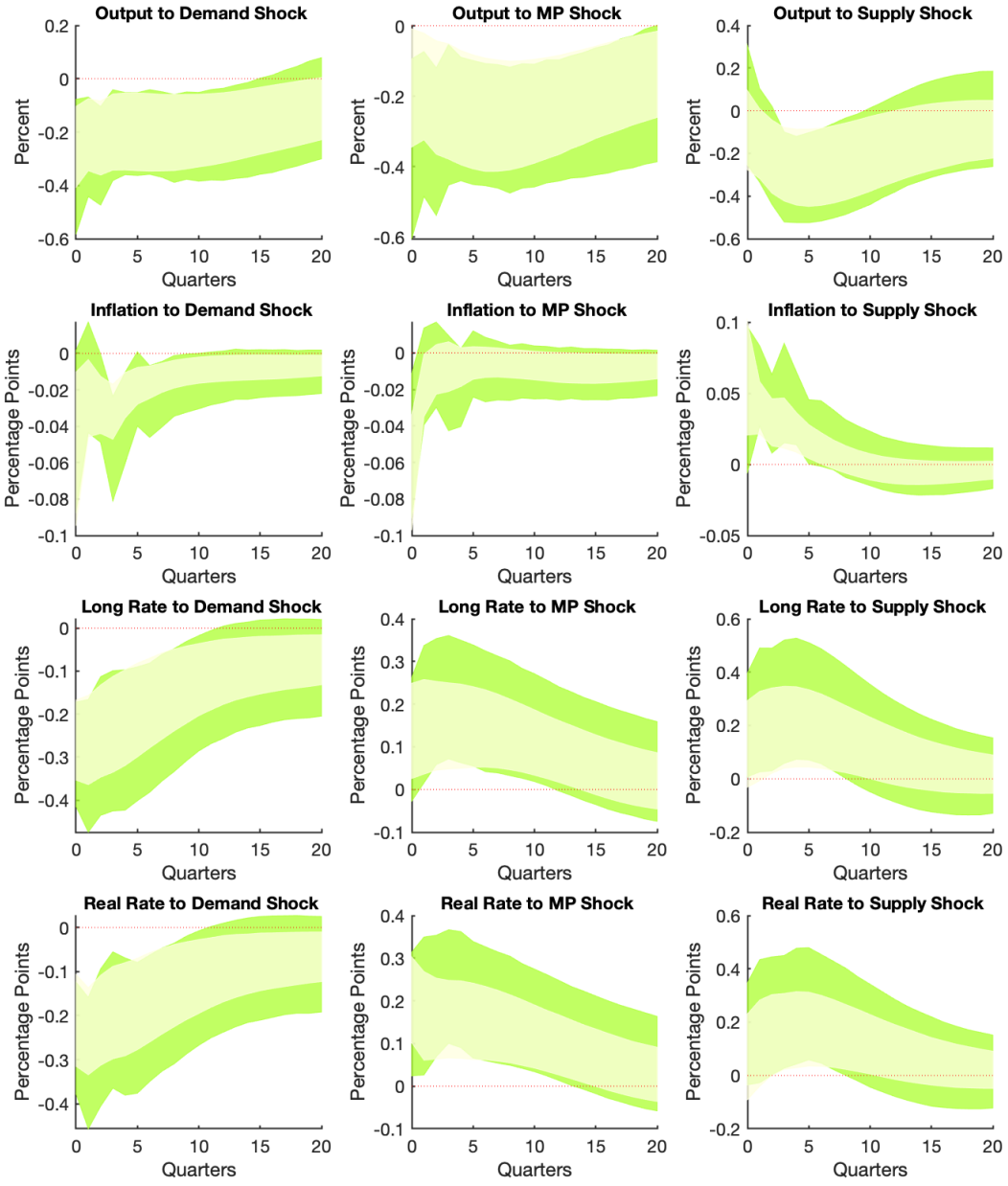


Figure 8: The dark (light) green areas show the equal-tailed 68 percent marginal posterior probability bands of individual impulse responses implied by the uniform joint prior distribution for the vector of impulse responses (Minnesota prior).

Next, we compare marginal and joint inference. The former is based on the unconditional

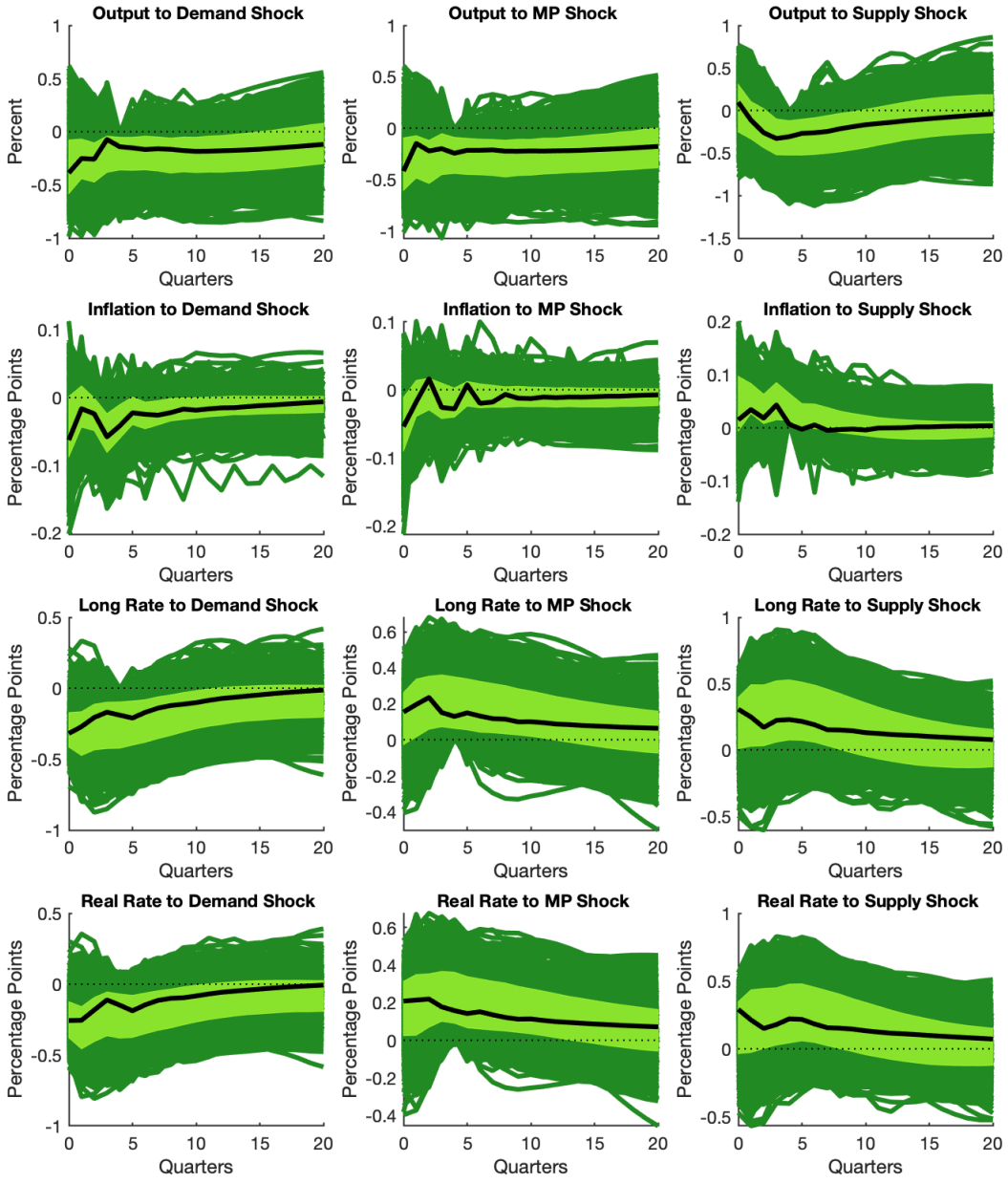


Figure 9: Bayes estimator of the joint posterior distribution for the vector of impulse responses (black lines) and its 68 percent credible set (dark green lines) under the absolute value loss function. Light green areas show the equal-tailed 68 percent unconditional prior distributions of individual impulse responses. Both posteriors are implied by the uniform joint prior distribution for the vector of impulse responses.

posterior distributions of individual impulse responses. The latter is based on the joint posterior distribution for the vector of impulse responses. Figure 9 compares the Bayes

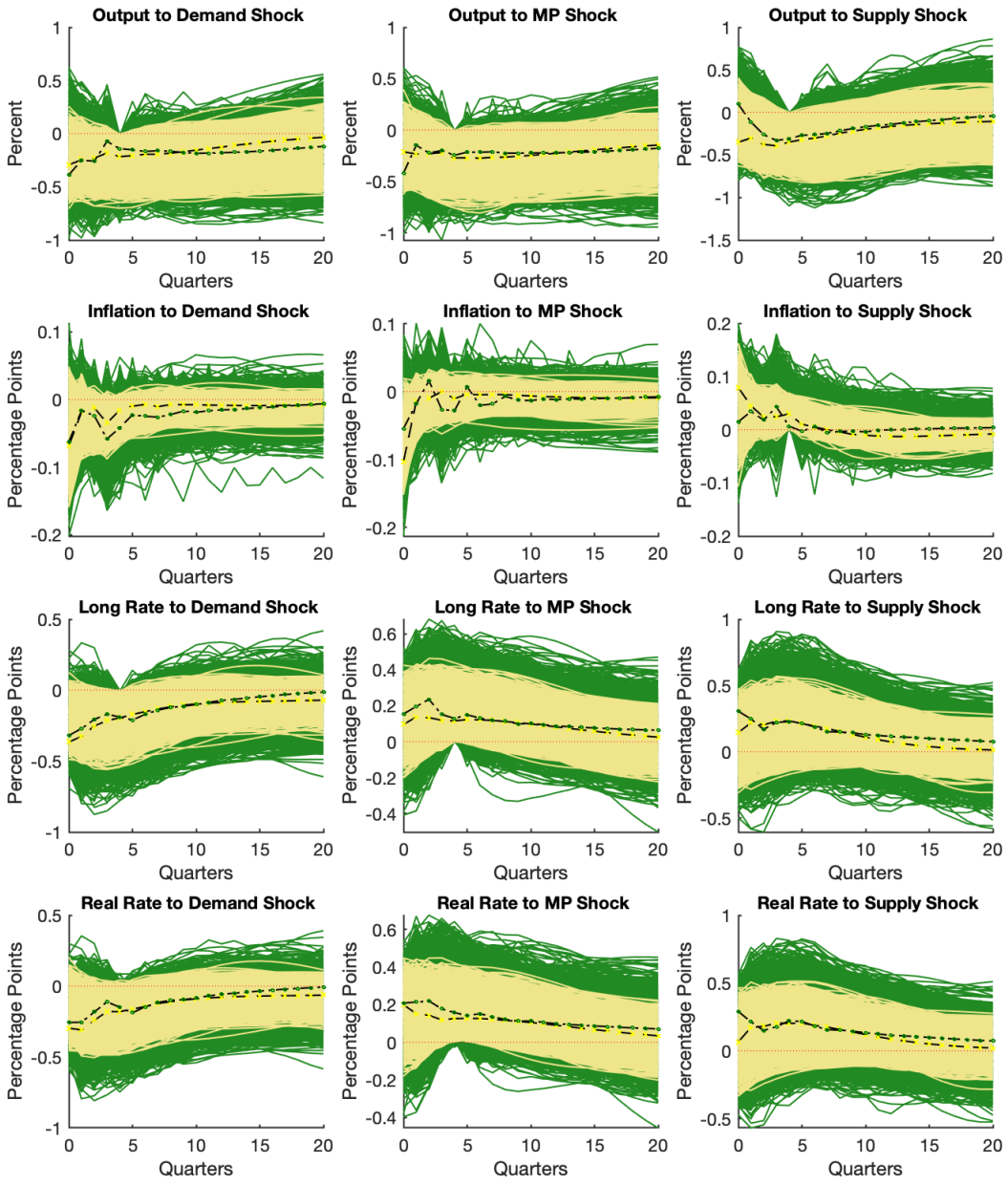


Figure 10: Bayes estimator of joint posterior impulse responses (dashed-dotted black lines with green round markers) and the 68 percent credible sets under the absolute value loss function under uniform joint prior distribution for the vector of impulse responses (dashed-dotted black lines with green round markers for the estimator and green lines for the credible sets) and under the Minnesota prior (dashed black lines with yellow cross markers for the estimator and yellow lines for the credible sets).

estimator of the joint posterior distribution for the vector of impulse responses (black lines) and its 68 percent credible set (dark green lines) under the absolute value loss function

following Inoue and Kilian (2022a) with the commonly used 68 percent point-wise posterior probability bands (light green areas), which are normally used to report unconditional posterior distributions of individual impulse responses. The figure makes clear that: (1) joint inference protects the researcher from excessively optimistic conclusions about the uncertainty surrounding the posteriors, as Inoue and Kilian (2022a) highlight; and (2) this particular model does not seem tightly identified by the imposed sign and zero restrictions. These conclusions are robust to using the sup-t Bayesian joint credible sets proposed by Montiel Olea and Plagborg-Møller (2019), see Appendix D.

We conclude this section by comparing the joint posterior distribution for the vector of impulse responses implied by two different joint priors. Figure 10 shows the Bayes estimators of joint posterior impulse responses and the 68 percent credible sets under the absolute value loss function when using a uniform joint prior distribution for the vector of impulse responses (dashed-dotted black lines with green round markers for the estimator and green lines for the credible sets) and when using the joint prior distribution for the vector of impulse responses induced by the Minnesota prior (dashed black lines with yellow cross markers for the estimator and yellow lines for the credible sets). Focusing on the Bayes estimators, the Minnesota prior exacerbates the negative effects on inflation caused by negative demand shocks and dampens the effects on inflation caused by unfavorable supply shocks. Although the 68 percent credible sets are much narrower when using the Minnesota prior, a visual inspection reveals that even when using a Minnesota prior 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 shown in Appendix D. As mentioned above, this is clearly in line with the conclusions in Inoue and Kilian (2022a).

6 Uniform Joint Priors for Objects of Interest

There are cases in which the vector of objects of interest is not the vector of impulse responses. In this section, we show that the insights of the previous sections apply to a general class of objects of interest parameterizations, which are transformations of the orthogonal reduced-form parameterization.[16] Denote the vector of objects of interest by $\boldsymbol{\Upsilon}$

[16] Alternatively, one could define the objects of interest parameterization as being a transformation of either the structural parameterization or the IR parameterization.

and the transformation from $\boldsymbol{\Upsilon}$ to the orthogonal reduced-form parameterization by ϕ_o. We assume that ϕ_o is invertible and that both ϕ_o and its inverse are continuously differentiable. In general, $\boldsymbol{\Upsilon}$ is a smooth manifold of dimension $n^2 + nm$, but here we focus on the case in which $\boldsymbol{\Upsilon}$ is an open subset of $\mathbb{R}^{n^2+nm}$ and use Lebesgue measure over $\boldsymbol{\Upsilon}$. As in the case of the IR parameterization, statements about the marginal distribution of individual objects of interest directly translate into statements about the marginal distribution of individual parameters in the objects of interest parameterization. Similarly, statements about the joint distribution of the vector of objects of interest also directly translate into statements about the distribution over the objects of interest parameterization.

If $\pi(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ is any density over the orthogonal reduced-form parameterization, the induced density over the objects of interest parameterization is $\pi(\phi_o(\boldsymbol{\Upsilon}))v_{\phi_o}(\boldsymbol{\Upsilon})$, where $v_{\phi_o}(\boldsymbol{\Upsilon})$ is the volume element induced by ϕ_o.[17] If $\pi(\boldsymbol{\Upsilon})$ is any density over the objects of interest parameterization, then the induced density over the orthogonal reduced-form parameterization will be $\pi(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \pi(\phi_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, where $v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ is the volume element of ϕ_o^{-1} evaluated at $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. Hence, a uniform joint prior distribution for the vector of objects of interest induces a prior over the orthogonal reduced-form parameterization such that $\pi(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) \propto v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. Because, in general, the volume element $v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ depends on $\mathbf{Q}$, the induced prior over the set of orthogonal matrices is not uniform.

To relate this to the previous sections, where the objects of interest are the vector of impulse responses, the mapping ϕ_o corresponds to ϕ_h and the volume element $v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ corresponds to the volume element $v_{\phi_h^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, which can be explicitly computed to be $2^{-\frac{n(n+1)}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{m-3}{2}}$, and hence, it does not depend on $\mathbf{Q}$. In general, it may not be possible to analytically compute the volume element and it may be the case that the volume element depends on $\mathbf{Q}$.

6.1 Conditional Joint Prior for Objects of Interest

As mentioned in Section 4 for the case of the IR parameterization, because the posterior reproduces the prior over the identified set, a researcher may want a uniform joint prior distribution over the identified set for the vector of objects of interest. Oftentimes, we will

[17] As in Section 5, it is easy to verify that the hypotheses of Theorem 1 in Appendix C are satisfied and so the theorem is applicable.

refer to this prior as the conditional joint prior distribution for the vector of objects of interest.[18] In this section, we characterize the class of priors over the orthogonal reduced-form parameterization that induces a uniform joint prior distribution over the identified set for the vector of objects of interest. As expected, the uniform joint prior distribution for the vector of objects of interest belongs to such a class.

The composite $\hat{\phi}_o = f_p \circ \phi_o$ maps the objects of interest parameterization to the reduced-form parameters. Given the reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$, the set $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$ will be the submanifold that is the support of the joint distribution of the object of interest parameterization conditional on $(\mathbf{B}, \mathbf{\Sigma})$. As noted in the case that the objects of interest were the impulse responses, the submanifold structure induces a natural measure over $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$, which is called the volume measure (as before, see Arias, Rubio-Ramírez, and Waggoner, 2018, for details.) If $\pi(\mathbf{\Upsilon})$ is a density over the object of interest parameterization, then the density conditional on $(\mathbf{B}, \mathbf{\Sigma})$ with respect to the volume measure over $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$ will be proportional to $\pi(\mathbf{\Upsilon})$. The volume measure is the only measure, up to a scale factor, that has this property. Thus, conditional on $(\mathbf{B}, \mathbf{\Sigma})$, the density with respect to the volume measure over $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$ will be uniform if and only if $\pi(\mathbf{\Upsilon})$ is constant over $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$.

Proposition 7. *For every density over the objects of interest parameterization with respect to Lebesgue measure, the density with respect to the volume measure over $\hat{\phi}_o^{-1}(\mathbf{B}, \mathbf{\Sigma})$, 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 parameterization is such that $\pi(\mathbf{Q} \,|\, \mathbf{B}, \mathbf{\Sigma})$ is proportional to $v_{\phi_o^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.*

Proof. See Appendix B. □

An immediate implication of Proposition 7 is that a uniform joint prior distribution for the vector of objects of interest implies a uniform joint prior and posterior distributions over the identified set for the vector of objects of interest. It should be clear that Proposition 7 is a generalization of Proposition 2 for our class of objects of interest. Proposition 2 just shows that for the case when the objects of interest are the impulse responses, $\pi(\mathbf{Q} \,|\, \mathbf{B}, \mathbf{\Sigma})$ is uniform and independent of $(\mathbf{B}, \mathbf{\Sigma})$.

[18] 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 conditioning on the reduced-form parameters.

6.2 The Algorithm

The above discussion justifies using Algorithm 3 to independently draw from the posterior distribution over the objects of interest parameterization conditional on the sign restrictions for inference based on a uniform prior distribution over the objects of interest parameterization. The algorithm is a simple modification of Algorithms 1 and 2 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 over the objects of interest parameterization is uniform, then the posterior density at $\mathbf{\Upsilon} = \phi_o^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ will also be proportional to $NIW_{(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}})}(\mathbf{B},\mathbf{\Sigma})$.

Algorithm 3. *The following algorithm independently draws from the posterior distribution over the objects of interest parameterization conditional on the sign restrictions implied by a uniform prior distribution over the objects of interest parameterization.*

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* $\mathcal{O}(n)$ *using Proposition 1.*

3. *If* $\mathbf{\Upsilon} = \phi_o^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ *satisfies the sign restrictions, then set its importance weight to:*

$$\frac{NIW_{(\hat{\nu},\hat{\mathbf{\Phi}},\hat{\mathbf{\Psi}},\hat{\mathbf{\Omega}})}(\mathbf{B},\mathbf{\Sigma})}{NIW_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{B},\mathbf{\Sigma})v_{\phi_o}^{-1}(\mathbf{\Upsilon})}.$$

   Otherwise, set its importance weight to zero.

4. *Return to Step 1 until the required number of draws has been obtained.*

The choice of $(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})$ is important. 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 3 to work with any joint posterior distribution for the vector of objects of interest provided that Step 3 is modified accordingly.

6.3 An Example

To illustrate Algorithm 3, consider a simplified version of the two-variable SVAR described in Baumeister and Hamilton (2015). In particular, let

$$\Delta n_t = k^d + \beta^d \Delta w_t + b_w^d \Delta w_{t-1} + b_n^d \Delta n_{t-1} + \sigma^d u_t^d, \tag{4}$$

$$\Delta n_t = k^s + \alpha^s \Delta w_t + b_w^s \Delta w_{t-1} + b_n^s \Delta n_{t-1} + \sigma^s u_t^s, \tag{5}$$

where the vector $(u_t^d, u_t^s)'$, conditional on past information and the initial conditions, is Gaussian with mean zero and covariance matrix $\boldsymbol{I}_2$. Letting $\mathbf{y}_t$ denote the endogenous variables (i.e., $\mathbf{y}_t = (\Delta w_t, \Delta n_t)'$), it should be clear that Baumeister and Hamilton's (2015) $(\mathbf{A}, \mathbf{D}, \mathbf{\Pi})$ parameterization of Equations (4) and (5) is:

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \mathbf{u}_t, \tag{6}$$

where $\mathbf{u}_t = (u_t^d, u_t^s)'$, $\mathbf{x}_{t-1} = (\mathbf{y}_{t-1}, 1)'$, and:

$$\mathbf{A} = \begin{bmatrix} -\beta^d & 1 \\ -\alpha^s & 1 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} \sigma^d & 0 \\ 0 & \sigma^s \end{bmatrix}, \text{and } \mathbf{\Pi}' = \begin{bmatrix} b_w^d & b_n^d & k^d \\ b_w^s & b_n^s & k^s \end{bmatrix}.$$

Our version of Baumeister and Hamilton's (2015) two-variable SVAR features one lag and a constant, and we assume that the objects of interest are the short-run wage elasticity of demand, β^d, the short-run wage elasticity of supply, α^s, the standard deviation of the structural demand and supply shocks, and the lag structural coefficients plus the constants $(\sigma^d, \sigma^s, b_w^d, b_n^d, k^d, b_w^s, b_n^s, k^s)'$. For compactness, we let $\mathbf{\Upsilon} = (\beta^d, \alpha^s, \sigma^d, \sigma^s, b_w^d, b_n^d, k^d, b_w^s, b_n^s, k^s)'$ denote the vector of objects of interest. Given $\mathbf{\Upsilon}$, we will construct a mapping ϕ_o from $\mathbf{\Upsilon}$ to the orthogonal reduced-form parameters, as the composite mapping $\phi_o = f_h \circ f_o$, where f_h was defined in Section 2 and f_o is a mapping from $\mathbf{\Upsilon}$ to the structural parameters defined below as follows:

$$f_o(\mathbf{\Upsilon}) = \left(\underbrace{\mathbf{A}'\boldsymbol{D}^{-1}}_{\mathbf{A}_0}, \underbrace{\mathbf{\Pi}\boldsymbol{D}^{-1}}_{\mathbf{A}_+} \right).$$

The inverse of the f_o mapping is given by:

$$f_o^{-1}(\mathbf{A}_0, \mathbf{A}_+) = (-\mathbf{A}(1,1), -\mathbf{A}(2,1), \text{diag}(\mathbf{D}), \text{vec}(\mathbf{\Pi}))\,,$$

where:

$$\mathbf{D} = \begin{bmatrix} \mathbf{A}_0^{-1}(2,1) & 0 \\ 0 & \mathbf{A}_0^{-1}(2,2) \end{bmatrix}, \mathbf{A} = \mathbf{D}\mathbf{A}_0', \text{and } \mathbf{\Pi} = \mathbf{A}_+ \mathbf{D}.$$

We use Algorithm 3 to obtain draws from the posterior implied by a uniform joint prior distribution for the vector of objects of interest. When applying Algorithm 3, we set $(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega}) = \left(\hat{\nu}, \hat{\mathbf{\Phi}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Omega}}\right)$.

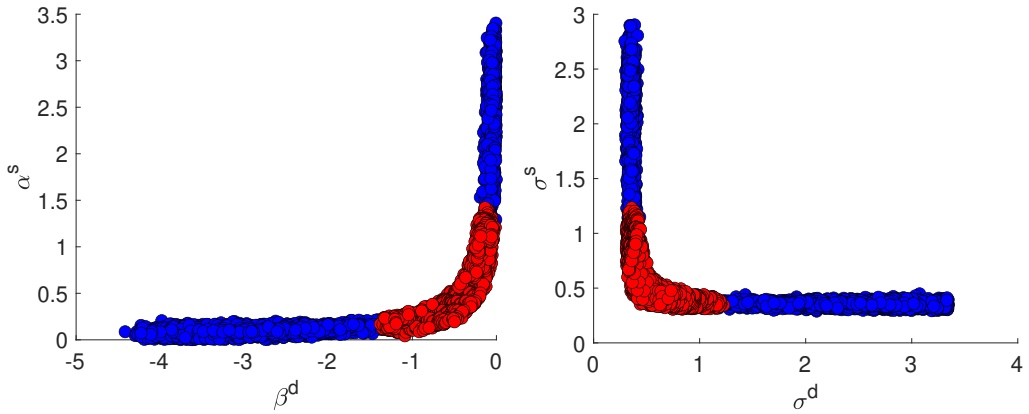
Finally, following Baumeister and Hamilton (2015), we impose the following sign restrictions: $\beta^d < 0$ and $\alpha^s > 0$.[19] We will compare the results to ones obtained with the posterior distribution over the objects of interest parameterization conditional on the sign restrictions associated with a conjugate uniform-normal-inverse-Wishart prior distribution over the orthogonal reduced-form parameterization where the normal-inverse-Wishart part of the prior is a standard Minnesota. To draw from this posterior, we use Algorithm 1 with the posterior distribution implied by the class of priors described in Section 3 where, following Giannone, Lenza and Primiceri's (2015) approach, we set $\bar{\nu} = 4$, $\lambda = 0.3055$, and $\bar{\mathbf{\Phi}} = \text{diag}(2.1232, 0.0598)$.[20]

Importantly, the aim of this section is not to argue that using a uniform joint prior distribution for the vector of objects of interest is preferred to using other priors. The results discussed below are meant to (1) emphasize that it is possible to conduct inference about joint posterior distribution for the vector of objects of interest implied by uniform prior distribution over the objects of interest parameterization and (2) highlight any difference with respect to a more typical choice of priors. As mentioned above, the algorithm can be used to work with any joint posterior distribution for the vector of objects of interest. In particular, one could work with the posterior described in Baumeister and Hamilton (2015).

Panel (a) of Figure 11 compares the 68 percent posterior joint credible sets for β^d and α^s

[19] In addition, we impose normalization on the standard deviation of the shocks (σ^d and σ^s must be positive) and a bound on their size (σ^d and σ^s must be smaller than 4 times the standard deviation of the more volatile time series in the system) to increase the efficiency of Algorithm 3. Without the bounds on the size of the shocks, the effective number of draws in Algorithm 3 collapses to one.

[20] Obviously, when executing Step 3 of the algorithm, we replace $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ with $\mathbf{\Upsilon} = \phi_o^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.

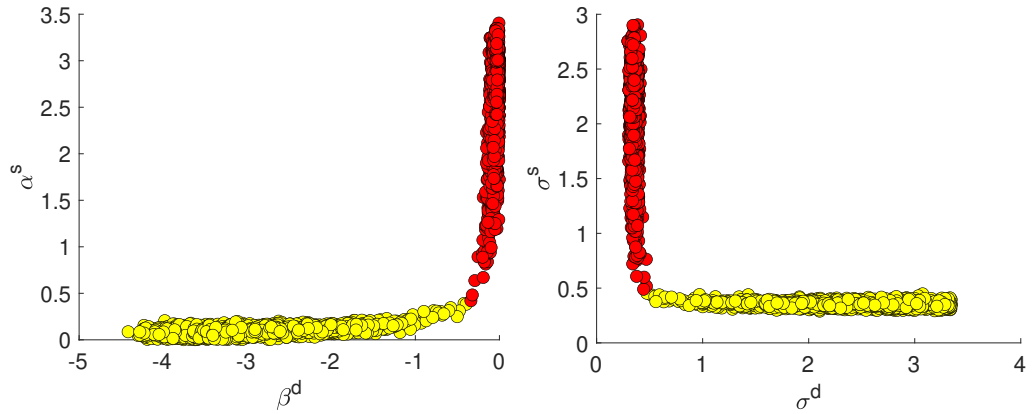


(a) Elasticities (b) Shock Sizes

Figure 11: Posterior distributions implied by a uniform joint prior distribution for the vector of objects of interest (blue) versus Minnesota prior (red). The 68 percent credible sets under the absolute value loss function using a uniform joint prior distribution for the vector of objects of interest and a Minnesota prior over the reduced-form parameters.

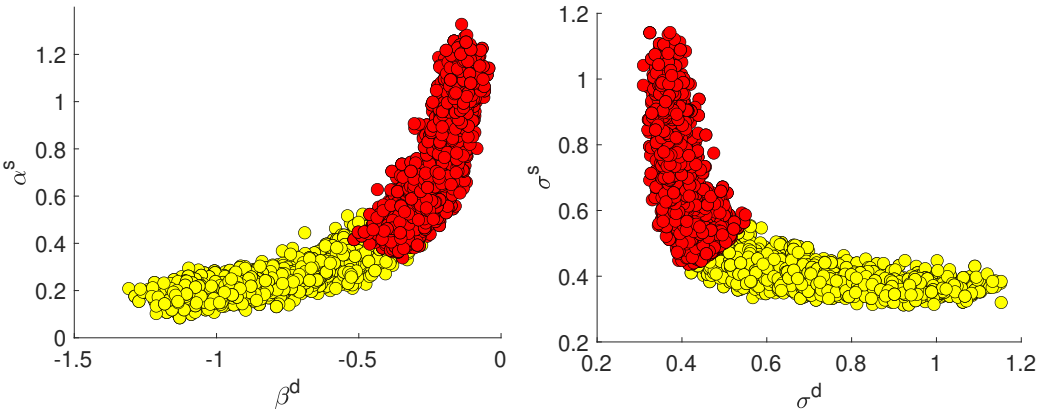
obtained using a uniform joint prior distribution for the vector of objects of interest (blue) with the ones obtained using the Minnesota prior described above (red). As the reader can see, for both priors the posterior is concentrated around low (absolute) values of either β^d or α^s. It is also clear from the figure that the uniform joint prior distribution for the vector of objects of interest implies much more uncertainty about the estimates. Panel (b) makes the same comparison for σ^d and σ^s obtaining similar results. Consequently, researchers using a uniform joint prior distribution for the vector of objects of interest are likely to face wider posterior joint credible sets.

To highlight one of the key advantages of working with joint credible sets, Figure 12 relies on colors (as suggested by Inoue and Kilian, 2022a) to show the relation between the posterior estimates of elasticities and standard deviations when using a uniform joint prior distribution for the vector of objects of interest. Figure 13 shows the same relation when using the Minnesota prior. Clearly, in both cases, large values for the standard deviation of the demand shock σ^d are associated with large values for the demand elasticity β^d. A similar conclusion emerges when assessing the relation between the supply elasticity and the standard deviation of the supply shock.



(a) Elasticities (b) Shock Sizes

Figure 12: The 68 percent credible sets under the absolute value loss function using a uniform joint prior distribution for the vector of objects of interest.



(a) Elasticities (b) Shock Sizes

Figure 13: The 68 percent credible sets under the absolute value loss function using a Minnesota prior.

7 Conclusion

Our paper demonstrates that there is nothing fundamentally wrong with the conventional method for Bayesian inference in SVARs identified with sign restrictions. Using an empirical example, we show that the concerns about the role played by the uniform prior over the set of orthogonal matrices in shaping posterior inference over impulse responses are overstated by ignoring reduced-form parameter uncertainty and focusing on marginal distributions. 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 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 offering a practical complementary alternative to 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 over the IR parameterization 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 over the objects of interest parameterization 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.

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*(2), 685–720.

Baumeister, C. and J. D. Hamilton (2015). Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica 83*(5), 1963–1999.

Baumeister, C. and J. D. Hamilton (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review 109*(5), 1873–1910.

Baumeister, C. and J. D. Hamilton (2022). Advances in Using Vector Autoregressions to Estimate Structural Magnitudes. *Working Paper*.

Bruder, S. and M. Wolf (2018). Balanced Bootstrap Joint Confidence Bands for Structural Impulse Response Functions. *Journal of Time Series Analysis 39*(5), 641–664.

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.

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*(3), 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*(4), 938–60.

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.

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

Inoue, A. and L. Kilian (2016). Joint Confidence Sets for Structural Impulse Responses. *Journal of Econometrics 192*(2), 421–432.

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

Inoue, A. and L. Kilian (2022a). Joint Bayesian Inference about Impulse Responses in VAR Models. *Journal of Econometrics, Forthcoming*.

Inoue, A. and L. Kilian (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*(3), 782–798.

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015b). Confidence Bands for Impulse Responses: Bonferroni vs. Wald. *Oxford Bulletin of Economics and Statistics 77*(6), 800–821.

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2018). Calculating Joint Confidence Bands for Impulse Response Functions Using Highest Density Regions. *Empirical Economics 55*(4), 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), 1–17.

Plagborg-Møller, M. (2019). Bayesian Inference on Structural Impulse Response Functions. *Quantitative Economics 10*(1), 145–184.

Rubio-Ramírez, J., D. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.

Shin, M. and M. Zhong (2020). A New Approach to Identifying the Real Effects of Uncertainty Shocks. *Journal of Business & Economic Statistics 38*(2), 367–379.

Sims, C. A. and T. Zha (1999). Error Bands for Impulse Responses. *Econometrica 67*(5), 1113–1155.

Stewart, G. (1980). The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators. *SIAM Journal on Numerical Analysis 17*(3), 403–409.

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.

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*(4), 1–32.

Appendix Not for Publication

A 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:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & \pm 1 \end{bmatrix} \tag{7}$$

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 $\mathbf{P}$
4. and finally multiplying by the Householder matrix $\mathbf{H}_n(\mathbf{x}(\mathbf{B},\mathbf{\Sigma}) - \mathbf{e}_4)$

is 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 parameterization 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 that 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 Propositions 2 and 3 directly apply to the IR parameterization identified with sign restrictions. It can be shown that they also apply to the model in Watson (2020) with an alternative IR parameterization defined as $(\mathbf{L}_0, \mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \mathbf{L}_\infty, \mathbf{c})$. The mapping from this alternative IR parameterization to the structural parameterization is one-to-one and onto, although we do have to restrict the structural parameters so that $\mathbf{L}_\infty$ is well defined. Using this alternative IR parameterization the zero restrictions define a lower dimensional linear subspace where the volume measure is Lebesgue.

B Proofs of Proposition 2, 4, and 7

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

$$p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \frac{\pi((f_h \circ f_{ir})^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}{2^{\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-\frac{np+2}{2}}}.$$

Because $\phi = f_p \circ (f_h \circ f_{ir})$, π is constant over $\phi^{-1}(\mathbf{B}, \mathbf{\Sigma})$ if and only if $p(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = p(\mathbf{B}, \mathbf{\Sigma}) p(\mathbf{Q} | \mathbf{B}, \mathbf{\Sigma})$ does not depend on $\mathbf{Q}$. If $p(\mathbf{Q} | \mathbf{B}, \mathbf{\Sigma})$ 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 4. Proposition 1 in Arias, Rubio-Ramírez, and Waggoner (2018) gives that $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+) = 2^{\frac{n(n+1)}{2}} |\det(\mathbf{A}_0)|^{-(2n+m+1)}$. Because $\mathbf{A}_k$ does not depend on $\mathbf{L}_j$ for $j > k$, the derivative of f_{ir} is a block lower triangular $(n^2(p+1)+1) \times (n^2(p+1)+1)$ matrix of the

form:

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

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_{ir} is the absolute value of the determinant of the above matrix, which is $|\det(\mathbf{L}_0)|^{-2n(p+1)}$. Thus the volume element of $\phi_h = f_h \circ f_{ir}$ is $2^{\frac{n(n+1)}{2}} \left|\det\left(\left(\mathbf{L}_0^{-1}\right)'\right)\right|^{-(2n+m+1)} |\det(\mathbf{L}_0)|^{-2n(p+1)} = 2^{\frac{n(n+1)}{2}} |\det(\mathbf{L}_0)|^{-(m-3)}$. □

Proof of Proposition 7. If $\pi(\boldsymbol{\Upsilon})$ is any density over the objects of interest parameterization with respect to the Lebesgue measure, then the induced density over the orthogonal reduced-form parameterization with respect to volume measure will be:

$$\pi(\mathbf{B}, \boldsymbol{\Sigma})\pi(\mathbf{Q} \,|\, \mathbf{B}, \boldsymbol{\Sigma}) = \pi(\phi_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}).$$

If $\pi(\boldsymbol{\Upsilon})$ is constant over $\hat{\phi}_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma})$, then $\pi(\phi_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))$ will not depend on $\mathbf{Q}$ and it must be the case that $\pi(\mathbf{Q} \,|\, \mathbf{B}, \boldsymbol{\Sigma})$ is proportional to $v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, though the proportionality constant, which is equal to $\pi(\phi_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))/\pi(\mathbf{B}, \boldsymbol{\Sigma})$, could depend on $\mathbf{B}$ and $\boldsymbol{\Sigma}$. If $\pi(\mathbf{Q} \,|\, \mathbf{B}, \boldsymbol{\Sigma})$ is proportional to $v_{\phi_o^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, then $\pi(\phi_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))$ cannot depend on $\mathbf{Q}$ and so is constant over $\hat{\phi}_o^{-1}(\mathbf{B}, \boldsymbol{\Sigma})$. □

C The Change of Variable Theorem

We will use the change of variable theorems outlined in Arias, Rubio-Ramírez, and Waggoner (2018). In particular, we will use Theorem 2 of that paper, which is reproduced here as Theorem 1.

Theorem 1. *Let $U \subset \mathbb{R}^b$ be an open set, let $\mathcal{V} \subset \mathbb{R}^a$ be a b-dimensional smooth manifold, and let $\gamma : U \to \mathcal{V}$ be a one-to-one and continuously differentiable function. If $A \subset \gamma(U)$ and*

$\lambda : A \to \mathbb{R}$ *is an integrable function, then:*

$$\int_A \lambda(\boldsymbol{v}) d_{\mathcal{V}} \boldsymbol{v} = \int_{\gamma^{-1}(A)} \lambda(\gamma(\boldsymbol{u})) |\det(D\gamma(\boldsymbol{u})' D\gamma(\boldsymbol{u}))|^{\frac{1}{2}} d\boldsymbol{u}. \tag{8}$$

D Sup-t Bayesian Credible Sets

Montiel Olea and Plagborg-Møller (2019) propose a method for joint inference on impulse responses that is equal to the Cartesian product of the commonly used equal-tailed posterior probability bands, where the tail probability is set to achieve the desired simultaneous credibility for the vector of impulse responses of interest. A key difference between this method and the one proposed by Inoue and Kilian (2022a) is that in the latter case the credible sets are constrained to contain impulse responses consistent with a given draw of the structural parameters. Even so, the results implied by both methods are similar as can be seen comparing Figure 14 with Figure 9, and Figure 15 with Figure 10.

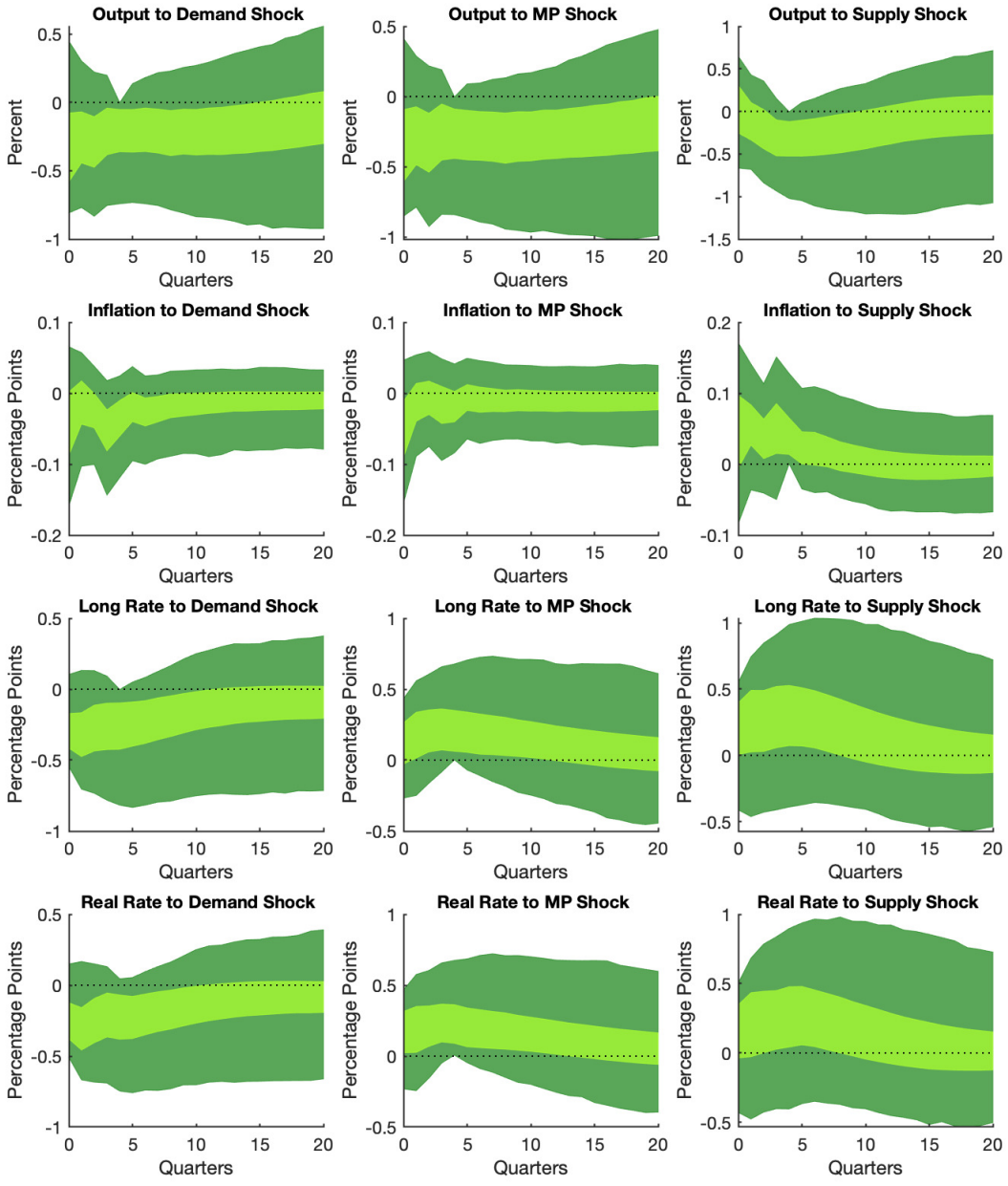


Figure 14: Sup-t 68 percent Bayesian credible set (dark green areas) based on Montiel Olea and Plagborg-Møller (2019). Light green areas show the equal-tailed 68 percent unconditional prior distributions of individual impulse responses. Both posteriors are implied by the uniform joint prior distribution for the vector of impulse responses.

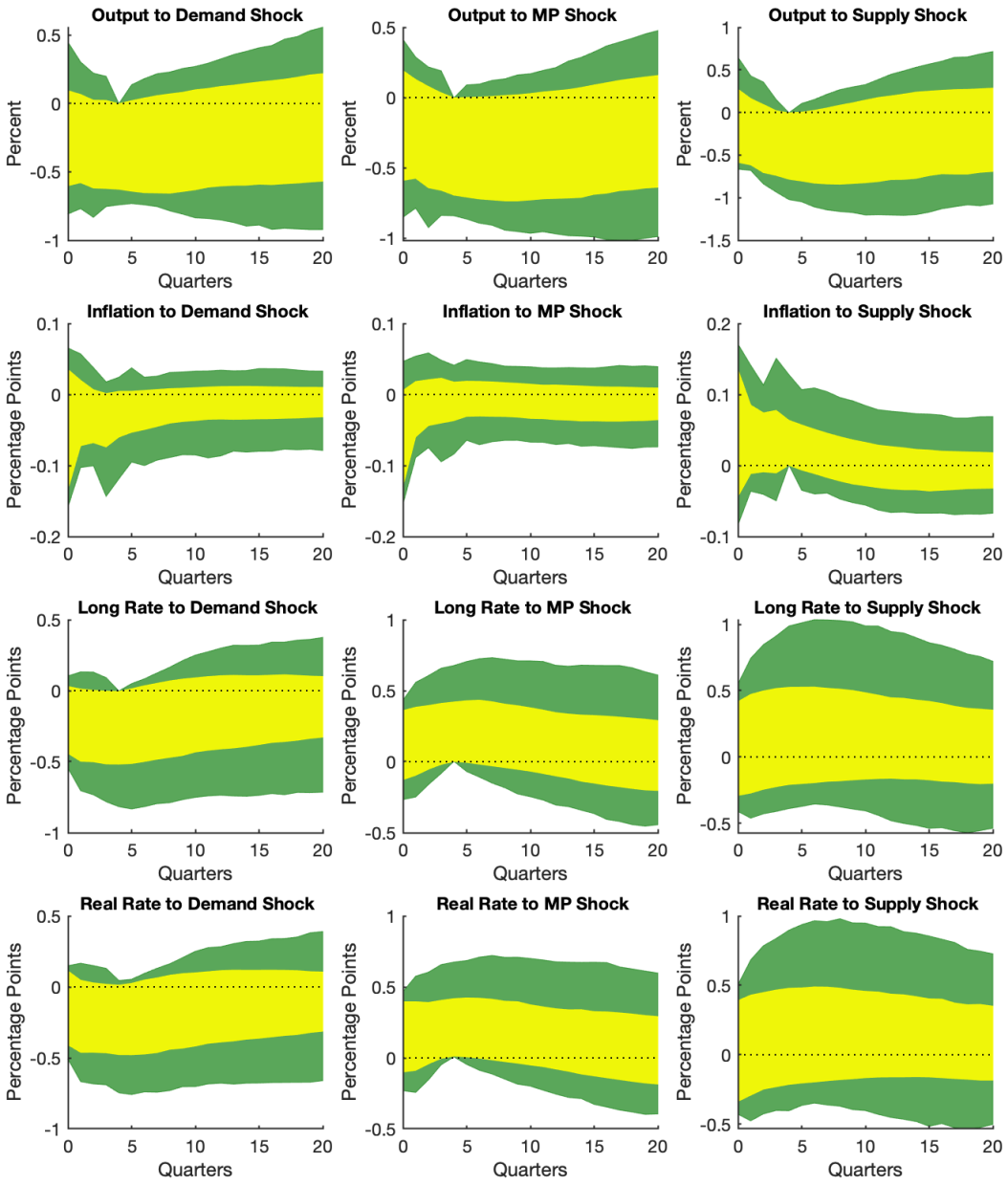


Figure 15: Sup-t 68 percent Bayesian 68 percent credible sets based on Montiel Olea and Plagborg-Møller (2019) under a uniform joint prior distribution for the vector of impulse responses (dark green areas) and under the Minnesota prior (yellow areas).