FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Inference Based on Scale, Label, and Economic Restrictions

Jonas E. Arias
Federal Reserve Bank of Philadelphia

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

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

WP 26-36

PUBLISHED
July 2026

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2026.36

Inference Based on Scale, Label, and Economic Restrictions

Jonas E. Arias

Federal Reserve Bank of Philadelphia

Juan Rubio-Ramírez*

Emory University

Federal Reserve Bank of Atlanta

Daniel F. Waggoner

Emory University and Federal Reserve Bank of Atlanta, Emeritus

July 14, 2026

Abstract

The results of nearly one hundred prominent studies in empirical macroeconomics have been called into question by Baumeister and Hamilton (2018). We show that their concern about distributional asymmetry for a typical question of interest under a uniform prior with respect to the Haar measure is actually driven by an unacknowledged sign restriction. We also demonstrate that such a prior induces symmetric prior distributions over individual impulse responses conditional on the reduced-form parameters, or more generally when the prior over the reduced-form covariance matrix rules out correlation among the residuals, as in the typical implementation of the Minnesota prior. Furthermore, we provide a theory for avoiding the pitfalls of Baumeister and Hamilton's critique. Key to our theory is a proposition establishing that any restriction can be decomposed into three types: scale, label, and economic. We use this theory to develop an algorithm for inference based on the unit modulus normalization that tackles a practical problem commonly faced by users of Bayesian SVAR methods.

JEL classification: C11, C32.

Keywords: structural vector autoregressions, unit modulus normalization.

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 The Concern

In a highly influential paper, Baumeister and Hamilton (2018) analyze a simple three-variable structural vector autoregression (SVAR) of the U.S. economy comprising the output gap, inflation, and the fed funds rate. They study how the output gap responds to a monetary policy shock normalized to raise the fed funds rate by 25 basis points on impact. Figure 1 reproduces panel (A) from Figure 1 in their paper depicting the impulse response of the output gap: the solid blue line depicts the point-wise posterior median and the blue-shaded area shows the 68 percent point-wise posterior probability bands. The figure suggests that the output gap is more likely to increase in response to a monetary policy contraction that raises the fed funds rate by 25 basis points.

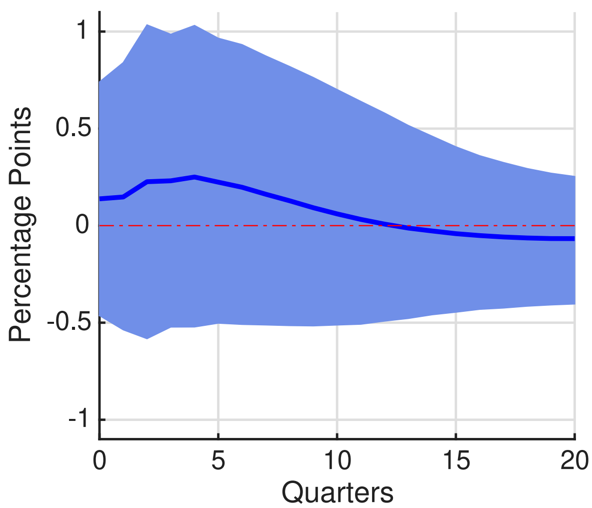
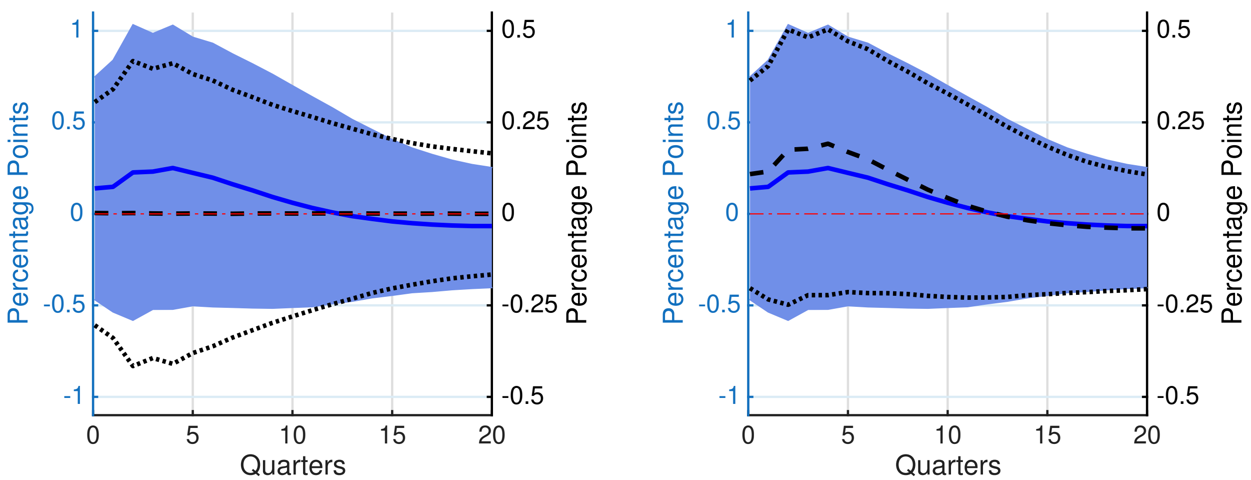


Figure 1: Impulse Response of the Output to a Monetary Policy Shock. The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent equal-tailed point-wise probability bands.

Baumeister and Hamilton (2018) find these impulse responses worrisome because, in their words, they "did not impose any sign restrictions at all, but simply kept every single draw." Baumeister and Hamilton (2018) obtain these responses by transforming draws generated with the conventional method of Uhlig (2005) and Rubio-Ramírez, Waggoner, and Zha (2010) so that the monetary policy shock raises the fed funds rate by 25 basis points on impact. They attribute the positive response of output to an "implicit Bayesian prior distribution" over the orthogonal matrices associated with the conventional method.

Importantly, Figure 1 uses the unit effect normalization as defined in Stock and Watson (2016), instead of the unit variance normalization for which the conventional method was designed. Specifically, under the unit variance normalization, the structural shocks are scaled to have unit variance, meaning the magnitudes of the impulse responses represent the effects of a typical realization of the shocks. In contrast, the unit effect normalization normalizes the shocks by fixing the contemporaneous effect of the structural shocks on a selected variable to some given values. This raises a natural question: what would the results have looked like if they had been reported under the unit variance normalization instead?



(a) Unit Effect vs. Unit Variance

(b) Unit Effect vs. Unit Variance + Sign

Figure 2: Impulse Responses of the Output to a Monetary Policy Shock. The solid blue (dashed) curves represent the point-wise posterior medians, and the blue-shaded areas (dotted lines) represent the 68 percent equal-tailed point-wise probability bands obtained under the unit effect (variance) normalization, with and without a sign restriction, panels (a) and (b), respectively.

Figure 2a illustrates this by comparing both normalizations. The solid blue line and the blue-shaded area reproduce the point-wise posterior median and 68 percent probability bands under the unit effect normalization, while the dashed and dotted black lines show the point-wise posterior median and 68 percent probability bands under the unit variance normalization. To facilitate the exposition, we present the results using two y-axes—the left y-axis corresponds to the unit effect normalization and the right y-axis to the unit variance normalization. Under the unit variance normalization, the posterior median of the impulse response of output is centered on zero and the point-wise posterior probability bands are symmetric around the median.

The Sign Restriction

As Stock and Watson (2016) highlight, the unit effect normalization entails fixing both the scale and the label of some impulse responses, whereas the unit variance normalization only determines the scale. In the case at hand, the unit effect normalization restricts the initial response of the fed funds rate to positive 25 basis points, while under the unit variance normalization, the initial response of the fed funds rate to a typical realization of the shock can be either positive or negative. Because Baumeister and Hamilton (2018) label their shock as a contractionary monetary policy shock by imposing a sign restriction requiring the fed funds rate to increase contemporaneously, they retain only those impulse responses that satisfy the restriction. In contrast, the unit variance normalization does not impose the label restriction, and all the draws are retained. To illustrate the effects of this label restriction, Figure 2b adds the aforementioned sign restriction to the unit variance normalization; we label this case as "unit variance normalization + sign" in the figure. Once the sign restriction is added to the unit variance normalization, the positive effects on output that concerned Baumeister and Hamilton (2018) reemerge; the results using the unit effect or the unit variance normalization + sign are almost identical up to scale.

The Mathematics Behind the Asymmetry

Let us try to understand how the prior over the orthogonal reduced-form parameterization and the sign restriction implicit in the unit effect normalization interact to shape the prior and posterior over the other impulse responses. To make things simpler, our starting point is an SVAR without lags written under the unit variance normalization to which we add the sign restriction discussed above. Let $\mathbf{\Sigma}$ be the variance-covariance matrix of the reduced-form innovations. Let $\mathbf{L}_0$ be the matrix of contemporaneous impulse responses to the structural shocks, so that $\mathbf{L}_0(i,j)$ is the contemporaneous response of the i^{th} variable to the j^{th} structural shock. The relationship between $\mathbf{\Sigma}$ and $\mathbf{L}_0$ is $\mathbf{\Sigma} = \mathbf{L}_0\mathbf{L}_0'$. Alternatively, if $\mathbf{\Sigma} = r(\mathbf{\Sigma})r(\mathbf{\Sigma})'$ is the Cholesky decomposition of $\mathbf{\Sigma}$, where $r(\mathbf{\Sigma})$ is lower triangular with positive diagonal, then $\mathbf{L}_0 = r(\mathbf{\Sigma})\mathbf{Q}$, for any symmetric and positive definite matrix $\mathbf{\Sigma}$ and orthogonal matrix $\mathbf{Q}$. So, we can parameterize the structural contemporaneous impulse responses as either $\mathbf{L}_0$, where $\mathbf{L}_0$ is invertible, or $(\mathbf{\Sigma},\mathbf{Q})$, where $\mathbf{\Sigma}$ is symmetric and positive definite and $\mathbf{Q}$ is orthogonal. We will refer to the former as the impulse response parameterization and to the latter as the orthogonal reduced-form parameterization. Priors over the orthogonal reduced-form parameterization induce priors over the impulse response parameterization, and vice versa. Let $\mathbf{P}_j^-$ be the diagonal orthogonal matrix defined by $\mathbf{P}_j^-(j,j) = -1$ and $\mathbf{P}_j^-(i,i) = 1$, for $i \neq j$. We have the following result. Throughout this paper, all densities over the orthogonal matrices are with respect to the Haar measure.

Proposition 1. *Fix $1 \leq j \leq n$. If the prior $p(\mathbf{\Sigma},\mathbf{Q}) = p(\mathbf{\Sigma})p(\mathbf{Q} \mid \mathbf{\Sigma})$ satisfies $p(\mathbf{Q} \mid \mathbf{\Sigma}) = p(\mathbf{Q}\mathbf{P}_j^- \mid \mathbf{\Sigma})$, then the induced marginal prior and posterior distributions over $\mathbf{L}_0(i,j)$ are symmetric about the origin for $1 \leq i \leq n$.*

Proof. See Appendix. □

The proposition gives us only a sufficient condition, hence, if the prior does not satisfy $p(\mathbf{Q} \mid \mathbf{\Sigma}) = p(\mathbf{Q}\mathbf{P}_j^- \mid \mathbf{\Sigma})$, then the induced marginal prior over $\mathbf{L}_0(i,j)$ could be symmetric about the origin, but this would require a delicate relationship between $p(\mathbf{\Sigma})$ and $p(\mathbf{Q} \mid \mathbf{\Sigma})$, a relationship that would likely not carry over to the posterior since $p(\mathbf{Q} \mid \mathbf{\Sigma})$ does not update with data. In any case, Proposition 1 makes clear that, in the absence of sign restrictions and regardless of the prior and posterior over the reduced-form, a uniform prior over the orthogonal matrices implies that both the marginal prior and posterior over $\mathbf{L}_0(i,j)$ are symmetric about the origin.

Suppose we use a uniform prior over the orthogonal matrices and analyze how introducing a sign restriction affects the symmetry of the prior over $\mathbf{L}_0(i,j)$. Continuing with the example in Baumeister and Hamilton (2018), suppose that $\mathbf{L}_0(\text{ffr,mp})$ is the response of the fed funds rate to a monetary policy shock and that $\mathbf{L}_0(\text{y,mp})$ is the response of the output gap to a monetary policy shock. How does restricting the response of the fed funds rate to a monetary policy shock to be positive affect the output gap response? Can it induce asymmetry in the output gap response? The example in the previous section shows that this can indeed occur. Is the uniform prior over the orthogonal matrices responsible for such asymmetry? The following proposition shows, at least for two cases considered in Baumeister and Hamilton (2018), that any asymmetry is driven by the interaction between the sign or label restriction and the reduced-form covariance structure, rather than by the uniform prior over orthogonal matrices.

Proposition 2. *Suppose the prior distribution over* $\mathbf{Q}$*, conditional on* $\mathbf{\Sigma}$*, is uniform.*

(i) Suppose the distribution over $\mathbf{\Sigma}$ *is inverse Wishart. The marginal distribution of* $\mathbf{L}_0(y,mp)$*, conditional on* $\mathbf{L}_0(ffr,mp) > 0$*, is symmetric if and only if the expected correlation between the innovations to the fed funds rate and the output gap is zero.*

(ii) Suppose $\mathbf{\Sigma}$ *is fixed. The marginal distribution of* $\mathbf{L}_0(y,mp)$*, conditional on* $\mathbf{L}_0(ffr,mp) >$ 0*, is symmetric if and only if the correlation between the innovations to the fed funds rate and the output gap is zero.*

Proof. See Appendix. □

If the inverse Wishart parameters in Proposition 2 (i) are ν and $\mathbf{\Phi}$, then the expected value of $\mathbf{\Sigma}$ is $\mathbf{\Phi}/(\nu - n - 1)$. Thus, the expected correlation between the innovations of the fed funds rate and the output gap is zero if and only if $\mathbf{\Phi}(\text{ffr,y}) = 0$. If we use a natural conjugate prior, then Proposition 2 (i) applies to both the prior and the posterior. Suppose that $\bar{\mathbf{\Phi}}$ and $\tilde{\mathbf{\Phi}}$ are the prior and posterior values of $\mathbf{\Phi}$. Proposition 2 (i) implies that any asymmetry in the prior marginal distribution of $\mathbf{L}_0(\text{y,mp})$, conditional on $\mathbf{L}_0(\text{ffr,mp}) > 0$, is completely driven by the value of $\bar{\mathbf{\Phi}}(\text{ffr,y})$, not the uniform prior over the orthogonal matrices. Furthermore, even if $\bar{\mathbf{\Phi}}(\text{ffr,y}) = 0$, one would not expect $\tilde{\mathbf{\Phi}}(\text{ffr,y}) = 0$ because of the correlation between the fed funds rate and the output gap in the data. So, Proposition 2 (i) makes clear that when we introduce the sign restriction implicit in the unit effect normalization,

(a) the prior over the reduced-form parameters, not the uniform prior over the orthogonal matrices, is responsible if the prior marginal distribution of $\mathbf{L}_0(\text{y,mp})$, conditional on $\mathbf{L}_0(\text{ffr,mp}) > 0$, is asymmetric,

(b) even if the prior is symmetric, the posterior will, in general, not be.

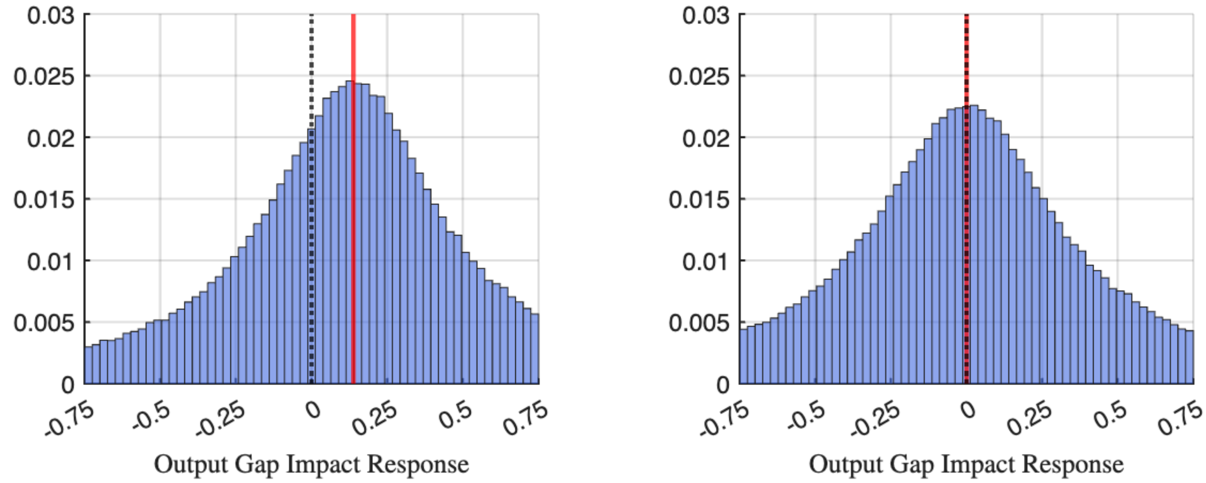
Proposition 2 (ii) shows that analogous results hold if we fix $\mathbf{\Sigma}$. To conclude, while Baumeister and Hamilton (2018) blame the positive response of output on the "implicit Bayesian prior distribution" over the orthogonal matrices associated with the conventional method, the results of this section show that this is clearly not true. In the case analyzed in that paper, the positive response of output is due to the nonzero expected posterior correlation between the innovations of the fed funds rate and the output gap.

Illustration

To illustrate the results above, we reproduce the histogram in panel D of Figure 1 in Baumeister and Hamilton (2018) as Figure 3a.[1] This histogram shows the contemporaneous response of the output gap to a monetary policy shock identified using the unit variance normalization + sign approach. As in panel D of Figure 1 in Baumeister and Hamilton (2018), the reduced-form parameters are set to their maximum likelihood estimates. At this value, the covariance between the fed funds rate and the output gap, $\mathbf{\Sigma}$(ffr,y), equals 0.05. The red lines in each panel correspond to the median of the distribution under consideration and the dotted black line denotes the origin.

Proposition 2 (ii) makes clear that the asymmetry in the response is driven by the sign restriction combined with the fact that the correlation between the innovations to the fed funds rate and the output gap is positive. To illustrate this point, Figure 3b shows the distribution of the contemporaneous response of the output gap in the alternative case in which we artificially set $\mathbf{\Sigma}$(ffr,y) to zero. In this alternative case, the contemporaneous response of output is symmetric around zero. Hence, the figure reaffirms that, contrary to the claims in Baumeister and Hamilton (2018), the uniform prior over the orthogonal matrices is not the channel that causes the asymmetry in the response of output to a monetary policy shock under the unit effect normalization; rather, it is the posterior over $\mathbf{\Sigma}$.

[1]The reader should notice that the axes are different. We focus on $[-0.75, 0.75]$ while Baumeister and Hamilton (2018) focus on $[-2, 2]$. We did that so that the positive median can be seen more clearly.



(a) $\boldsymbol{\Sigma}$(ffr,y) = 0.05 (b) $\boldsymbol{\Sigma}$(ffr,y) = 0

Figure 3: Unit Effect Normalization and $\boldsymbol{\Sigma}_{\mathrm{ffr,y}}$

This Paper

The results above make clear that Baumeister and Hamilton (2018)'s troubling output response is not caused by the uniform prior over orthogonal matrices. Instead, it comes from a sign restriction hidden inside the unit-effect normalization, interacting with the reduced-form covariance structure. Does this mean that we should give up on pinning down the impact effect of the structural shock and use only the unit variance normalization? The answer is no, as choosing the impact of the structural shock facilitates direct estimation of impulse responses in the native units of the data, such as percentage points for interest rates, enhancing the interpretability of results for policy analysis. This paper proposes an alternative normalization that restricts the impact of the structural shock while not imposing the sign restriction that the unit effect normalization imposes. We call such a normalization the "unit modulus normalization" and show how to do Bayesian inference under that normalization.

But before moving forward, we need to address an important practical issue for the case under study. In particular, we answer the question of whether, in general, we can distinguish restrictions that solely determine the scale of the model from other types of restrictions. The answer is yes. In Section 2, we develop a general theory of restrictions that classifies them as scale, label, or economic restrictions. We will use this framework to define the unit modulus normalization in Section 3 and to present an algorithm for inference under such a normalization in Section 4.

2 General Restrictions

One important conclusion from Section 1 is that not all normalizations are alike: while the unit variance normalization only restricts the scale, the unit effect normalization pins down both the scale and the sign of a given impulse response. Pinning down the sign of a selected impulse response implies choosing a label for the shock, which in the context of our example determines whether the monetary policy shock is contractionary or expansionary. As shown, depending on the prior over the reduced-form parameters, the sign restriction associated with choosing the label of the shock can have important effects on the prior distribution of the impulse responses. Hence, the researcher must be able to choose any scale without imposing a label. More generally, we need a theory that allows us to disentangle labeling restrictions from scale restrictions and from economic restrictions. To address this gap, we develop a general theory of restrictions that classifies them as scale, label, and economic restrictions. Using our theory, we define a new normalization labeled "unit modulus" that restricts the scale of the shock as in the unit effect normalization without assigning a label.

The Model

Consider a reduced-form vector autoregression (VAR) of the form:

$$\mathbf{y}_t' = \mathbf{y}_{t-1}'\mathbf{B}_1 + \cdots + \mathbf{y}_{t-p}'\mathbf{B}_p + \mathbf{d} + \mathbf{u}_t', \text{ for } 1 \leq t \leq T, \tag{1}$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of endogenous variables, $\mathbf{u}_t$ is an $n \times 1$ vector of reduced-form shocks, $\mathbf{B}_\ell$ is an $n \times n$ matrix of parameters for $1 \leq \ell \leq p$, $\mathbf{d}$ is a $1 \times n$ vector of parameters, p is the lag length, and T is the sample size. The vector $\mathbf{u}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, is Gaussian with mean zero and covariance matrix $\mathbf{\Sigma}$. For a more compact notation, let $\mathbf{B} = \begin{bmatrix}\mathbf{B}_1' \cdots \mathbf{B}_p' & \mathbf{d}'\end{bmatrix}'$ and $\mathbf{x}_t' = \begin{bmatrix}\mathbf{y}_{t-1}' \cdots \mathbf{y}_{t-p}' & 1\end{bmatrix}$ for $1 \leq t \leq T$

and hence Equation (1) can be written as:

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

The reduced-form parameters are $(\mathbf{B}, \mathbf{\Sigma})$, where $\mathbf{B} \in \mathbb{R}^{m\times n}$ and $\mathbb{R}^{m\times n}$ is the set of $m \times n$ matrices, with $m = np + 1$ denoting the number of regressors per equation, and $\mathbf{\Sigma} \in \mathcal{S}_n^+$, where $\mathcal{S}_n^+$ denotes the set of $n \times n$ symmetric and positive definite matrices. A reduced-form VAR is identified, so the reduced-form parameters can be estimated from the data.

Often, one is interested in the effects of the structural shocks rather than the reduced-form shocks. We assume that the reduced-form shocks are a function of the structural shocks of the form:

$$\mathbf{u}_t = \mathbf{L}_0\boldsymbol{\varepsilon}_t, \text{ for } 1 \leq t \leq T, \tag{2}$$

where $\boldsymbol{\varepsilon}_t$ is the $n\times 1$ vector of structural shocks and $\mathbf{L}_0$ is an $n\times n$ invertible matrix. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, is Gaussian with mean zero and covariance matrix $\mathbf{\Omega}$, where $\mathbf{\Omega} \in \mathcal{D}_n^+$ where $\mathcal{D}_n^+$ is the set of $n\times n$ positive diagonal matrices. The element in row i and column j of $\mathbf{L}_0$ is the contemporaneous effect on the i^{th} variable of the structural shock $\boldsymbol{\varepsilon}_t = \mathbf{e}_j$, where $\mathbf{e}_j$ is the j^{th} column of $\mathbf{I}_n$, the $n \times n$ identity matrix. Substituting Equation (2) into Equation (1) gives a structural-form representation of the reduced-form VAR model. The structural-form parameters are $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$. Note that there are many other ways to represent a structural-form.[2] There is a mapping $g(\cdot)$ from these structural-form parameters onto the reduced-form parameters given by

$$g(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) = (\mathbf{B}, \underbrace{\mathbf{L}_0\mathbf{\Omega}\mathbf{L}_0'}_{\mathbf{\Sigma}}).$$

No structural-form parameterization is identified. In general, two structural-form parameters

[2]For instance, one of the most common structural-forms of the reduced-form VAR, sometimes called the structural form, is

$$\mathbf{y}_t'\mathbf{A}_0 = \mathbf{y}_{t-1}'\mathbf{A}_1 + \cdots + \mathbf{y}_{t-p}'\mathbf{A}_p + \mathbf{c} + \boldsymbol{\varepsilon}_t', \text{ for } 1 \leq t \leq T.$$

Its relationship to the reduced-form is given by $\mathbf{B}_\ell = \mathbf{A}_\ell\mathbf{A}_0^{-1}$, for $1 \leq \ell \leq p$, $\mathbf{d} = \mathbf{c}\mathbf{A}_0^{-1}$, $\mathbf{u}_t = (\mathbf{A}_0^{-1})'\boldsymbol{\varepsilon}_t$, and $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, is standard normal. Arias, Rubio-Ramirez, and Waggoner (2025) describe another popular parameterization where the structural-form parameters consist of contemporaneous and dynamic impulse responses.

are observationally equivalent if and only if they map to the same reduced-form parameters. Proposition 3 gives necessary and sufficient conditions for observational equivalence for our structural-form parameterization. The proof of this proposition is a straightforward extension of well-known ideas in the literature, but for completeness, a proof is given in the Appendix. Before stating the results, we develop some notation. For $\mathbf{\Omega} \in \mathcal{D}_n^+$, the matrix $\mathbf{\Omega}^{\frac{1}{2}}$ is the unique positive diagonal matrix such that $\mathbf{\Omega}^{\frac{1}{2}}\mathbf{\Omega}^{\frac{1}{2}} = \mathbf{\Omega}$. The set of all $n \times n$ orthogonal matrices is $\mathcal{O}_n$ and the set of all $n \times n$ diagonal orthogonal matrices is denoted by $\mathcal{O}_n^D$. A diagonal orthogonal matrix has plus and minus ones along its diagonal and zero in all other entries.

Proposition 3. $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ *and* $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$ *are observationally equivalent if and only if* $\tilde{\mathbf{B}} = \mathbf{B}$ *and there exists a* $\mathbf{Q} \in \mathcal{O}_n$ *such that* $\tilde{\mathbf{L}}_0 = \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}}$.

Proof. See Appendix. □

Proposition 3 is useful for determining if two given structural-form parameters, $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$, are observationally equivalent. One need only check whether $\mathbf{B} = \tilde{\mathbf{B}}$ and whether $\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{L}_0^{-1}\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}^{\frac{1}{2}}$ is orthogonal. Furthermore, a corollary of Proposition 3 can be used to generate all observationally equivalent structural-form parameters to some given structural-form parameter.

Corollary 1. $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ *and* $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$ *are observationally equivalent if and only if the latter is of the form*

$$(\mathbf{B}, \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}),$$

where $\mathbf{Q} \in \mathcal{O}_n$ *and* $\mathbf{D} \in \mathcal{D}_n^+$.

Proof. See Appendix. □

Together, the proposition and the corollary lay a foundation for classifying the structural-form parameters into the three classes of observationally equivalent parameters. We will use these classes to define our three types of restrictions, which will allow us to partition all restrictions into scale, label, and economic restrictions.

Three Classes of Observationally Equivalent Parameters

Corollary 1 allows one to produce all the observationally equivalent parameters to $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ in terms of an orthogonal matrix $\mathbf{Q}$ and a positive diagonal matrix $\mathbf{D}$. This will be important for defining three classes of observationally equivalent parameters. These three classes will allow us to partition the set of all restrictions into what we call scale, label, and economic restrictions.

Class 1. $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) = \{(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega}\mathbf{D}) \mid \mathbf{D} \in \mathcal{D}_n^+\}$.

The elements in $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ rescale the variance and contemporaneous effects of the structural shocks. If we scale the variance of the structural shocks by $\mathbf{D} \in \mathcal{D}_n^+$, then to maintain observational equivalence we must scale the contemporaneous effects of the structural shocks by $\mathbf{D}^{-\frac{1}{2}}$. So, each element of $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ corresponds to a choice of scale for the variance and the contemporaneous effects of the structural shocks.

Class 2. $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) = \{(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \boldsymbol{\Omega}) \mid \mathbf{P} \in \mathcal{O}_n^D\}$.

Elements of $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ change the sign of some of the columns of $\mathbf{L}_0$. Equivalently, this changes the sign of the contemporaneous effect of some of the structural shocks for all the variables. So, each element of $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ corresponds to a choice of sign of the contemporaneous effect of each structural shock for all the variables.

Class 3. $\mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) = \{(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})\} \cup \{(\mathbf{B}, \mathbf{L}_0\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{N}\boldsymbol{\Omega}^{-\frac{1}{2}}, \boldsymbol{\Omega}) \mid \mathbf{N} \in \mathcal{O}_n \backslash \mathcal{O}_n^D\}$, *where* $\mathcal{O}_n \backslash \mathcal{O}_n^D$ *denotes the complement of* $\mathcal{O}_n^D$ *in* $\mathcal{O}_n$.

Because $\mathcal{O}_n \backslash \mathcal{O}_n^D$ excludes the identity matrix, we need to explicitly include $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ in the definition. What do elements of $\mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ look like? The non-zero off-diagonal elements of $\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{N}\boldsymbol{\Omega}^{-\frac{1}{2}}$ occur in the same positions as the non-zero off-diagonal elements of $\mathbf{N}$. Suppose that the j^{th} column of $\mathbf{N}$ has non-zero off-diagonal elements. This implies that the response of the variables to the j^{th} shock in $(\mathbf{B}, \mathbf{L}_0\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{N}\boldsymbol{\Omega}^{-\frac{1}{2}}, \boldsymbol{\Omega})$ has a different economic interpretation than the response of the variables to the same shock in $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$, in the sense that the contemporaneous responses are not just a rescaling or a change in sign. However, if the off-diagonal elements of $\mathbf{\Omega}^{\frac{1}{2}}\mathbf{N}\mathbf{\Omega}^{-\frac{1}{2}}$ are small relative to the diagonal elements, then the economic interpretations of $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\mathbf{B}, \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{N}\mathbf{\Omega}^{-\frac{1}{2}}, \mathbf{\Omega})$ will be similar.

It is easy to verify that $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) = \{(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})\}$. In this sense, we have divided the observationally equivalent parameters into three distinct classes that we will use to define scale, label, and economic restrictions in the next section.

Three Types of Restrictions: Scale, Label, and Economic

We define scale, label, and economic restrictions here and, in the next section, show how restrictions can be decomposed into a combination of these types. First, we review the general definition of a restriction and some related terminology. A restriction is simply a subset $\mathcal{R}$ of the set of all structural-form parameters. The restriction is satisfied by $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ if and only if $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$. If $\mathcal{R}_1$ and $\mathcal{R}_2$ are two restrictions, then $\mathcal{R}_1 \cap \mathcal{R}_2$ is the restriction that is satisfied if and only if both $\mathcal{R}_1$ and $\mathcal{R}_2$ are satisfied. We refer to this as the combination of $\mathcal{R}_1$ and $\mathcal{R}_2$. A restriction is trivial if and only if its complement is of measure zero. A restriction $\mathcal{R}$ on the structural-form parameters induces a restriction $g(\mathcal{R})$ on the reduced-form parameters, which could be trivial.

What should a scale, label, or economic restriction look like? We assert that a scale restriction should not restrict the sign of the contemporaneous response of a shock, a label restriction should not restrict the scale of the variance or contemporaneous effects of a shock, and both scale and label restrictions should induce trivial restrictions over the reduced-form parameters. Economic restrictions may induce non-trivial restrictions over the reduced-form parameters, but they may not restrict the sign of the contemporaneous response of a shock or the scale of the variance or contemporaneous effects of a shock.

Definition 1 (**Scale Restriction**). *The restriction $\mathcal{R}$ is a scale restriction if and only if the following conditions hold*

S1 $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ *for every* $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$.

S2 $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{R} \neq \varnothing$, *for almost all structural-form parameters* $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$.

Condition S1 says that a scale restriction cannot restrict the sign of the contemporaneous response of a shock. Condition S2 says that for almost all structural-form parameters there is a rescaling of the shock variances and contemporaneous responses that will satisfy the restriction, which implies that the induced restriction over the reduced-form parameters is trivial. Often, we want the scale of each shock to be uniquely determined, which would imply that the set $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{R}$ is a singleton for almost all structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$; however, sometimes we want only to uniquely determine the scale on a subset of the shocks. Our definition of a scale restriction allows this.

Definition 2 (**Label Restriction**). *The restriction $\mathcal{R}$ is a label restriction if and only if the following conditions hold*

L1 $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ *for every* $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$.

L2 $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{R} \neq \varnothing$, *for almost all structural-form parameters* $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$.

Condition L1 says that a label restriction cannot restrict the scale of the variance or contemporaneous response of a shock. Condition L2 says that for almost all structural-form parameters there is a choice of sign for the contemporaneous responses to each shock that will satisfy the restriction, which implies that the induced restriction over the reduced-form parameters is trivial. Often, we want the label of each shock to be uniquely determined, which would imply that the set $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{R}$ is a singleton for almost all structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$; however, sometimes we want only to uniquely determine the label on a subset of the shocks. Our definition of a label restriction allows this.

We have the following result, which states that a non-trivial restriction cannot be both a scale and a label restriction.

Proposition 4. *If a restriction is both a scale and a label restriction, then it is trivial.*

Proof. See Appendix. □

Why do we call this type of restriction a label restriction rather than a sign restriction? Structural shocks always come in two orientations. A demand shock can either increase or decrease demand. A supply shock can either increase or decrease supply. A monetary policy shock is either an expansion or a contraction. For instance, suppose that the first shock is a monetary policy shock. For a given structural-form parameter, does a positive value of the first shock correspond to an expansion or a contraction? The answer to this question gives rise to a *labeling* of the first shock. Of course, if a positive value of the first shock under $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ is labeled an expansion, then a positive value of the first shock under $(\mathbf{B}, \mathbf{L}_0\mathbf{P}_1^-, \boldsymbol{\Omega})$ must be labeled a contraction. Typically, the label for a shock is chosen based on the sign of a given contemporaneous impulse response, but there are also cases in which the label of the shock is chosen based on the sign of some function of the contemporaneous impulse responses. For instance, Christiano, Eichenbaum, and Evans (1996) assume that if the impact of a unit monetary policy shock on the fed funds rate is positive before the reaction function of the Federal Reserve is taken into account, then this shock is a contraction.

A labeling of the j^{th} shock gives rise to a label restriction. Such a labeling consists of two labels, l_1 and l_2, a subset $\mathcal{L}_j$ of the structural-form parameters over which the labeling is defined, and a function $\ell_j(\cdot)$ from $\mathcal{L}_j$ into $\{l_1, l_2\}$. We require that if $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{L}_j$, then $(\mathbf{B}, \mathbf{L}_0\mathbf{P}_j^-, \boldsymbol{\Omega}) \in \mathcal{L}_j$ and $\ell_j(\mathbf{B}, \mathbf{L}_0\mathbf{P}_j^-, \boldsymbol{\Omega}) \neq \ell_j(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$. The interpretation of $\ell_j(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ is that it is the label assigned to a positive value of the j^{th} shock. We assume that a labeling cannot depend on the scale. This means that if $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{L}_j$, then $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \subset \mathcal{L}_j$ and $\ell_j(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) = \ell_j(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega}\mathbf{D})$ for every $\mathbf{D} \in \mathcal{D}_n^+$. Given a labeling of the j^{th} shock, we can define a restriction by

$$\mathcal{R}_j = \{(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \mid (\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \notin \mathcal{L}_j \text{ or } (\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{L}_j \text{ and } \ell_j(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) = l_1\}$$

Requiring the complement of $\mathcal{L}_j$ to be contained in $\mathcal{R}_j$ ensures that Condition L2 is satisfied and, since the labeling cannot depend on the scale, Condition L1 is satisfied, so $\mathcal{R}_j$ is a label restriction. For all structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$, either both $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ and $(\mathbf{B}, \mathbf{L}_0\mathbf{P}_j^-, \boldsymbol{\Omega})$ are in $\mathcal{R}_j$, or exactly one of these is in $\mathcal{R}_j$. In the latter case, a positive value of the j^{th} shock is given the label l_1 if $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R}_j$. We can use this idea to define a labeling of the j^{th} structural shock from any label restriction.

If $\mathcal{R}$ is a label restriction, define $\mathcal{L}_j$ to be the set of all structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ for which exactly one of $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ or $(\mathbf{B}, \mathbf{L}_0\mathbf{P}_j^-, \boldsymbol{\Omega})$ satisfies the restriction. The set $\mathcal{L}_j$ could be empty. The function $\ell_j(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ is l_1 if $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ satisfies the restriction and l_2 otherwise. The fact that $\mathcal{R}$ is a label restriction ensures that the requirements for the set $\mathcal{L}_j$ and the function ℓ_j discussed above are satisfied. If we go from a labeling to a label restriction and then back to a labeling, we recover the original labeling. It is in this sense that the term "label restriction" is justified. However, when we go from a label restriction to a labeling and then back to a label restriction, we do not necessarily end up where we started. This is because a generic label restriction could determine the label for more than one shock. Finally, while the labeling or restriction was defined in terms of the first label, it could just as easily be defined in terms of the second label.

Definition 3 (**Economic Restriction**). *The restriction $\mathcal{R}$ is an economic restriction if and only if the following condition holds*

E1 $\mathcal{C}_S(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ *and* $\mathcal{C}_L(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ *for every* $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$.

Condition E1 says that an economic restriction can restrict neither the scale of the variance nor the contemporaneous response of a shock, nor the sign of the contemporaneous response of a shock. Unlike scale and label restrictions, we do not require $\mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \cap \mathcal{R} \neq \varnothing$, for almost all structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$, because we want to allow economic restrictions to restrict the reduced-form parameters.

Proposition 5. *If an economic restriction is either a scale restriction or a label restriction, then it is trivial.*

Proof. See Appendix. □

One might wonder why we did not include $\mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ for every $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$ in the first condition of either a scale restriction or a label restriction. The reason is that if we did, then some of the most commonly used scale or label restrictions would no longer be scale or label restrictions. Forcing $\mathcal{C}_E(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \subset \mathcal{R}$ for every $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$ would not give label or scale restrictions enough flexibility to encompass a sufficiently wide range of restrictions.

A Decomposition Theorem

In this section, we show how to decompose a restriction into a combination of these three types of restrictions, at least for a large class of restrictions. In particular, we consider restrictions that do not impose a relation between the label and scale. Formally, we have the following definition.

Definition 4 (**Relation Between the Label and Scale**). *A restriction $\mathcal{R}$ imposes a relation between the label and scale if and only if there is a structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \notin \mathcal{R}$, a $\mathbf{P} \in \mathcal{O}_n^D$, and a $\mathbf{D} \in \mathcal{D}_n^+$ such that both $(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \boldsymbol{\Omega}) \in \mathcal{R}$ and $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega}\mathbf{D}) \in \mathcal{R}$.*

Conditions S1, L1, and E1 imply that scale, label, and economic restrictions do not impose a relation between the label and scale. These are the prototypical restrictions that do not impose a relation between the label and scale. The following proposition implies that combinations of scale, label, and economic restrictions will not impose a relation between the label and scale.

Proposition 6. *The intersection of restrictions that do not impose a relation between the label and scale will not impose a relation between the label and scale.*

Proof. See Appendix. □

The converse of this proposition is also true, as the following proposition states.

Proposition 7. *A restriction $\mathcal{R}$ does not impose a relation between the label and scale if and only if it can be decomposed as $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$, where $\mathcal{R}_e$ is an economic restriction, $\mathcal{R}_\ell$ is a label restriction, and $\mathcal{R}_s$ is a scale restriction.*

Proof. See Appendix. □

It is useful to have explicit expressions for $\mathcal{R}_e$, $\mathcal{R}_\ell$, and $\mathcal{R}_s$.

$$\mathcal{R}_e = \{(\mathbf{B}, \mathbf{L}_0\mathbf{P}\mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega}\mathbf{D}) \mid (\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R},\ \mathbf{D} \in \mathcal{D}_n^+, \text{ and } \mathbf{P} \in \mathcal{O}_n^D\}, \tag{3}$$

$$\mathcal{R}_\ell = \underbrace{\{(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega}\mathbf{D}) \mid (\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R} \text{ and } \mathbf{D} \in \mathcal{D}_n^+\}}_{\mathcal{R}_d} \cup \mathcal{R}_e^c, \tag{4}$$

$$\mathcal{R}_s = \underbrace{\{(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \boldsymbol{\Omega}) \mid (\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R} \text{ and } \mathbf{P} \in \mathcal{O}_n^D\}}_{\mathcal{R}_p} \cup \mathcal{R}_e^c, \tag{5}$$

where the superscript c denotes the complement in the set of all structural-form parameters.

The studies we are aware of do not use restrictions that impose a relation between the label and scale, so this proposition applies to a broad class of useful restrictions. However, if one wanted to allow such a relation, then a version of this proposition would still hold, but one of Conditions S1 and L1 would have to be omitted. Both cannot be omitted because then Proposition 4 would not hold. One must decide whether to place the relation between the label and scale in the label restriction or in the scale restriction. If Condition S1 is omitted from the scale restriction definition, then $\mathcal{R}_s$ should be defined as $\mathcal{R} \cup \mathcal{R}_d^c$. If Condition L1 is omitted from the label restriction definition, then $\mathcal{R}_\ell$ should be defined as $\mathcal{R} \cup \mathcal{R}_p^c$. In particular, we have the following result.

Proposition 8. *Suppose that either S1 or L1 is omitted from the definition of a scale or label restriction. For any restriction $\mathcal{R}$, there exists an economic restriction $\mathcal{R}_e$, a label restriction $\mathcal{R}_\ell$, and a scale restriction $\mathcal{R}_s$ such that $\mathcal{R} = \mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$.*

Proof. See Appendix. □

It is important to verify that a restriction $\mathcal{R}$ does not impose a relation between the label and scale before applying Proposition 7, because $\mathcal{R}_e$, $\mathcal{R}_\ell$, and $\mathcal{R}_p$ can always be defined using Equations (3)-(5), but $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_p \neq \mathcal{R}$ if $\mathcal{R}$ imposes a relation between the label and scale. Fortunately, in practice, restrictions are specified by a set of equalities and inequalities, and for a restriction defined by a single equality or inequality it is often easy to determine if S1 or L1 holds. If S1 or L1 holds for each equality or inequality, then from the discussion following Definition 4 and Proposition 6, $\mathcal{R}$ will not impose a relation between the label and scale.

Finally, the decomposition in Proposition 7 is not unique, as we shall explicitly see in the next section. Furthermore, as shown in the proof of Proposition 7, for the decomposition given by Equations (3)-(5), both S2 and L2 will hold for all $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$, not just for almost all. For this reason, the decomposition given by Equations (3)-(5) is often complicated, but in many cases it is straightforward to devise simpler decompositions.

Examples of Decompositions

This section provides examples of how Proposition 7 works in practice.

Example 1. *Suppose that the first shock is a demand shock, the first variable is price, and the second variable is quantity. Consider the restriction $\mathcal{R}$ that requires both $\mathbf{L}_0(1,1) > 0$ and $\mathbf{L}_0(2,1) > 0$, so that the contemporaneous response to a positive demand shock strictly increases both price and quantity.*

We decompose this restriction using Equations (3)-(5). It is easy to see that $\mathcal{R}_d = \mathcal{R}$ and $\mathcal{R}_p$ requires $\mathbf{L}_0(1,1)\mathbf{L}_0(2,1) > 0$. So, $\mathcal{R}_e = \mathcal{R}_p$ and $\mathcal{R}_e^c$ requires $\mathbf{L}_0(1,1)\mathbf{L}_0(2,1) \leq 0$. From this, $\mathcal{R}_\ell = \mathcal{R}_d \cup \mathcal{R}_e^c$ either requires $\mathbf{L}_0(1,1) > 0$ and $\mathbf{L}_0(2,1) > 0$ or requires $\mathbf{L}_0(1,1)\mathbf{L}_0(2,1) \leq 0$. Since $\mathcal{R}_p = \mathcal{R}_e$, the restriction $\mathcal{R}_p \cup \mathcal{R}_e^c = \mathcal{R}_s$ is trivial. If $\mathbf{L}_0(1,1)$ and $\mathbf{L}_0(2,1)$ have the same sign, then there is a labeling for the demand shock. If both are positive, then a positive shock increases demand, and if both are negative, then a positive shock decreases demand. However, if they do not have the same sign, then the labeling is not defined. This is reflected in the definition of $\mathcal{R}_\ell$.

There are other choices we could make for the label restriction. For instance, we could require either $\mathbf{L}_0(1,1) > 0$ or $\mathbf{L}_0(2,1) > 0$. It is instructive to compare these three labeling restrictions. Clearly, all three of these agree if the signs of the price and quantity responses are the same, but disagree if the signs are different. The label restriction $\mathcal{R}_\ell$ says the label is not defined if the signs are different, but both of the other choices define a label if the signs are different, though the labelings differ. Which of these three we choose does not matter in terms of the decomposition, but it does matter in terms of the implications of the label restriction.

Returning to the example from the introduction, consider the following cases.

Example 2. *Suppose that the first shock is a monetary policy shock, the first variable is the fed funds rate, the second output gap, and the third variable is inflation. Consider the restriction $\mathcal{R}$ that requires $\mathbf{L}_0(1,1)$ to be positive, so that the contemporaneous response to a positive monetary policy shock strictly increases the fed funds rate, and the restriction that the structural shocks have unit variance.*

The restriction $\mathcal{R}_d$ requires $\mathbf{L}_0(1,1) > 0$ and the restriction $\mathcal{R}_p$ requires $\mathbf{L}_0(1,1) \neq 0$ and $\mathbf{\Omega} = \mathbf{I}_3$. The restriction $\mathcal{R}_e$ requires $\mathbf{L}_0(1,1) \neq 0$, so its complement requires $\mathbf{L}_0(1,1) = 0$. The restriction $\mathcal{R}_\ell$ requires either $\mathbf{L}_0(1,1) > 0$ or $\mathbf{L}_0(1,1) = 0$, which can be more compactly written as $\mathbf{L}_0(1,1) \geq 0$. The restriction $\mathcal{R}_s$ requires either $\mathbf{L}_0(1,1) \neq 0$ and $\mathbf{\Omega} = \mathbf{I}_3$ or $\mathbf{L}_0(1,1) = 0$. This is an example where Equations (3)-(5) do not produce the simplest result. If $\tilde{\mathcal{R}}_\ell$ is the restriction that requires $\mathbf{L}_0(1,1) > 0$ and $\tilde{\mathcal{R}}_s$ is the restriction that requires $\mathbf{\Omega} = \mathbf{I}_3$, then it is easy to see that $\tilde{\mathcal{R}}_\ell$ is a label restriction, $\tilde{\mathcal{R}}_s$ is a scale restriction, and $\tilde{\mathcal{R}}_\ell \cap \tilde{\mathcal{R}}_s = \mathcal{R}$.

Example 3. *Suppose that the first shock is a monetary policy shock, the first variable is the fed funds rate, the second output gap, and the third variable is inflation. Consider the restriction $\mathcal{R}$ that requires $\mathbf{L}_0(1,1)$ to be positive, so that the contemporaneous response to a positive monetary policy shock strictly increases the fed funds rate, and the restriction that requires the diagonal elements of $\mathbf{L}_0$ to be 1.*

The restriction $\mathcal{R}$ is equal to the restriction that requires only $\mathbf{L}_0(i,i) = 1$, for all i. For this reason, it is tempting, but incorrect, to claim that this example consists of only a normalization of the scale. Let us decompose $\mathcal{R}$ using Equations (3)-(5). The restriction $\mathcal{R}_d$ requires $\mathbf{L}_0(i,i) > 0$, for all i, and the restriction $\mathcal{R}_p$ requires $\mathbf{L}_0(i,i)^2 = 1$, for all i. The restriction $\mathcal{R}_e$ requires $\mathbf{L}_0(i,i) \neq 0$, for all i, so its complement requires $\mathbf{L}_0(i,i) = 0$, for some i. The restriction $\mathcal{R}_\ell$ either requires $\mathbf{L}_0(i,i) > 0$, for all i, or requires $\mathbf{L}_0(i,i) = 0$, for some i. The restriction $\mathcal{R}_s$ either requires $\mathbf{L}_0(i,i)^2 = 1$, for all i, or requires $\mathbf{L}_0(i,i) = 0$, for some i. This is another example where Equations (3)-(5) do not produce the simplest result. If $\tilde{\mathcal{R}}_\ell$ is the label restriction that requires $\mathbf{L}_0(i,i) > 0$, for every i, and $\tilde{\mathcal{R}}_s$ is the scale restriction that requires $\mathbf{L}_0(i,i)^2 = 1$, then $\tilde{\mathcal{R}}_\ell \cap \tilde{\mathcal{R}}_s = \mathcal{R}$.

Now let us consider a more involved example.

Example 4. *Suppose that the first shock is a demand shock, the second shock is a supply shock, the first variable is price, and the second variable is quantity. Consider the restriction $\mathcal{R}$ that requires: (i) the ratio between $\mathbf{L}_0(1,1)$ and $\mathbf{L}_0(2,1)$ to be positive so that the supply curve is upward sloping; (ii) the determinant of $\mathbf{L}_0$ times $\mathbf{L}_0(2,1)^{-1}$ is negative so that a supply shock increases prices holding the quantity constant; (iii) the structural shocks have unit variance.*

While we could use Equations (3)-(5) to produce a decomposition, sometimes it is easier to directly work with the raw restrictions using what we have learned about economic, label, and scale restrictions, though such a decomposition could be different from the one produced by Equations (3)-(5). Condition (i) is equivalent to $\mathbf{L}_0(1,1)\mathbf{L}_0(2,1) > 0$, which is an economic restriction. Condition (ii) is equivalent to $\det(\mathbf{L}_0)\mathbf{L}_0(2,1) < 0$, which is a label restriction. Condition (iii) is $\mathbf{\Omega} = \mathbf{I}_2$, which is a scale restriction. We can write $\mathcal{R}$ as a combination of the economic restriction (i), the label restriction (ii), and the scale restriction (iii).

Consider now an example based on the identification strategy proposed by Arias, Caldara, and Rubio-Ramírez (2019) inspired by Christiano, Eichenbaum, and Evans (1996).

Example 5. *Suppose that the first shock is a monetary policy shock, the first variable is the fed funds rate, the second output gap, and the third variable is inflation. Consider the restriction* $\mathcal{R}$ *that requires: (i)* $c_1(\mathbf{L}_0) = \frac{\mathbf{L}_0(1,2)\mathbf{L}_0(3,3)-\mathbf{L}_0(1,3)\mathbf{L}_0(3,2)}{\mathbf{L}_0(2,2)\mathbf{L}_0(3,3)-\mathbf{L}_0(2,3)\mathbf{L}_0(3,2)} > 0$ *so that the fed funds rate reacts positively to the output gap; (ii)* $c_2(\mathbf{L}_0) = \frac{\mathbf{L}_0(1,3)\mathbf{L}_0(2,2)-\mathbf{L}_0(1,2)\mathbf{L}_0(2,3)}{\mathbf{L}_0(2,2)\mathbf{L}_0(3,3)-\mathbf{L}_0(2,3)\mathbf{L}_0(3,2)} > 0$ *so that the fed funds rate reacts positively to inflation; (iii)* $c_3(\mathbf{L}_0) = \frac{\det(\mathbf{L}_0)}{\mathbf{L}_0(2,2)\mathbf{L}_0(3,3)-\mathbf{L}_0(2,3)\mathbf{L}_0(3,2)} < 0$ *so that a positive monetary policy shock strictly increases the fed funds rate before the Fed reaction function is taken into account.*

It is easy to see that changing the scale or sign of a column of $\mathbf{L}_0$ will have no effect on $c_1(\mathbf{L}_0)$ or $c_2(\mathbf{L}_0)$. This implies that both (i) and (ii) are economic restrictions. However, changing the sign of the first column of $\mathbf{L}_0$ does change the sign of $c_3(\mathbf{L}_0)$, but changing the sign of any of the other columns of $\mathbf{L}_0$ will have no effect on the sign of $c_3(\mathbf{L}_0)$, nor will changing the scale of any of the columns of $\mathbf{L}_0$. So, (iii) is a label restriction. Thus, $\mathcal{R}$ is the combination of the economic restrictions $c_1(\mathbf{L}_0) > 0$ and $c_2(\mathbf{L}_0) > 0$, and the label restriction $c_3(\mathbf{L}_0) < 0$.

3 A Tale of Two Normalizations

As mentioned above, despite the issues highlighted with the unit effect normalization, it is convenient to have a normalization that allows the researcher to choose the impact of the structural shock since it facilitates direct estimation of impulse responses in the native units of the data, such as percentage points for interest rates, enhancing the interpretability of results for policy analysis. Additionally, pinning down either the variance of the structural shocks or their effect answers two different economic questions. To make things concrete, suppose that the price and quantity of some commodity are among the variables. Choosing the variance of the structural shocks allows us to answer questions of the following type.

Question 1. *What is the contemporaneous effect on the quantity of a one-standard-deviation positive demand shock?*

On the contrary, pinning down the effects of the structural shocks allows the researcher to answer questions of the following type.

Question 2. *What is the contemporaneous effect on the quantity of a demand shock that upon impact raises the price by one unit?*

These are interesting yet different economic questions: The first asks what the effect is of a statistically typical positive demand shock, regardless of whether that effect is economically plausible, and the second asks what the effect is of a demand shock with a typical positive effect on price, regardless of the probability of observing such a shock.

Having developed a framework to distinguish and classify scale, label, and economic restrictions, we can address scale normalization in a principled way and thus answer both questions above. As mentioned several times during the paper, there are two restrictions often used in the literature to uniquely determine $\mathbf{\Omega}$: the unit variance and the unit effect normalizations. While the unit variance normalization intends to answer questions of Type 1, the unit effect normalization has the objective of answering questions of Type 2. We define them below and show that while the unit variance normalization is a scale restriction, the unit effect is not. We also provide an alternative normalization that we call unit modulus normalization that shares objectives with the unit effect normalization but is itself a scale restriction.

Definition 5. *The unit variance normalization restricts the structural-form parameters so that* $\mathbf{\Omega} = \mathbf{I}_n$.

It is easy to see that the unit variance normalization satisfies both S1 and S2 and thus is a scale restriction. Before defining the unit effect normalization, we introduce some additional notation. For any $n \times n$ matrix $\mathbf{X}$, let $\text{diag}(\mathbf{X})$ be the $n \times n$ diagonal matrix with the diagonal of $\mathbf{X}$ along its diagonal.

Definition 6. *The unit effect normalization restricts* $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ *so that* $\text{diag}(\mathbf{L}_0) = \mathbf{I}_n$.

The unit effect normalization is not a scale restriction because it does not satisfy S1, though it satisfies S2. It will be a combination of a scale restriction, defined next, and a label restriction.

Definition 7. *The unit modulus normalization restricts* $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ *so that* $\text{diag}(\mathbf{L}_0)^2 = \mathbf{I}_n$.

It is easy to see that the unit modulus normalization satisfies both S1 and S2 and hence, it is a scale restriction. We refer to parameters satisfying these restrictions as unit modulus parameters, and to SVARs written in terms of those parameters as the unit modulus parameterization. The unit effect normalization is the combination of the unit modulus normalization and the label restriction $\text{diag}(\mathbf{L}_0) > 0$. More generally, one can define the $\boldsymbol{\kappa}$-modulus normalization as the restriction that requires $\text{diag}(\mathbf{L}_0)^2 = \boldsymbol{\kappa}^2$, the $\boldsymbol{\kappa}$-effect normalization as the restriction that requires $\text{diag}(\mathbf{L}_0) = \boldsymbol{\kappa}$, and the $\boldsymbol{\kappa}$-variance normalization as the restriction that requires $\boldsymbol{\Omega} = \boldsymbol{\kappa}$, for any fixed $\boldsymbol{\kappa} \in \mathcal{D}_n^+$. Notwithstanding, for ease of exposition we maintain the unit-modulus terminology unless explicitly stated otherwise. The unit modulus normalization provides a solution that allows the researcher to control the size of the shocks without assigning a label, thereby isolating the scale from interpretive assumptions. However, conducting Bayesian inference in SVARs under this parameterization raises new theoretical challenges, which the next section addresses.

4 Bayesian Inference

In this section, we develop theory and an algorithm for Bayesian inference under the unit modulus normalization. Let $\mathcal{B}$ denote the set of all $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ such that $\text{diag}(\mathbf{L}_0)^2 = \mathbf{I}_n$, and denote elements of this set by $(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$. We will consider priors with respect to the volume measure on $\mathcal{B}$ induced by Lebesgue measure. Under this parameterization, two features shape the analysis. First, a flat prior over $(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$ induces an improper posterior kernel, so prior elicitation plays a central role. Second, the two-step approach, which places a uniform prior over the orthogonal matrices before transforming to the unit modulus parameterization, induces a proper posterior that is not flat over observationally equivalent parameter values.

A Mapping and an Improper Posterior

The theory introduced in this section relies on a mapping between the unit modulus parameterization and the orthogonal reduced-form parameters. Let us begin by defining this mapping in two stages:

$$(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega}) \xrightarrow{f} (\mathbf{B}, \underbrace{\mathbf{K}_0\mathbf{\Omega}^{\frac{1}{2}}}_{\mathbf{L}_0}) \xrightarrow{\phi} (\mathbf{B}, \underbrace{\mathbf{L}_0\mathbf{L}_0'}_{\mathbf{\Sigma}}, \underbrace{r(\mathbf{L}_0\mathbf{L}_0')^{-1}\mathbf{L}_0}_{\mathbf{Q}}),$$

so that $(\phi \circ f)(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega}) = (\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$. Define $\mathcal{A}$ to be the set of all $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ such that all the diagonal elements of $r(\mathbf{\Sigma})\mathbf{Q}$ are non-zero. Over $\mathcal{A}$, we can define $(\phi \circ f)^{-1} = f^{-1} \circ \phi^{-1}$ by:

$$(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) \xrightarrow{\phi^{-1}} (\mathbf{B}, \underbrace{r(\mathbf{\Sigma})\mathbf{Q}}_{\mathbf{L}_0}) \xrightarrow{f^{-1}} (\mathbf{B}, \underbrace{\mathbf{L}_0(\mathrm{diag}(\mathbf{L}_0)^2)^{-\frac{1}{2}}}_{\mathbf{K}_0}, \underbrace{\mathrm{diag}(\mathbf{L}_0)^2}_{\mathbf{\Omega}}).$$

It follows from Lemma 4 in the Appendix that the volume element of this mapping is:

$$v_{(\phi\circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n-1)}{2}} |\det(\mathbf{\Sigma})|^{-\frac{1}{2}} |\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}.$$

Under a flat prior over $(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$, the induced posterior kernel over the orthogonal reduced-form parameterization is proportional to:

$$p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma}) v_{(\phi\circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}), \tag{6}$$

where $p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma}) = NIW(\mathbf{B}, \mathbf{\Sigma}; \hat{\nu}, \hat{\mathbf{\Phi}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Theta}})$, $\hat{\nu} = T - m - n - 1$, $\hat{\mathbf{\Theta}} = (\mathbf{X}_T'\mathbf{X}_T)^{-1}$, $\hat{\mathbf{\Psi}} = \hat{\mathbf{\Theta}}\mathbf{X}_T'\mathbf{Y}_T$, and $\hat{\mathbf{\Phi}} = \mathbf{Y}_T'\mathbf{Y}_T - \hat{\mathbf{\Psi}}'\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{\Psi}}$, with $\mathbf{Y}_T = [\mathbf{y}_1 \ \cdots \ \mathbf{y}_T]'$ and $\mathbf{X}_T = [\mathbf{x_1}, \ldots, \mathbf{x_T}]'$ the data. Equation (6) implies that, unlike under the unit variance normalization, a flat prior over the unit modulus parameterization does not yield a proper posterior. Consequently, prior elicitation under this parameterization requires special care. Proposition 9 formalizes this result:

Proposition 9. *For $n \geq 3$, $\int_{\mathcal{B}} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})\, d\mathbf{B}\, d\mathbf{K}_0\, d\mathbf{\Omega} = \infty$.*

Proof. See Appendix. □

The intuition behind Proposition 9 is that the impropriety comes from regions of the parameter space in which at least one diagonal entry of $\mathbf{\Omega}$ approaches zero. In those regions, at least one structural shock has variance arbitrarily close to zero. Because the complement of $\mathcal{A}$ has measure zero, we can study the posterior by integrating over all $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$. Therefore,

$$\int_{\mathcal{B}} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})\, d\mathbf{B}\, d\mathbf{K}_0\, d\mathbf{\Omega} = \int_{\mathbb{R}^{m\times n}\times\mathcal{S}_n^+\times\mathcal{O}_n} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma})\, v_{(\phi\circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})\, d\mathbf{B}\, d\mathbf{\Sigma}\, d\mathbf{Q}$$

$$= 2^{-\frac{n(n-1)}{2}} \int_{\mathbb{R}^{m\times n}\times\mathcal{S}_n^+} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma}) |\det(\mathbf{\Sigma})|^{-\frac{1}{2}} \left(\int_{\mathcal{O}_n} |\det(\operatorname{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}\, d\mathbf{Q} \right) d\mathbf{B}\, d\mathbf{\Sigma}.$$

If some diagonal entry of $\mathbf{\Omega}$ approaches zero, then the corresponding diagonal entry of $r(\mathbf{\Sigma})\mathbf{Q}$ also approaches zero. As a result, $|\det(\operatorname{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|$ approaches zero and the Jacobian term $|\det(\operatorname{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}$ diverges, for $n \geq 3$.

From Proposition 9, it follows that a prior over the unit-modulus parameters that is proportional to a normal inverse-Wishart density, $NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Theta})$, will also induce an improper posterior. Corollary 2 establishes this result.

Corollary 2. *For $n \geq 3$, $\int_{\mathcal{B}} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})\, NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Theta}), d\mathbf{B}\, d\mathbf{K}_0\, d\mathbf{\Omega} = \infty$.*

Proof. Follows directly from the fact that $p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})\, NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Theta})$ and $p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$ have the same functional form. □

The Prior Induced by the Two-Step Approach

Before introducing our proposed priors over the unit modulus normalization, we analyze the prior over this normalization induced by the two-step approach employed by Baumeister and Hamilton (2018). They start from a uniform-normal-inverse-Wishart prior over the orthogonal reduced-form parameters, denoted by $UNIW(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}})$, and then map those draws into the unit modulus parameterization through $(\phi \circ f)^{-1}$.[3] Then, by a change of variables, the implied prior over the unit modulus normalization is proportional to:

$$NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}}) v_{\phi\circ f}(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega}). \tag{7}$$

[3]This prior specification allows for the weak prior defined by Uhlig (2005) and used by Baumeister and Hamilton (2018) in their application.

where $v_{\phi\circ f}(\mathbf{B},\mathbf{K}_0,\mathbf{\Omega}) = \left(v_{(\phi\circ f)^{-1}}((\phi\circ f)(\mathbf{B},\mathbf{K}_0,\mathbf{\Omega}))\right)^{-1}$. Although it is straightforward to show that the posterior induced by Equation (7) is proper, the prior is not constant over observationally equivalent parameters. Any parameter observationally equivalent to $(\mathbf{B},\mathbf{K}_0,\mathbf{\Omega})$ can be written as $(\mathbf{B},\mathbf{K}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}},\mathbf{\Omega}\mathbf{D})$, where $\mathbf{Q}\in\mathcal{O}_n$ and $\mathbf{D} = \mathrm{diag}(\mathbf{K}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Omega}^{-\frac{1}{2}})^2\in\mathcal{D}_n^+$. There always exist a $\mathbf{Q}$ such that $\mathbf{D}\neq\mathbf{I}_n$. Since the ratio between $v_{\phi\circ f}(\mathbf{B},\mathbf{K}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}},\mathbf{\Omega}\mathbf{D})$ and $v_{\phi\circ f}(\mathbf{B},\mathbf{K}_0,\mathbf{\Omega})$ is $|\det(\mathbf{D})|^{\frac{n-2}{2}}$, the prior induced by the two-step approach tilts mass across observationally equivalent unit modulus parameters.

Because the two-step approach induces a prior over the unit modulus parameterization that is not constant over observationally equivalent parameters, it is desirable to develop priors that are flat across observationally equivalent structural parameters under the unit modulus normalization. The challenge is that the density in Equation (6) is improper, so obtaining such priors while preserving a proper posterior is not straightforward. Our solution is to consider truncated priors of two types. The first is a prior that is flat across observationally equivalent structural parameters allowing researchers to encode familiar beliefs over reduced-form parameters, such as a flat prior as in Uhlig (2005) or Minnesota-style beliefs. The second is a truncated flat prior directly imposed over the unit modulus parameterization. This latter prior does not inherit the features that made Minnesota-style priors popular, such as favoring persistence, but it may be attractive to researchers interested in robustness to prior assumptions.

Uniform Prior Over Observationally Equivalent Parameters

In this subsection, we describe a prior that is constant across observationally equivalent points in the unit modulus parameterization while allowing flexibility across non-equivalent regions of the parameter space. While each researcher is entitled to their own prior beliefs, uniform priors over observationally equivalent structural parameters have the advantage of disentangling the problem of causal inference from the estimation problem (Arias, Rubio-Ramirez and Waggoner, 2025). Proposition 3 in Arias, Rubio-Ramirez, and Waggoner (2025) implies that if the prior over the orthogonal reduced-form parameters is proportional to $NIW(\mathbf{B},\mathbf{\Sigma};\bar{\nu},\bar{\mathbf{\Phi}},\bar{\mathbf{\Psi}},\bar{\mathbf{\Theta}})v_{(\phi\circ f)^{-1}}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$, then the induced prior over the unit modulus parameterization is proportional to $NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}})$. Consequently, it assigns the same density to observationally equivalent $(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$. However, Corollary 2 implies that this prior still leads to an improper posterior, as the posterior density diverges as $\mathbf{\Omega}$ approaches zero. To obtain a proper posterior, we truncate away from values of $\mathbf{\Omega}$ that are too close to zero. Proposition 10 establishes this result.

Proposition 10. *A prior over $(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$ of the form*

$$p(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega}) \propto \prod_{i=1}^{n} [\omega_i > \underline{\ell}]\, NIW(\mathbf{B}, \mathbf{K}_0\mathbf{\Omega}\mathbf{K}_0'; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}}),$$

where ω_i is the the i^{th} diagonal element of $\mathbf{\Omega}$, induces a proper posterior under the unit modulus parameterization.

Proof. See Appendix. □

The indicator function $[\omega_i > \underline{\ell}]$ is equal to 1 when $\omega_i > \underline{\ell}$ and zero otherwise and $\underline{\ell} \in \mathbb{R}_+$, which denotes the positive reals. This prior can be interpreted as a truncated version of a prior that is otherwise flat across observationally equivalent unit modulus parameters. Economically, the threshold $\underline{\ell}$ rules out shocks whose one-standard-deviation effects on the variables of interest are too small to be meaningful. As $\underline{\ell} \to 0$, the posterior approaches the region that generates impropriety. As $\underline{\ell}$ increases, however, one risks excluding shocks with small but still economically relevant standard deviations. Therefore, the choice of $\underline{\ell}$ must be application specific. Importantly, the constraint $\prod_{i=1}^{n}[\omega_i > \underline{\ell}] = 1$ defines a set of positive measure with respect to the volume measure on $\mathcal{B}$. For easy exposition, we have considered a common bound $\underline{\ell}$, but it is easy to extend our results to consider individual bounds $\underline{\ell}_i \in \mathbb{R}_+$ for $i = 1, \ldots, n$.

Uniform Prior

Researchers who wish to work with a flat prior directly over $(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$ under the unit modulus normalization can do so by imposing the same lower-bound restriction on the standard deviations of the structural shocks described above. Proposition 11 shows that this truncation yields a proper posterior.

Proposition 11. *A prior over* $(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$ *of the form*

$$p(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega}) \propto \prod_{i=1}^{n} [\omega_i > \underline{\ell}]$$

induces a proper posterior under the unit modulus parameterization.

Proof. The proof follows from the same argument as Proposition 10, combined with the fact that the remaining kernel over $(\mathbf{B}, \boldsymbol{\Sigma})$ is proportional to $p(\mathbf{Y}_T \mid \mathbf{B}, \boldsymbol{\Sigma})|\det(\boldsymbol{\Sigma})|^{-\frac{1}{2}}$, which has a proper normal-inverse-Wishart form. $\square$

This prior can be interpreted as a truncated flat prior over the unit modulus parameterization. It preserves flatness over the admissible region of the parameter space while ruling out draws with arbitrarily small structural-shock standard deviations, which are exactly the draws responsible for the impropriety result in Proposition 9. As before, the constraint $\prod_{i=1}^{n} [\omega_i > \underline{\ell}] = 1$ defines a set of positive measure with respect to the volume measure on $\mathcal{B}$.

An Algorithm

In this section, we exploit the central idea of the elliptical-slice-within-Gibbs sampler proposed by Arias, Rubio-Ramírez, Rudolf, and Shin (2026) to conduct inference over the structural-form parameters under the unit-modulus normalization; adapting it to the $\boldsymbol{\kappa}$-modulus normalization is straightforward. This approach is useful here because while as shown below the posteriors implied by our proposed priors do not permit direct sampling, the posterior kernel accommodates the elliptical slice sampling. The algorithm works for any prior over $(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$ that induces a proper posterior, such as the ones discussed above. The basic idea is to draw from the posterior distribution over the orthogonal reduced-form parameters, map those draws back to $\mathcal{B}$, and retain the draws that satisfy the restrictions. Importantly, differently than Arias, Rubio-Ramírez, Rudolf, and Shin (2026) we use a joint implementation that avoids the need of the Gibbs sampler.

Consider any prior over $(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$ whose support is a set of positive measure with respect to the volume measure on $\mathcal{B}$. Let us derive the induced prior over the orthogonal reduced-form parameterization. The prior over $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ induced by the prior $p(\mathbf{B}, \mathbf{K}_0, \boldsymbol{\Omega})$ is

$$p(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = p((\phi \circ f)^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))v_{(\phi \circ f)^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}). \tag{8}$$

We will combine the prior in Equation (8) with the likelihood, which is proportional to a normal-inverse Wishart density $NIW(\mathbf{B}, \boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}})$. The density $NIW(\mathbf{B}, \boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}})$ can be written as $IW(\boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}})N(\mathbf{B}; \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}}, \boldsymbol{\Sigma})$, where $IW(\boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}})$ is the inverse Wishart density with parameters $\hat{\nu}$ and $\hat{\boldsymbol{\Phi}}$ and $N(\mathbf{B}; \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}}, \boldsymbol{\Sigma})$ is a matrix normal density which implies that $\mathrm{vec}(\mathbf{B})$ follows a normal density with mean $\mathrm{vec}(\hat{\boldsymbol{\Psi}})$ and variance $\boldsymbol{\Sigma} \otimes \hat{\boldsymbol{\Theta}}$. The resulting posterior, $p(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q} \mid \mathbf{Y}_T)$, is proportional to

$$NIW(\mathbf{B}, \boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}})p((\phi \circ f)^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}))v_{(\phi \circ f)^{-1}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}). \tag{9}$$

Our objective will be to draw from the posterior of the orthogonal reduced-form parameters defined in Equation (9). Rather than partitioning this posterior into conditional densities and updating $\mathbf{B}$, $\boldsymbol{\Sigma}$, and $\mathbf{Q}$ in separate Gibbs blocks, we update them jointly. To do so, we express $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ as a deterministic transformation of three independent Gaussian matrices whose joint law pushes forward to $UNIW(\mathbf{B}, \boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}})$. Specifically, let $\mathbf{X} \sim N(\mathbf{0}_{n\times n}, \mathbf{I}_n, \mathbf{I}_n)$, $\mathbf{R} \sim N(\mathbf{0}_{n\times\hat{\nu}}, \hat{\boldsymbol{\Phi}}^{-1}, \mathbf{I}_{\hat{\nu}})$, and $\mathbf{U} \sim N(\mathbf{0}_{m\times n}, \mathbf{I}_m, \mathbf{I}_n)$ be mutually independent, and define the mapping $\mathbf{T}$ by

$$\mathbf{Q} = \gamma(\mathbf{X}), \qquad \boldsymbol{\Sigma} = \varsigma(\mathbf{R}) = (\mathbf{R}\mathbf{R}')^{-1}, \qquad \mathbf{B} = \hat{\boldsymbol{\Psi}} + \mathbf{L}_{\hat{\boldsymbol{\Theta}}}\,\mathbf{U}\,r(\boldsymbol{\Sigma})', \tag{10}$$

where γ is the QR-decomposition mapping, $r(\boldsymbol{\Sigma})$ is the lower-triangular Cholesky factor of $\boldsymbol{\Sigma}$ with positive diagonal, and $\mathbf{L}_{\hat{\boldsymbol{\Theta}}}$ is a fixed matrix satisfying $\mathbf{L}_{\hat{\boldsymbol{\Theta}}}\mathbf{L}'_{\hat{\boldsymbol{\Theta}}} = \hat{\boldsymbol{\Theta}}$. By construction, $\mathbf{Q}$ is uniformly (Haar) distributed, $\boldsymbol{\Sigma} \sim IW(\hat{\nu}, \hat{\boldsymbol{\Phi}})$, and, conditional on $\boldsymbol{\Sigma}$, $\mathrm{vec}(\mathbf{B})$ is normal with mean $\mathrm{vec}(\hat{\boldsymbol{\Psi}})$ and variance $\boldsymbol{\Sigma} \otimes \hat{\boldsymbol{\Theta}}$; that is, $\mathbf{B} \mid \boldsymbol{\Sigma} \sim N(\mathbf{B}; \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Theta}}, \boldsymbol{\Sigma})$. The whitening of $\mathbf{B}$ is the key step: it removes the dependence of the covariance of $\mathbf{B}$ on $\boldsymbol{\Sigma}$, so that the three latent blocks are jointly Gaussian under a single fixed reference measure. Collecting the latent blocks into $\mathbf{Z} = (\mathbf{X}, \mathbf{R}, \mathbf{U})$, the posterior in Equation (9) becomes, in the latent coordinates,

$$p_{\mathbf{Z}}(\mathbf{Z} \mid \mathbf{Y}_T) \propto p((\phi \circ f)^{-1}(\mathbf{T}(\mathbf{Z})))\, v_{(\phi\circ f)^{-1}}(\mathbf{T}(\mathbf{Z})) \\ \times N(\mathbf{X}; \mathbf{0}_{n\times n}, \mathbf{I}_n, \mathbf{I}_n)\, N(\mathbf{R}; \mathbf{0}_{n\times\hat{\nu}}, \hat{\mathbf{\Phi}}^{-1}, \mathbf{I}_{\hat{\nu}})\, N(\mathbf{U}; \mathbf{0}_{m\times n}, \mathbf{I}_m, \mathbf{I}_n). \quad (11)$$

This is a product of a fixed Gaussian reference law and the non-Gaussian factor $L(\mathbf{Z}) = p((\phi \circ f)^{-1}(\mathbf{T}(\mathbf{Z})))v_{(\phi\circ f)^{-1}}(\mathbf{T}(\mathbf{Z}))$. Note that the NIW likelihood, including the matrix-normal density of $\mathbf{B}$ given $\mathbf{\Sigma}$, is fully absorbed into the Gaussian reference and therefore does not appear in $L(\mathbf{Z})$. We can thus sample $\mathbf{Z}$ with a single elliptical slice move, as shown in Algorithm 1.

Algorithm 1. *This algorithm draws samples from a joint elliptical slice sampler targeting the distribution of interest determined by Equation (11).*

1. *Set $i = 1$ and assign initial values $\mathbf{Z}^0 = (\mathbf{X}^0, \mathbf{R}^0, \mathbf{U}^0)$ such that $\mathbf{T}(\mathbf{Z}^0)$ satisfies the restrictions.*
2. *Draw the elliptical direction $\mathbf{V} = (\mathbf{V}_{\mathbf{X}}, \mathbf{V}_{\mathbf{R}}, \mathbf{V}_{\mathbf{U}})$ from the reference law, i.e., $\mathbf{V}_{\mathbf{X}} \sim N(\mathbf{0}_{n\times n}, \mathbf{I}_n, \mathbf{I}_n)$, $\mathbf{V}_{\mathbf{R}} \sim N(\mathbf{0}_{n\times\hat{\nu}}, \hat{\mathbf{\Phi}}^{-1}, \mathbf{I}_{\hat{\nu}})$, and $\mathbf{V}_{\mathbf{U}} \sim N(\mathbf{0}_{m\times n}, \mathbf{I}_m, \mathbf{I}_n)$.*
3. *Using elliptical slice sampling, draw $\mathbf{Z}^i$ on the ellipse*

$$\mathbf{Z}(\theta) = \mathbf{Z}^{i-1}\cos\theta + \mathbf{V}\sin\theta, \qquad \theta \in [0, 2\pi], \quad (12)$$

   with target slice factor $L(\mathbf{Z}) = p((\phi \circ f)^{-1}(\mathbf{T}(\mathbf{Z})))v_{(\phi\circ f)^{-1}}(\mathbf{T}(\mathbf{Z}))$.
4. *Set $(\mathbf{B}^i, \mathbf{\Sigma}^i, \mathbf{Q}^i) = \mathbf{T}(\mathbf{Z}^i)$.*
5. *If $i < I$, let $i = i + 1$ and return to Step 2.*

In Step 3, elliptical slice sampling proceeds in the standard way: one draws a slice level $\log u + \log L(\mathbf{Z}^{i-1})$ with $u \sim U(0, 1)$, samples an initial angle $\theta \sim U(0, 2\pi)$ together with an initial bracket, and accepts the first θ for which $\log L(\mathbf{Z}(\theta))$ exceeds the slice level, shrinking the bracket toward the origin otherwise. Because the reference measure in Equation (11) is

Gaussian and the factor $L(\mathbf{Z})$ enters only through the slice level, the elliptical slice move leaves $p_{\mathbf{Z}}$ invariant, and the parameter-dependent Jacobians implied by the map $\mathbf{T}$ need never be evaluated: they are precisely what makes the latent Gaussian reference push forward to the $NIW \times$ Haar posterior. Mapping $\mathbf{Z}^i$ through $\mathbf{T}$ in Step 4 then yields a draw $(\mathbf{B}^i, \mathbf{\Sigma}^i, \mathbf{Q}^i)$ from Equation (9). It is straightforward to adapt the code to retain only the draws that satisfy sign restrictions of the type $[\boldsymbol{S}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) > 0]$. In this case, one needs only to include the indicator $[\boldsymbol{S}(\mathbf{T}(\mathbf{Z})) > 0]$ in the slice factor $L(\mathbf{Z})$ evaluated in Step 3; the current state always satisfies the restrictions, so $\theta = 0$ is feasible and the shrinkage procedure terminates. Also, the algorithm has been written in terms of the unit modulus normalization, but it can be easily adapted to any $\boldsymbol{\kappa}$-modulus normalization by changing the function f. In order to fix ideas, it should be clear that when using the observationally equivalent prior defined above, we have that:

$$p((\phi \circ f)^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{(\phi \circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \iota_{\underline{\ell}}(\mathbf{\Sigma}, \mathbf{Q}) NIW(\mathbf{B}, \mathbf{\Sigma}; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}}) v_{(\phi \circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}),$$

where $\iota_{\underline{\ell}}(\mathbf{\Sigma}, \mathbf{Q}) = \prod_{i=1}^{n} [(e_i' r(\mathbf{\Sigma}) \mathbf{Q} e_i)^2 > \underline{\ell}]$, while we have that:

$$p((\phi \circ f)^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{(\phi \circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \iota_{\underline{\ell}}(\mathbf{\Sigma}, \mathbf{Q}) v_{(\phi \circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}),$$

when using the uniform prior defined above. Finally, when using the prior induced by the two-step approach we have:

$$p((\phi \circ f)^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{(\phi \circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = NIW(\mathbf{B}, \mathbf{\Sigma}; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}})$$

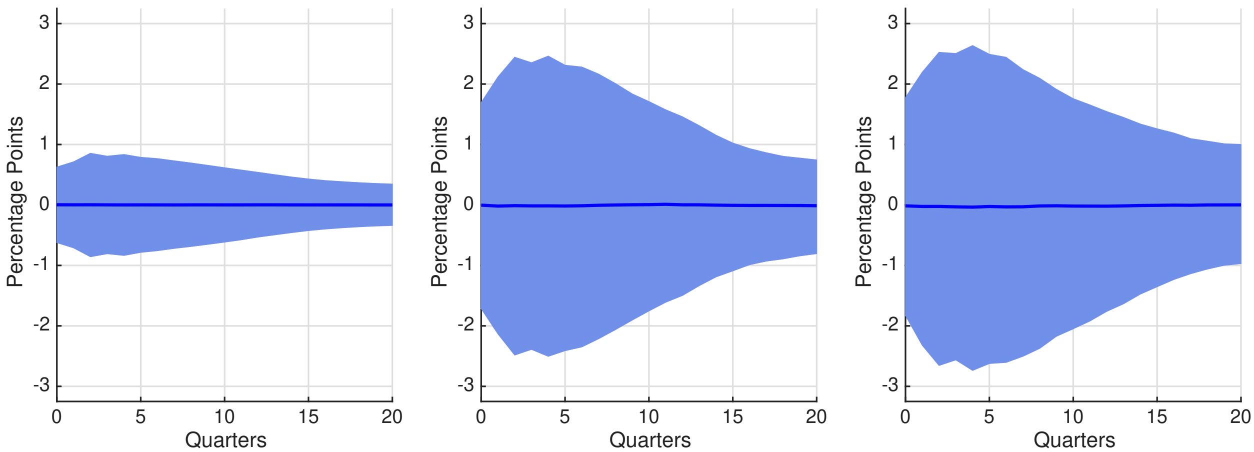
and Algorithm 1 targets the same posterior as the two-step approach used by Baumeister and Hamilton (2018).

5 Illustration

This section has three objectives. First, it illustrates that Algorithm 1 can be used to draw from the posteriors implied by any of the three priors described in Section 4 under the $\boldsymbol{\kappa}$-modulus normalization, with $\boldsymbol{\kappa} = 0.25$, so that a monetary policy shock causes a ± 25-basis-point change in the fed funds rate upon impact. Figure 4 shows that the three priors imply a symmetric response of output to a unit monetary policy shock under the $\boldsymbol{\kappa}$-modulus normalization. Hence, Baumeister and Hamilton's (2018) concern is not a consequence of the prior. Second, the section reinforces that the concern arises from a sign restriction on the impulse response of the fed funds rate implied by the $\boldsymbol{\kappa}$-effect normalization. In this case, we add a sign restriction on the response of the fed funds rate to a monetary policy shock so that a unit monetary policy shock causes a 25-basis-point increase in the fed funds rate upon impact. Figure 5 shows that the sign restriction is responsible for the concern regardless of the prior used. Finally, the section shows the consequences of using priors that are not uniform over observationally equivalent parameters. Figures 4 and 5 show that such priors produce narrower posterior probability bands, and Figure 6 explains why.

We use the same model specification as Baumeister and Hamilton (2018): a 3-variable SVAR of the U.S. economy featuring the output gap, inflation, and the fed funds rate. The output gap is measured as the log difference between observed and potential real GDP estimated by the Congressional Budget Office, and inflation is measured as the log difference of the PCE deflator. The SVAR is specified at quarterly frequency and estimated over the sample 1986:Q1 to 2008:Q4. The model includes 4 lags and a constant. For both the prior implied by the two-step approach and the observationally equivalent prior, we specify $NIW(\mathbf{B}, \mathbf{\Sigma}; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}})$ as the weak prior of Uhlig (2005). When using the observationally equivalent prior and the uniform prior, we restrict the impact of a one-standard-deviation shock on the fed funds rate to be at least 1 basis point.

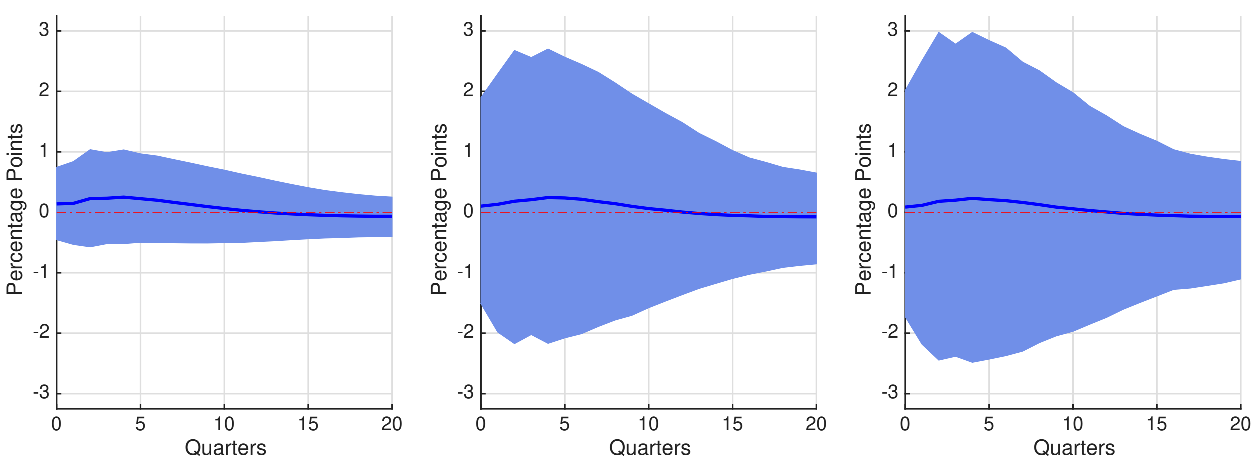
Figure 4 compares posterior impulse responses of output to a monetary policy shock under the $\boldsymbol{\kappa}$-modulus normalization. Panel (a) reports the posterior under the two-step approach prior, panel (b) reports the posterior under the observationally equivalent prior, and panel (c) reports the posterior under the uniform prior. In each panel, the solid blue lines represent the



(a) Two-step Approach (b) Observationally Equivalent (c) Uniform

Figure 4: Impulse Response of the Output to a Monetary Policy Shock. The solid blue lines represent the point-wise posterior medians, and the blue-shaded areas represent the 68 percent equal-tailed point-wise probability bands.

point-wise posterior medians, and the blue-shaded areas represent the 68 percent equal-tailed point-wise probability bands. The point-wise posterior medians are symmetric around zero, while the bands are wider under priors that are uniform over observationally equivalent parameters.



(a) Two-step Approach (b) Observationally Equivalent (c) Uniform

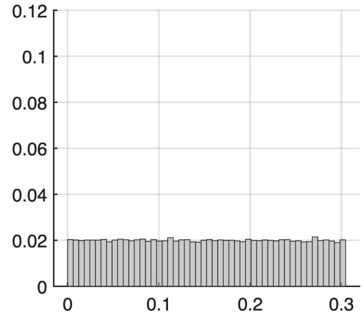
Figure 5: Impulse Response of the Output to a Monetary Policy Shock. The solid blue lines represent the point-wise posterior medians, and the blue-shaded areas represent the 68 percent equal-tailed point-wise probability bands.

Figure 5 compares posterior impulse responses of output to a monetary policy shock under the $\boldsymbol{\kappa}$-effect normalization. Panel (a) reports the posterior under the two-step approach prior, panel (b) reports the posterior under the observationally equivalent prior, and panel (c) reports the posterior under the uniform prior. In each panel, the solid blue lines represent the point-wise posterior medians, and the blue-shaded areas represent the 68 percent equal-tailed point-wise probability bands. Unlike under the $\boldsymbol{\kappa}$-modulus normalization, the point-wise posterior medians are no longer symmetric once the sign associated with the $\boldsymbol{\kappa}$-effect normalization is introduced. Thus, Figures 4 and 5 show that the output gap is more likely to increase only after imposing this sign restriction.

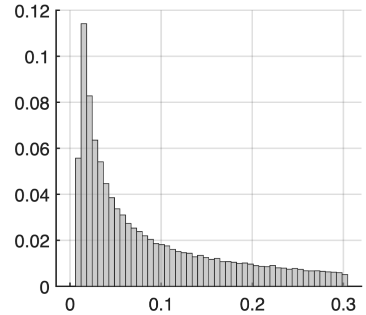
To shed light on the differences between the posterior probability bands, we fix the reduced-form parameters at the maximum likelihood estimates and draw from the posterior over the orthogonal matrices conditional on these reduced-form parameters. Because the likelihood is unaffected by the orthogonal matrices conditional on the reduced-form parameters, this exercise isolates the prior over the orthogonal matrices. Then, we inspect the implied distribution over $\mathbf{L}_0$. Recall that while the prior over the orthogonal matrices in the two-step approach is uniform, it is proportional to the volume element under both the observationally equivalent prior and the uniform prior. Thus, we should expect the implied distributions over $\mathbf{L}_0$ to differ. For our purposes, it suffices to focus on $\mathbf{L}_0(\text{ffr,mp})$.

Figure 6 reports the implied distribution of $\mathbf{L}_0(\text{ffr,mp})$ under the three priors. Across the panels, both the observationally equivalent prior and the uniform prior put more probability mass on values of $\mathbf{L}_0(\text{ffr,mp})$ close to zero relative to the two-step approach. When these draws are mapped to the unit modulus normalization via $\mathbf{K}_0 = \mathbf{L}_0(\text{diag}(\mathbf{L}_0)^2)^{-\frac{1}{2}}$, draws with small values of $\mathbf{L}_0(\text{ffr,mp})$ generate large normalized responses for the other variables in the column of $\mathbf{L}_0$ associated with the monetary-policy shock. This is why the observationally equivalent and uniform priors produce wider probability bands.

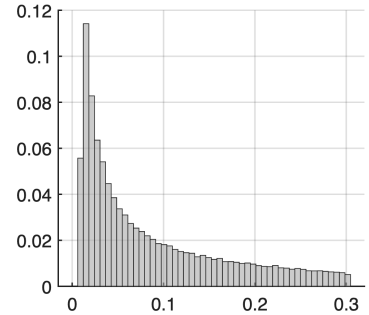
Let us also highlight that the results from both the observationally equivalent prior and the uniform prior are sensitive to the choice of $\underline{\ell}$. As discussed above, from an economic perspective, the threshold $\underline{\ell}$ rules out shocks whose one-standard-deviation impact effects on the variables of interest are too small to be meaningful. As $\underline{\ell}$ increases, the posterior probability bands narrow; however, one risks excluding shocks with small but still economically relevant standard deviations. For example, when we restrict the impact of a one-standard-deviation shock on the fed funds rate to be at least 10 basis points instead of 1 basis point,



(a) Two-step Approach



(b) Observationally Equivalent



(c) Uniform

Figure 6: Implied Distribution over $\mathbf{L}_0$(ffr,mp)

the probability bands under the observationally equivalent prior and the uniform prior are very similar to those implied by the two-step approach. A corollary of this result is that our theory allows researchers to explicitly confront the problem of conducting Bayesian inference under the unit modulus parameterization.

6 Conclusion

This paper revisits the influential critique of Baumeister and Hamilton (2018), who argue that conventional Bayesian methods for sign-identified SVARs deliver economically implausible results even when no sign restrictions are imposed. We show that the troubling pattern is driven not by the uniform prior over orthogonal matrices, as previously suggested, but by an unacknowledged sign restriction hidden inside the unit-effect normalization that interacts with the reduced-form covariance structure. Building on this observation, we develop a general theory of restrictions for structural VARs. Its centerpiece is a decomposition result: any restriction can be uniquely separated into an economic, a label, and a scale component. Guided by this framework, we introduce the unit modulus normalization—a pure scale restriction that pins down the impact size of structural shocks without committing to a label. We characterize priors that yield a proper posterior under this normalization and develop an elliptical-slice-sampling algorithm to sample from it. Applied to the SVAR studied by Baumeister and Hamilton, our method makes the troubling positive output response to a contractionary monetary policy shock disappear.

Appendix

We first prove the following lemma that will be useful in the proofs of several of our results. Let $\psi(\mathbf{L}_0) = (\mathbf{L}_0\mathbf{L}_0', r(\mathbf{L}_0\mathbf{L}_0')^{-1}\mathbf{L}_0) = (\mathbf{\Sigma}, \mathbf{Q})$, so that $\phi(\mathbf{B}, \mathbf{L}_0) = (\mathbf{B}, \psi(\mathbf{L}_0)) = (\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.

Lemma 1. *The volume elements of ϕ and ψ are $v_\phi(\mathbf{B}, \mathbf{L}_0) = v_\psi(\mathbf{L}_0) = 2^{\frac{n(n+1)}{2}}|\det(\mathbf{L}_0)|$.*

Proof. Since $\phi(\mathbf{B}, \mathbf{L}_0) = (\mathbf{B}, \psi(\mathbf{L}_0))$, the volume element of ψ will be equal to the volume element of ϕ. We can factor the mapping ϕ as

$$(\mathbf{B}, \mathbf{L}_0) \longrightarrow (\underbrace{\mathbf{B}(\mathbf{L}_0^{-1})'}_{\mathbf{A}_+}, \underbrace{(\mathbf{L}_0^{-1})'}_{\mathbf{A}_0}) \longrightarrow (\underbrace{\mathbf{A}_+\mathbf{A}_0^{-1}}_{\mathbf{B}}, \underbrace{(\mathbf{A}_0\mathbf{A}_0')^{-1}}_{\mathbf{\Sigma}}, \underbrace{r((\mathbf{A}_0\mathbf{A}_0')^{-1})'\mathbf{A}_0}_{\mathbf{Q}}).$$

From Proposition 1 in Arias, Rubio-Ramírez, and Waggoner (2018), the volume element for the second mapping is $2^{\frac{n(n+1)}{2}}|\det(\mathbf{A}_0)|^{-(2n+(np+1)+1)}$. Note that the function h in that paper is equal to the transpose of the function r. The derivative of the first mapping is

$$\begin{bmatrix} \mathbf{L}_0^{-1} \otimes \mathbf{I}_{np+1} & -(\mathbf{I}_n \otimes \mathbf{B})\mathbf{K}_{n,n}(\mathbf{L}_0' \otimes \mathbf{L}_0)^{-1} \\ 0 & -\mathbf{K}_{n,n}(\mathbf{L}_0' \otimes \mathbf{L}_0)^{-1} \end{bmatrix}$$

where $\mathbf{K}_{n,n}$ is the $n^2 \times n^2$ commutation matrix, so the volume element of the first mapping is $|\det(\mathbf{L}_0)|^{-(np+1)-2n}$. Thus, the volume element of either ϕ or ψ is $2^{\frac{n(n+1)}{2}}|\det(\mathbf{L}_0)|$. $\square$

Proof of Proposition 1. We first prove the following lemma.

Lemma 2. *Let $x \in \mathbb{R}$, $\mathbf{y} \in \mathbb{R}^{n-1}$, and $\mathbf{z} \in \mathcal{D}(x, \mathbf{y}) \subset \mathbb{R}^{n(n-1)}$ such that $\mathcal{D}(x, \mathbf{y}) = \mathcal{D}(-x, -\mathbf{y})$. If $p(x, \mathbf{y}, \mathbf{z})$ is a density satisfying $p(x, \mathbf{y}, \mathbf{z}) = p(-x, -\mathbf{y}, \mathbf{z})$, then the marginal distribution of x is symmetric about the origin.*

Proof. Because $\mathcal{D}(x, \mathbf{y}) = \mathcal{D}(-x, -\mathbf{y})$ and $p(x, \mathbf{y}, \mathbf{z}) = p(-x, -\mathbf{y}, \mathbf{z})$, the marginal density satisfies $p(x) = \int_{\mathbf{y}\in\mathbb{R}^{n-1}} \int_{\mathbf{z}\in\mathcal{D}(x,\mathbf{y})} p(x, \mathbf{y}, \mathbf{z}) = \int_{\mathbf{y}\in\mathbb{R}^{n-1}} \int_{\mathbf{z}\in\mathcal{D}(-x,-\mathbf{y})} p(-x, -\mathbf{y}, \mathbf{z})d\mathbf{z} = p(-x)$. $\square$

Let $p(\mathbf{\Sigma}, \mathbf{Q}) = p(\mathbf{\Sigma})p(\mathbf{Q} \mid \mathbf{\Sigma})$ be any prior density satisfying $p(\mathbf{Q} \mid \mathbf{\Sigma}) = p(\mathbf{Q}\mathbf{P}_j^- \mid \mathbf{\Sigma})$, which is equivalent to $p(\mathbf{\Sigma}, \mathbf{Q}) = p(\mathbf{\Sigma}, \mathbf{Q}\mathbf{P}_j^-)$. The function ψ induces a prior density over $\mathbf{L}_0$ given by $p(\mathbf{L}_0) = v_\psi(\mathbf{L}_0)p(\psi(\mathbf{L}_0))$, where, by Lemma 1, $v_\psi(\mathbf{L}_0) = 2^{\frac{n(n+1)}{2}}|\det(\mathbf{L}_0)|$. If $\mathbf{Y}$ is the data, then the posterior density is proportional to $p(\mathbf{L}_0)p(\mathbf{Y} \mid \mathbf{\Sigma} = \mathbf{L}_0\mathbf{L}_0')$. From these representations, it is easy to see that $p(\mathbf{L}_0\mathbf{P}_j^-) = p(\mathbf{L}_0)$ and $p(\mathbf{L}_0\mathbf{P}_j^- \mid \mathbf{Y}) = p(\mathbf{L}_0 \mid \mathbf{Y})$. So, we can apply Lemma 2 with $x = \mathbf{L}_0(i,j)$, with $\mathbf{y}$ consisting of all the elements of the j^{th} column of $\mathbf{L}_0$ other than $\mathbf{L}_0(i,j)$, and with $\mathbf{z}$ consisting of all the elements in $\mathbf{L}_0$ not in the j^{th} column. □

Proof of Proposition 2. Since we can reorder the variables and shocks, we can assume without loss of generality that the first variable is the fed fund rate, the second variable is the output gap, and the first shock is the monetary policy shock. To have a more compact notation, let $\mathbf{L}_0(i,j) = \ell_{ij}$, $\mathbf{\Sigma}(i,j) = \sigma_{ij}$, $r(\mathbf{\Sigma})(i,j) = r_{ij}$, $\mathbf{Q}(i,j) = q_{ij}$, and $\mathbf{\Phi}(i,j) = \phi_{ij}$. We prove each part separately below. □

Proof of Proposition 2 (i). Assume that $p(\mathbf{Q} \mid \mathbf{\Sigma})$ is uniform and $\mathbf{\Sigma} \sim IW(\nu, \mathbf{\Phi})$. A direct computation gives $r_{11} = \sigma_{11}^{\frac{1}{2}}$, $r_{22} = (\sigma_{22} - \sigma_{21}^2\sigma_{11}^{-1})^{\frac{1}{2}}$, and $r_{21} = \sigma_{21}\sigma_{11}^{-\frac{1}{2}}$, where all square roots are non-negative. Because $\ell_{11} = r_{11}q_{11} = \sigma_{11}^{\frac{1}{2}}q_{11}$ and

$$\ell_{21} = r_{21}q_{11} + r_{22}q_{21} = \sigma_{11}^{-\frac{1}{2}}(\sigma_{21}q_{11} + (\sigma_{11}\sigma_{22} - \sigma_{21}^2)^{\frac{1}{2}}q_{21}),$$

we analyze the marginal distribution over $(\sigma_{11}, \sigma_{22}, \sigma_{21}, q_{11}, q_{21})$, conditional on $\ell_{11} > 0$. Let $\mathcal{W}$ be the set of all $\mathbf{w} = (\sigma_{11}, \sigma_{22}, \sigma_{21}, q_{11}, q_{21})$ such that $0 < \sigma_{11}$, $0 < \sigma_{22}$, $\sigma_{21}^2 < \sigma_{11}\sigma_{22}$, $q_{11}^2 + q_{21}^2 \leq 1$, and $0 < q_{11}$. The first three inequalities hold because $\mathbf{\Sigma}_{11}$ is symmetric and positive definite, the fourth inequality holds because $\mathbf{Q}$ is orthogonal, and the last inequality holds because we are conditioning on $\ell_{11} > 0$. Because $\mathbf{\Sigma}$ and $\mathbf{Q}$ are independent, the density over $\mathcal{W}$ can be written as $p(\mathbf{w}) = p(\sigma_{21} \mid \sigma_{22}, \sigma_{11})p(\sigma_{11}, \sigma_{22})p(q_{11}, q_{21})$. Because the density over $\mathbf{Q}$ is invariant to left multiplication by $\mathbf{P}_2^-$,

R1 $p(q_{11}, q_{21}) = p(q_{11}, -q_{21})$.

For $\mathcal{S} \subset \mathcal{W}$, let $P(\mathcal{S})$ denote the probability of being in $\mathcal{S}$. Let $\mathcal{L}_+(c) = \{\mathbf{w} \in \mathcal{W} \mid \ell_{21} > c\}$ and $\mathcal{L}_-(c) = \{\mathbf{w} \in \mathcal{W} \mid \ell_{21} < -c\}$. The distribution of ℓ_{21}, conditional on $\ell_{11} > 0$, is symmetric if and only if $P(\mathcal{L}_+(c)) = P(\mathcal{L}_-(c))$, for every c. We show that $P(\mathcal{L}_+(c)) = P(\mathcal{L}_-(c))$, for every c, if and only if $\phi_{21} = 0$. We will need the following lemma.

Lemma 3. *If $\mathbf{\Sigma} \sim IW(\nu, \mathbf{\Phi})$ and $\mathbf{J}$ is a $k \times n$ matrix of rank k, then*

$$\mathbf{J\Sigma J}' \sim IW(\nu - (n-k), \mathbf{J\Phi J}').$$

Proof. Follows directly from Proposition 8.9 in Eaton (2007) and the fact that $\mathbf{\Sigma} \sim IW(\nu, \mathbf{\Phi})$ if and only if $\mathbf{\Sigma}^{-1} \sim W(\nu, \mathbf{\Phi}^{-1})$. □

We apply Lemma 3 with $\mathbf{J}$ equal to the first two rows of $\mathbf{I}_n$, so $\mathbf{\Sigma}_{11} = \mathbf{J\Sigma J}'$ is inverse Wishart with parameters $\nu - (n-2)$ and $\mathbf{\Phi}_{11} = \mathbf{J\Phi J}'$. A direct computation gives

$$\mathbf{\Sigma}_{11}^{-1} = \frac{1}{\det(\mathbf{\Sigma}_{11})}\begin{bmatrix} \sigma_{22} & -\sigma_{21} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix}.$$

The trace of $\mathbf{\Phi}_{11}\mathbf{\Sigma}_{11}^{-1}$ is equal to $(\phi_{11}\sigma_{22} + \phi_{22}\sigma_{11} - 2\phi_{21}\sigma_{21})/\det(\mathbf{\Sigma}_{11})$. Thus, the density of σ_{21}, conditional on σ_{11} and σ_{22}, is proportional to

$$|\det(\mathbf{\Sigma}_{11})|^{-\frac{\nu+3}{2}} e^{-\frac{1}{2}\operatorname{tr}(\mathbf{\Phi}_{11}\mathbf{\Sigma}_{11}^{-1})} = |\sigma_{11}\sigma_{22} - \sigma_{21}^2|^{-\frac{\nu+3}{2}} e^{-\frac{\phi_{11}\sigma_{22}+\phi_{22}\sigma_{11}}{2(\sigma_{11}\sigma_{22}-\sigma_{21}^2)}} e^{\frac{\phi_{21}\sigma_{21}}{\sigma_{11}\sigma_{22}-\sigma_{21}^2}}.$$

So, we have the following.

R2 If $\phi_{21} = 0$, then $p(\sigma_{21} \mid \sigma_{11}, \sigma_{22}) = p(-\sigma_{21} \mid \sigma_{11}, \sigma_{22})$.

R3 If $\phi_{21} \neq 0$, then $\phi_{21}\sigma_{21} > 0$ implies $p(\sigma_{21} \mid \sigma_{11}, \sigma_{22}) > p(-\sigma_{21} \mid \sigma_{11}, \sigma_{22})$.

Let $m_\sigma(\mathbf{w})$ change the sign of σ_{21}, $m_q(\mathbf{w})$ change the sign of q_{21}, and $m_{\sigma q}(\mathbf{w})$ change the sign of both σ_{21} and q_{21}. Because R1 holds, $P(\mathcal{S}) = P(m_q(\mathcal{S}))$, for every $\mathcal{S} \subset \mathcal{W}$. If $\phi_{21} = 0$, so that R1 and R2 hold, then $P(\mathcal{S}) = P(m_{\sigma q}(\mathcal{S}))$, for every $\mathcal{S} \subset \mathcal{W}$. If $\phi_{21} \neq 0$, so that R1 and R3 hold, then $P(\mathcal{S}) \neq P(m_{\sigma q}(\mathcal{S}))$, for every $\mathcal{S} \subset \mathcal{W}$ such that $\sigma_{21} > 0$ for every $\mathbf{w} \in \mathcal{S}$.

Assume that $\phi_{21} = 0$. The term ℓ_{21} is a function of $\mathbf{w}$ and $\ell_{21}(m_{\sigma q}(\mathbf{w})) = -\ell_{21}(\mathbf{w})$, which implies that $m_{\sigma q}(\mathcal{L}_+(c)) = \mathcal{L}_-(c)$, for every c. So, $P(\mathcal{L}_+(c)) = P(m_{\sigma q}(\mathcal{L}_+(c))) = P(\mathcal{L}_-(c))$, for every c.

Assume that $\phi_{21} \neq 0$. It suffices to show that $P(\mathcal{L}_+(0)) \neq P(\mathcal{L}_-(0))$. Note that $\mathbf{w} \in \mathcal{L}_+(0)$

if and only if $\sigma_{21} > b(\mathbf{w}) = -q_{21}(\sigma_{11}\sigma_{22} - \sigma_{21}^2)^{\frac{1}{2}}/q_{11}$. Let

$$\mathcal{A} = \{\mathbf{w} \in \mathcal{W} \mid |b(\mathbf{w})| < \sigma_{21}\} \cup \{\mathbf{w} \in \mathcal{W} \mid 0 < q_{21} \text{ and } -b(\mathbf{w}) = \sigma_{21}\},$$

$$\mathcal{B} = \{\mathbf{w} \in \mathcal{W} \mid 0 < q_{21} \text{ and } b(\mathbf{w}) < \sigma_{21} < -b(\mathbf{w})\}.$$

It is easy to see that $\mathcal{L}_+(0) = \mathcal{A} \cup \mathcal{B}$ and $\mathcal{L}_-(0) = m_{\sigma q}(\mathcal{A}) \cup m_q(\mathcal{B})$. Since $P(\mathcal{B}) = P(m_q(\mathcal{B}))$ and $P(\mathcal{A}) \neq P(m_{\sigma q}(\mathcal{A}))$, the results follow. $\square$

Proof of Proposition 2 (ii). Since $\sigma_{21} = r_{11}r_{21}$, the element $\sigma_{21} \neq 0$ if and only if $r_{21} \neq 0$. Since $\ell_{21} = r_{21}q_{11} + r_{22}q_{21}$, the support of the distribution of ℓ_{21}, conditional on $\ell_{11} > 0$, is $\left[-r_{22}^2, \sqrt{r_{21}^2 + r_{22}^2}\right]$, if $r_{21} > 0$, and is $\left[-\sqrt{r_{21}^2 + r_{22}^2}, r_{22}^2\right]$, if $r_{21} < 0$. Thus, the support is not symmetric and so the distribution cannot be symmetric. If $r_{21} = 0$, then $\ell_{21} = r_{22}q_{21}$. Since $p(\mathbf{Q} \mid \mathbf{\Sigma})$ is uniform, the distribution of q_{21}, conditional on $q_{11} > 0$, is symmetric. Thus, the distribution of ℓ_{21}, conditional on $\ell_{11} > 0$, is symmetric. $\square$

Proof of Proposition 3. If the structural-form parameters $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$ are observationally equivalent then $g(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) = g(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$, or equivalently, $\tilde{\mathbf{B}} = \mathbf{B}$ and $\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}\tilde{\mathbf{L}}_0' = \mathbf{L}_0\mathbf{\Omega}\mathbf{L}_0'$. Multiplying $\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}\tilde{\mathbf{L}}_0' = (\mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}})(\mathbf{\Omega}^{\frac{1}{2}}\mathbf{L}_0')$ on the left by $(\mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}})^{-1}$ and on the right by $(\mathbf{\Omega}^{\frac{1}{2}}\mathbf{L}_0')^{-1}$, gives

$$\left(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{L}_0^{-1}\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}^{\frac{1}{2}}\right)\left(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{L}_0^{-1}\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}^{\frac{1}{2}}\right)' = \mathbf{I}_n$$

So, $\mathbf{Q} = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{L}_0^{-1}\tilde{\mathbf{L}}_0\tilde{\mathbf{\Omega}}^{\frac{1}{2}}$ is an orthogonal matrix. Thus, $\tilde{\mathbf{L}}_0 = \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}}$.

All that remains to be shown is that $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\mathbf{B}, \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}}, \tilde{\mathbf{\Omega}})$ are observationally equivalent for any $n \times n$ orthogonal matrix $\mathbf{Q}$ and any $n \times n$ positive diagonal matrix $\tilde{\mathbf{\Omega}}$. Since

$$(\mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}})\tilde{\mathbf{\Omega}}(\mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}})' = \mathbf{L}_0\mathbf{\Omega}\mathbf{L}_0',$$

both $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\mathbf{B}, \mathbf{L}_0\mathbf{\Omega}^{\frac{1}{2}}\mathbf{Q}\tilde{\mathbf{\Omega}}^{-\frac{1}{2}}, \tilde{\mathbf{\Omega}})$ map to the same reduced-form parameters, which implies they are observationally equivalent. $\square$

Proof of Corollary 1. Given $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\mathbf{\Omega}})$, define $\mathbf{D} = \mathbf{\Omega}^{-1}\tilde{\mathbf{\Omega}}$, so that $\tilde{\mathbf{\Omega}} = \mathbf{\Omega}\mathbf{D}$.

By Proposition 3, $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{L}}_0, \tilde{\boldsymbol{\Omega}})$ are observationally equivalent if and only if $\tilde{\mathbf{B}} = \mathbf{B}$ and there exists a $\mathbf{Q} \in \mathcal{O}_n$ such that $\tilde{\mathbf{L}}_0 = \mathbf{L}_0 \boldsymbol{\Omega}^{\frac{1}{2}} \mathbf{Q} \tilde{\boldsymbol{\Omega}}^{-\frac{1}{2}} = \mathbf{L}_0 \boldsymbol{\Omega}^{\frac{1}{2}} \mathbf{Q} \boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}}$, where the last equality follows from substitution and the fact that multiplication of diagonal matrices is commutative. □

Proof of Propositions 4 and 5. If $\mathcal{R}$ is a scale restriction, then for almost every structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$, there exists a $\mathbf{D}_1 \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0 \mathbf{D}_1^{-\frac{1}{2}}, \boldsymbol{\Omega} \mathbf{D}_1) \in \mathcal{R}$. If $\mathcal{R}$ is either a label restriction or an economic restriction, then $(\mathbf{B}, \mathbf{L}_0 \mathbf{D}_1^{-\frac{1}{2}} \mathbf{D}_2^{-\frac{1}{2}}, \boldsymbol{\Omega} \mathbf{D}_1 \mathbf{D}_2) \in \mathcal{R}$ for every $\mathbf{D}_2 \in \mathcal{D}_n^{++}$. Taking $\mathbf{D}_2 = \mathbf{D}_1^{-1}$ gives that $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R}$ for almost all structural-form parameters. Thus, the restriction is trivial. If $\mathcal{R}$ is a label restriction, then for almost every structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega})$, there exists a $\mathbf{P}_1 \in \mathcal{O}_n^D$ such that $(\mathbf{B}, \mathbf{L}_0 \mathbf{P}_1, \boldsymbol{\Omega}) \in \mathcal{R}$. If $\mathcal{R}$ is either a scale restriction or an economic restriction, then $(\mathbf{B}, \mathbf{L}_0 \mathbf{P}_1 \mathbf{P}_2, \boldsymbol{\Omega}) \in \mathcal{R}$ for every $\mathbf{P}_2 \in \mathcal{O}_n^D$. Taking $\mathbf{P}_2 = \mathbf{P}_1$ gives that $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R}$ for almost all structural-form parameters. Thus, the restriction is trivial. □

Proof of Proposition 6. Let $\mathcal{R}_i$ be a restriction that does not impose a relation between the label and scale for every i in some index set $\mathcal{I}$. Suppose $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \notin \bigcap_{i \in \mathcal{I}} \mathcal{R}_i$ and there exists $\mathbf{P} \in \mathcal{O}_n^D$ and $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0 \mathbf{P}, \boldsymbol{\Omega}) \in \bigcap_{i \in \mathcal{I}} \mathcal{R}_i$ and $(\mathbf{B}, \mathbf{L}_0 \mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega} \mathbf{D}) \in \bigcap_{i \in \mathcal{I}} \mathcal{R}_i$. Since $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \notin \bigcap_{i \in \mathcal{I}} \mathcal{R}_i$, there exists an $i \in \mathcal{I}$ such that $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \notin \mathcal{R}_i$. Since both $(\mathbf{B}, \mathbf{L}_0 \mathbf{P}, \boldsymbol{\Omega})$ and $(\mathbf{B}, \mathbf{L}_0 \mathbf{D}^{-\frac{1}{2}}, \boldsymbol{\Omega} \mathbf{D})$ are in $\mathcal{R}_i$, this contradicts the fact that $\mathcal{R}_i$ does not impose a relation between the label and scale. □

Proof of Proposition 7. It follows from Proposition 6 that if $\mathcal{R}_e$ is an economic restriction, $\mathcal{R}_\ell$ is a label restriction, and $\mathcal{R}_s$ is a scale restriction, then $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$ does not impose a relation between the label and scale. To complete the proof of Proposition 7, we must show that if $\mathcal{R}$ is a restriction that does not impose a relation between the label and scale, then $\mathcal{R} = \mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$, where $\mathcal{R}_e$ is an economic restriction, $\mathcal{R}_\ell$ is a label restriction, and $\mathcal{R}_s$ is a scale restriction.

Let $\mathcal{R}$ be any restriction that does not impose a relation between the label and scale and let $\mathcal{R}_e$, $\mathcal{R}_\ell$, $\mathcal{R}_s$, $\mathcal{R}_d$, and $\mathcal{R}_p$ be defined by Equations (3)-(5). First we show that $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s = \mathcal{R}$. Since $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s = \mathcal{R}_d \cap \mathcal{R}_p$, it suffices to show that $\mathcal{R}_d \cap \mathcal{R}_p = \mathcal{R}$. It follows directly from the definition of $\mathcal{R}_d$ and $\mathcal{R}_p$ that $\mathcal{R} \subset \mathcal{R}_d \cap \mathcal{R}_p$. Now suppose that $(\mathbf{B}, \mathbf{L}_0, \boldsymbol{\Omega}) \in \mathcal{R}_d \cap \mathcal{R}_p$.

This implies that there exists a $\mathbf{P} \in \mathcal{O}_n^D$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \mathbf{\Omega}) \in \mathcal{R}$ and a $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R}$. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \notin \mathcal{R}$, then the restriction $\mathcal{R}$ would impose a relation between the label and scale. So, it must be the case that $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}$ and $\mathcal{R}_d \cap \mathcal{R}_p \subset \mathcal{R}$. Thus, $\mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s = \mathcal{R}$. To complete the proof, we show that $\mathcal{R}_e$ is an economic restriction, $\mathcal{R}_\ell$ is a label restriction, and $\mathcal{R}_s$ is a scale restriction.

It is easy to see that $\mathcal{R}_e$ satisfies E1, so $\mathcal{R}_e$ is an economic restriction. It is also easy to see that $\mathcal{R}_d$ satisfies L1, that $\mathcal{R}_p$ satisfies S1, and that $\mathcal{R}_e^c$ satisfies both L1 and S1. Since both $\mathcal{R}_d$ and $\mathcal{R}_e^c$ satisfy L1, so does $\mathcal{R}_d \cup \mathcal{R}_e^c = \mathcal{R}_\ell$. Since both $\mathcal{R}_p$ and $\mathcal{R}_e^c$ satisfy S1, so does $\mathcal{R}_p \cup \mathcal{R}_e^c = \mathcal{R}_s$. All that remains to be shown is that L2 and S2 hold.

To show that L2 holds, it suffices to show that for every structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ there exists $\mathbf{P} \in \mathcal{O}_n^D$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \mathbf{\Omega}) \in \mathcal{R}_\ell$. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_e^c$, then we can take $\mathbf{P}$ to be the identity. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_e$, then there exists a $\mathbf{P} \in \mathcal{O}_n^D$ and a $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{P}\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R}$. Thus, $(\mathbf{B}, \mathbf{L}_0\mathbf{P}, \mathbf{\Omega}) \in \mathcal{R}_d \subset \mathcal{R}_\ell$ and L2 holds.

To show that S2 holds, it suffices to show that for every structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$ there exists $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R}_s$. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_e^c$, then we can take $\mathbf{D}$ to be the identity. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_e$, then there exists $\mathbf{P} \in \mathcal{O}_n^D$ and a $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{P}\mathbf{D}^{\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}^{-1}) \in \mathcal{R}$. Thus, $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R}_p \subset \mathcal{R}_s$ and S2 holds. □

Proof of Proposition 8. Suppose that S1 was omitted from the definition of a scale restriction. If L1 was omitted from the definition of a label restriction, then the proof is similar. In this case, $\mathcal{R}_e$ and $\mathcal{R}_\ell$ are given by Equations (3) and (4), but $\mathcal{R}_s = \mathcal{R} \cup \mathcal{R}_d^c$. Since $\mathcal{R}_e$ and $\mathcal{R}_\ell$ are as in the proof of Proposition 7, it suffices to show that $\mathcal{R} = \mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$ and $\mathcal{R}_s$ satisfies S2. We first show that $\mathcal{R} = \mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$. From the definitions of $\mathcal{R}_e$ and $\mathcal{R}_\ell$, it is easy to see that $\mathcal{R}_e \cap \mathcal{R}_\ell = \mathcal{R}_d$. From the definition of $\mathcal{R}_s$, it is easy to see that $\mathcal{R}_d \cap \mathcal{R}_s = \mathcal{R}$. So, $\mathcal{R} = \mathcal{R}_e \cap \mathcal{R}_\ell \cap \mathcal{R}_s$.

To complete the proof we must show that $\mathcal{R}_s$ satisfies S2. It suffices to show that for every structural-form parameter $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega})$, there exists a $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R}_s$. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_d^c$, then we can take $\mathbf{D} = \mathbf{I}_n$. If $(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) \in \mathcal{R}_d$, then there exists a $\mathbf{D} \in \mathcal{D}_n^{++}$ such that $(\mathbf{B}, \mathbf{L}_0\mathbf{D}^{-\frac{1}{2}}, \mathbf{\Omega}\mathbf{D}) \in \mathcal{R} \subset \mathcal{R}_s$. □

Proof of Proposition 9. The following lemma will allow us to explicitly compute the volume element of $f^{-1} \circ \phi^{-1}$.

Lemma 4. *The volume elements of ϕ^{-1} and f^{-1} are*

$$v_{\phi^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-\frac{1}{2}}, \tag{13}$$

$$v_{f^{-1}}(\mathbf{B}, \mathbf{L}_0) = 2^{n} |\det(\operatorname{diag}(\mathbf{L}_0))|^{-(n-2)} \tag{14}$$

Proof. By Lemma 1, $v_\phi(\mathbf{B}, \mathbf{L}_0) = 2^{\frac{n(n+1)}{2}} |\det(\mathbf{L}_0)|$. Since

$$v_{\phi^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = v_\phi(\phi^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))^{-1} = 2^{-\frac{n(n+1)}{2}} |\det(r(\mathbf{\Sigma})\mathbf{Q})|^{-1} = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-\frac{1}{2}},$$

Equation (13) holds. We now show Equation (14) holds. We can parameterize $(\mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})$ as $(\operatorname{vec}(\mathbf{B})', \boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_n)$, where

$$\boldsymbol{\lambda}_j = (\mathbf{K}_0(1,j), \cdots, \mathbf{K}_0(j-1,j), \omega_j, \mathbf{K}_0(j+1,j), \cdots, \mathbf{K}_0(n,j))$$

and ω_j is the j^{th} diagonal element of $\mathbf{\Omega}$. Because $\mathbf{L}_0(i,j) = \mathbf{K}_0(i,j)\omega_j^{\frac{1}{2}}$, the derivative of f will be

$$\begin{bmatrix} \mathbf{I}_{n(np+1)} & 0 & \cdots & 0 \\ 0 & \mathbf{D}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_n \end{bmatrix},$$

where

$$\mathbf{D}_j = \begin{bmatrix} \omega_j^{\frac{1}{2}} & \cdots & 0 & \frac{1}{2}\mathbf{K}_0(1,j)\omega_j^{-\frac{1}{2}} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_j^{\frac{1}{2}} & \frac{1}{2}\mathbf{K}_0(j-1,j)\omega_j^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{2}\omega_j^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{2}\mathbf{K}_0(j+1,j)\omega_j^{-\frac{1}{2}} & \omega_j^{\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{2}\mathbf{K}_0(n,j)\omega_j^{-\frac{1}{2}} & 0 & \cdots & \omega_j^{\frac{1}{2}} \end{bmatrix}$$

From this it is easy to see that $\det(\mathbf{D}_j) = \frac{1}{2}\omega_j^{\frac{n-2}{2}}$. So, $v_f(\mathbf{B}, \mathbf{L}_0, \mathbf{\Omega}) = 2^{-n}|\det(\mathbf{\Omega})|^{\frac{(n-2)}{2}}$. Since $v_{f^{-1}}(\mathbf{B}, \mathbf{L}_0) = v_f(f^{-1}(\mathbf{B}, \mathbf{L}_0))^{-1} = 2^n|\det(\mathrm{diag}(\mathbf{L}_0))|^{-(n-2)}$, Equation (14) holds. □

It follows directly from Lemma 4 that

$$v_{(\phi\circ f)^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n-1)}{2}}|\det(\mathbf{\Sigma})|^{-\frac{1}{2}}|\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}.$$

Because the complement of $\mathcal{A}$ is of measure zero, the integral of a function over $\mathcal{A}$ is equal to the integral of that function over all $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$. Putting this all together, we have

$$\begin{aligned}\int_{\mathcal{B}} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{K}_0, \mathbf{\Omega})\, d\mathbf{B}\, d\mathbf{K}_0\, d\mathbf{\Omega} &= \int_{\mathbb{R}^{m\times n}\times\mathcal{S}_n^+\times\mathcal{O}_n} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma})\, v_{f^{-1}\circ\phi^{-1}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})\, d\mathbf{B}\, d\mathbf{\Sigma}\, d\mathbf{Q} \\ &= 2^{-\frac{n(n-1)}{2}} \int_{\mathbb{R}^{m\times n}\times\mathcal{S}_n^+} p(\mathbf{Y}_T \mid \mathbf{B}, \mathbf{\Sigma})|\det(\mathbf{\Sigma})|^{-\frac{1}{2}} \left(\int_{\mathcal{O}_n} |\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}\, d\mathbf{Q}\right) d\mathbf{B}\, d\mathbf{\Sigma}.\end{aligned}$$

We will show that for $n \geq 3$ and every $\mathbf{\Sigma}$, $\int_{\mathcal{O}_n} |\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}\, d\mathbf{Q} = \infty$. The intuition behind this is that if $x = |\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|$ and $\mathbf{Q}_0$ is an orthogonal matrix such that $r(\mathbf{\Sigma})\mathbf{Q}_0$ has at least one zero diagonal element, then, as $\mathbf{Q}$ approaches $\mathbf{Q}_0$, x approaches zero and $x^{-(n-2)}$ approaches infinity at least as fast as $1/x$, which ensures that the integral cannot converge to a finite number. We make this intuition rigorous.

Let $\mathbf{r}_i$ be the i^{th} row of $r(\mathbf{\Sigma})$ and $\mathbf{q}_i$ be the i^{th} column of $\mathbf{Q}$, so that

$$|\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)} = \prod_{i=1}^{n} |\mathbf{r}_i\mathbf{q}_i|^{-(n-2)}.$$

Because $\|\mathbf{q}_i\| = 1$, $|\mathbf{r}_i\mathbf{q}_i| \leq \|\mathbf{r}_i\|$ for every $\mathbf{q}_i$, which implies $|\mathbf{r}_i\mathbf{q}_i|^{-(n-2)} \geq \|\mathbf{r}_i\|^{-(n-2)}$ for every $\mathbf{q}_i$. Because $r(\mathbf{\Sigma})$ is lower triangular, $|\mathbf{r}_1\mathbf{q}_1| = \|\mathbf{r}_1\| \cdot |q_{11}|$, where q_{11} is the element in the first row and column of $\mathbf{Q}$. So,

$$\int_{\mathcal{O}_n} |\det(\mathrm{diag}(r(\mathbf{\Sigma})\mathbf{Q}))|^{-(n-2)}\, d\mathbf{Q} \geq \prod_{i=1}^{n} \|\mathbf{r}_i\|^{-(n-2)} \int_{\mathcal{O}_n} |q_{11}|^{-(n-2)}\, d\mathbf{Q}.$$

Let $\tilde{\mathbf{q}}_1$ be the last $n-1$ elements of $\mathbf{q}_1$. Given q_{11}, let $S(q_{11})$ be the set of all such $\tilde{\mathbf{q}}_1$. The set $S(q_{11})$ is the sphere of radius $(1-q_{11}^2)^{\frac{1}{2}}$ in $\mathbb{R}^{n-1}$ centered at the origin. The integral

$\int_{S(q_{11})} 1 d\tilde{\mathbf{q}}_1$, which will be of use in what follows, is proportional to the radius of $S(q_{11})$ raised to the power of $n-2$. Thus, there is a positive constant c_1 such that $\int_{S(q_{11})} 1 d\tilde{\mathbf{q}}_1 = c_1(1-q_{11}^2)^{\frac{n-2}{2}}$.

Let $\tilde{\mathbf{Q}}$ be the last $n-1$ columns of $\mathbf{Q}$. Given $\mathbf{q}_1$, let $S(\mathbf{q}_1)$ be the set of all such $\tilde{\mathbf{Q}}$. The integral $\int_{S(\mathbf{q}_1)} 1 d\tilde{\mathbf{Q}}$, which will also be of use in what follows, does not depend on $\mathbf{q}_1$. This is because given any two n-vectors, $\mathbf{q}_1$ and $\hat{\mathbf{q}}_1$, both of length one, there exists $\mathbf{Q}_0 \in \mathcal{O}_n$ such that $\hat{\mathbf{q}}_1 = \mathbf{Q}_0 \mathbf{q}_1$. This implies $S(\hat{\mathbf{q}}_1) = \mathbf{Q}_0 S(\mathbf{q}_1)$. Thus, there is a positive constant c_2 such that $\int_{S(\mathbf{q}_1)} 1 d\tilde{\mathbf{Q}} = c_2$, for every $\mathbf{q}_1$.

So,

$$\begin{aligned}\int_{\mathcal{O}_n} |q_{11}|^{-(n-2)}\, d\mathbf{Q} = \int_{-1}^{1} |q_{11}|^{-(n-2)} \left(\int_{S(q_{11})} \left(\int_{S(\mathbf{q}_1)} 1\, d\tilde{\mathbf{Q}} \right) d\tilde{\mathbf{q}}_1 \right) dq_{11} \\ = c_1 c_2 \int_{-1}^{1} (1-q_{11}^2)^{\frac{n-2}{2}} |q_{11}|^{-(n-2)} dq_{11}.\end{aligned}$$

For $0 < q_{11} < \varepsilon < 1$, $(1-q_{11}^2)^{\frac{n-2}{2}} |q_{11}|^{-(n-2)} \geq (1-\varepsilon^2)^{\frac{n-2}{2}}/q_{11}$. Since $1/q_{11}$ integrates to $\ln(q_{11})$, $\int_{\mathcal{O}_n} |q_{11}|^{-(n-2)}\, d\mathbf{Q} \geq c_1 c_2 (1-\varepsilon^2)^{\frac{n-2}{2}} \int_0^{\varepsilon} 1/q_{11}\, dq_{11} = \infty$. $\square$

Proof of Proposition 10. Let $\bar{\ell} = \underline{\ell}^{1/2}$. Under the mapping $(\phi \circ f)^{-1}$, the restriction $\omega_i > \underline{\ell}$ is equivalent to $|e_i' r(\boldsymbol{\Sigma}) \mathbf{Q} e_i| > \bar{\ell}$ for all $i = 1, \ldots, n$, because $\omega_i = (e_i' r(\boldsymbol{\Sigma}) \mathbf{Q} e_i)^2$. Let $\mathcal{O}_{\bar{\ell}}(\boldsymbol{\Sigma})$ denote the set of orthogonal matrices satisfying this restriction:

$$\mathcal{O}_{\bar{\ell}}(\boldsymbol{\Sigma}) = \{\mathbf{Q} \in \mathcal{O}_n : |e_i' r(\boldsymbol{\Sigma}) \mathbf{Q} e_i| > \bar{\ell} \text{ for all } i = 1, \ldots, n\}.$$

Given the proof of Proposition 9, it suffices to show that

$$\int_{\mathcal{O}_{\bar{\ell}}(\boldsymbol{\Sigma})} |\det(\operatorname{diag}(r(\boldsymbol{\Sigma})\mathbf{Q}))|^{-(n-2)} d\mathbf{Q}$$

is finite. Each element of $\mathcal{O}_{\bar{\ell}}(\boldsymbol{\Sigma})$ satisfies

$$|\mathbf{r}_i \cdot \mathbf{q}_i| = |e_i' r(\boldsymbol{\Sigma}) \mathbf{Q} e_i| > \bar{\ell} > 0, \text{ so } |\mathbf{r}_i \cdot \mathbf{q}_i|^{-(n-2)} < \bar{\ell}^{-(n-2)}.$$

Taking the product over $i = 1, \dots, n$:

$$\prod_{i=1}^{n} |\mathbf{r}_i \cdot \mathbf{q}_i|^{-(n-2)} < \bar{\ell}^{-n(n-2)}.$$

The set of orthogonal matrices $\mathcal{O}_n$ has finite Haar volume $V_n := \mathrm{vol}(\mathcal{O}_n) < \infty$. Hence,

$$\sup_{\mathbf{\Sigma} \in \mathcal{S}_n^+} \int_{\mathcal{O}_{\bar{\ell}}(\mathbf{\Sigma})} \prod_{i=1}^{n} |\mathbf{r}_i \cdot \mathbf{q}_i|^{-(n-2)} \, d\mathbf{Q} \;\leq\; \int_{\mathcal{O}_n} \bar{\ell}^{-n(n-2)} \, d\mathbf{Q} \;=\; V_n \, \bar{\ell}^{-n(n-2)}.$$

Therefore,

$$\int_{\mathcal{O}_{\bar{\ell}}(\mathbf{\Sigma})} \left|\det \operatorname{diag}(r(\mathbf{\Sigma})\mathbf{Q})\right|^{-(n-2)} \, d\mathbf{Q} \;\leq\; V_n \, \bar{\ell}^{-n(n-2)} < \infty.$$

Since the remaining kernel over $(\mathbf{B}, \mathbf{\Sigma})$ is proportional to $NIW(\mathbf{B}, \mathbf{\Sigma}; \bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Theta}}) |\det(\mathbf{\Sigma})|^{-1/2}$, which is integrable, the full posterior is proper. $\square$

References

Arias, J., J. F. Rubio-Ramirez, and D. F. Waggoner (2025). Uniform Priors for Impulse Responses. *Econometrica 93*, 695–718.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2019). The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure. *Journal of Monetary Economics 101*, 1–13.

Arias, J. E., J. F. Rubio-Ramírez, D. Rudolf, and M. Shin (2026). Large SVARs. Working paper.

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 (2018). Inference in Structural Vector Autoregressions

When the Identifying Assumptions Are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations. *Journal of Monetary Economics 100*, 48–65.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1996). The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds. *Review of Economics and Statistics 78*(1), 16–34.

Eaton, M. L. (2007). *Multivariate Statistics: A Vector Space Approach*, Volume 53 of *Institute of Mathematical Statistics Lecture Notes - Monograph Series*. Institute of Mathematical Statistics.

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

Stock, J. H. and M. W. Watson (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. *Handbook of Macroeconomics 2*, 415–525.

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.