Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications

Jonas E. Arias
Federal Reserve Board

Juan F. Rubio-Ramírez*
Emory University, BBVA Research,
and Federal Reserve Bank of Atlanta

Daniel F. Waggoner
Federal Reserve Bank of Atlanta

December 6, 2015

Abstract

In this paper we characterize agnostic priors and propose numerical algorithms for Bayesian inference when using them within SVARs identified by imposing sign and zero restrictions on a function of the structural parameters. As Baumeister and Hamilton (2015a) has made clear, since the data cannot tell apart models satisfying the sign and zero restrictions, priors play a crucial role in this environment. We will emphasize the importance of agnostic priors: any prior density that is equal across observationally equivalent parameters. If the prior is not agnostic, additional restrictions to the proclaimed sign and zero restrictions become part of the identification. While our numerical algorithms use agnostic priors we will show that existing ones use non-agnostic priors. We will use Beaudry, Nam and Wang (2011) work on the relevance of optimism shocks to show the dangers of using non-agnostic priors.

*Corresponding author: Juan F. Rubio-Ramírez <jrubior@emory.edu>, Economics Department, Emory University, Atlanta, GA 30322. We thank Paul Beaudry, Andrew Mountford, Deokwoo Nam, and Jian Wang for sharing supplementary material with us, and for helpful comments. We also thank Grátula Bedátula for her support and help. Without her this paper would have been impossible. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Board of Governors of the Federal Reserve System. This paper has circulated under the title "Algorithm for Inference with Sign and Zero Restrictions." Juan F. Rubio-Ramírez also thanks the NSF, the Institute for Economic Analysis (IAE) and the "Programa de Excelencia en Educación e Investigación" of the Bank of Spain, and the Spanish ministry of science and technology Ref. ECO2011-30323-c03-01 for support.

1 Introduction

SVARs identified by imposing sign and zero restrictions have become prominent. The facts that identification generally comes from fewer restrictions than in traditional identification schemes and that any conclusion is robust across the set of structural parameters consistent with the restrictions, have made the approach attractive to researchers. Most papers using this approach work in the Bayesian paradigm.[1] As Baumeister and Hamilton (2015a) has made clear that priors play a crucial role in this environment. If the researcher wants the identification to solely come from the sign and zero restrictions, she should be careful when choosing the prior.

When the SVAR is identified by imposing only sign restrictions, we emphasize the importance of agnostic priors: any prior density that is equal across observationally equivalent parameters. If the prior is not agnostic, additional restrictions to the proclaimed sign restrictions become part of the identification. In particular, we first show that a SVAR can be parameterized in three different ways. In addition to the typical structural and impulse response function (IRF) parameterizations, it can also be written as the product of the reduced-form parameters and the set of orthogonal matrices: we call such parameterization the orthogonal reduced-form parameterization. Second, we characterize agnostic priors in the three parameterizations and highlight that an agnostic prior over the orthogonal reduced-form parameterization is flat over the set of orthogonal matrices. Third, we describe the change of variable theorems that allows a researcher that details a prior in any of the three parameterizations to find the equivalent prior over any of the other two. Fourth, we demonstrate that if a prior is agnostic over one parameterization then it is agnostic over the three of them. Thus, when one only cares about the prior being agnostic, the choice of parameterization is irrelevant. Finally, we show that a prior is agnostic if and only if the implied posterior is agnostic: any posterior density that is equal across observationally equivalent parameters.

These results imply that current numerical algorithms for SVARs identified by imposing only sign restrictions, as described by Rubio-Ramírez, Waggoner and Zha (2010), implicitly involve agnostic priors. These algorithms are written over the orthogonal reduced-form parameterization and use conjugate priors over this parameterization.[2] Since the likelihood does not bear any information about the

[1] Exceptions are Moon and Schorfheide (2012), Moon, Schorfheide and Granziera (2013), and Gafarov and Montiel Olea (2015). Moon and Schorfheide (2012) analyze the differences between Bayesian credible sets and frequentist confidence sets in partially identified models. Moon, Schorfheide and Granziera (2013) develops methods of constructing error bands for impulse response functions of sign-restricted SVARs that are valid from a frequentist perspective. Gafarov and Montiel Olea (2015) is the paper closest to ours and it will be analyzed in more detail below.

[2] By conjugate priors we mean priors that have the same functional form that the likelihood.

orthogonal matrices any conjugate prior over the orthogonal reduced-form parameterization is flat over the set of orthogonal matrices and, therefore, agnostic. The conjugate posterior over the orthogonal reduced-form parameterization is normal-inverse-Wishart over the reduced-form parameters and flat over the set of orthogonal matrices, therefore it is extremely easy to draw from it. Therefore, although the current numerical algorithms draw from conjugate posteriors over the orthogonal reduced-form parameterization for practical convenience, they are implicitly drawing from the equivalent agnostic posterior over both the structural and the IRF parameterizations.

As Baumeister and Hamilton (2015a) highlight, a researcher may want to use flat priors over either the structural or the IRF parameterization. This is not a problem for current numerical algorithms. We show how a particular choice of conjugate prior over the orthogonal reduced-form parameterization implies an equivalent flat prior over the structural parameterization. A different choice of conjugate prior implies an equivalent flat prior over the IRF parameterization. The key is to notice that flat priors over either the structural or the IRF parameterization are agnostic and, hence, they imply an equivalent prior over the orthogonal reduced-form parameterization that it is flat over the set of orthogonal matrices. Thus, for any of the two choices, the conjugate posterior over the orthogonal reduced-form parameterization is still normal-inverse-Wishart over the reduced-form parameters and flat over the set of orthogonal matrices and it is still extremely easy to draw from it. In any case, one has to choose over which parameterization to be flat: the conjugate prior over the orthogonal reduced-form parameterization that implies a flat prior over the structural parameterization does not necessarily imply a flat prior over the IRF parameterization.

Next we adapt the current numerical algorithms to consider both sign and zero restrictions. When only sign restrictions are considered, current numerical algorithms work because the sign restrictions define an open set in the structural parameters. But the zero restrictions define a smooth manifold of measure zero in the set of structural parameters. We need to draw from a posterior conditional on the zero restrictions because the sign restrictions define an open set in the smooth manifold. If the researcher begins with an agnostic prior over a particular parameterization and then conditions on the zero restrictions, the resulting conditional prior over the particular parameterization will be conditionally agnostic: any prior density over the particular parameterization that is equal across observationally equivalent parameters that satisfy the zero restrictions. Using conditionally agnostic priors is important since, otherwise, additional restrictions to the proclaimed sign and zero restrictions become part of the identification.

But while agnostic priors survive transformation across parameterizations, this is not necessarily the case for conditionally agnostic priors. Using the change of variables theorem for smooth manifolds, we show that if one considers a conditionally agnostic prior over one parameterization and then transforms it into another parameterization, the resulting equivalent conditional prior is not conditionally agnostic over the new parameterization. This means that the researcher has to choose over which parameterization to impose the zero restrictions since it is not possible to have conditionally agnostic priors over two parameterizations at the same time. In general, it may be desirable to have a conditionally agnostic prior over either the structural or the IRF parameterizations. As before, it will be practical to detail a conjugate prior over the orthogonal reduced-form parameterization, but this is not a problem. The conjugate, and hence agnostic, prior over the orthogonal reduced-form parameterization just needs to be transformed over the chosen parameterization before conditioning on the zero restrictions. It is also the case that, given a parametrization, a conditional prior is conditionally agnostic if and only if the implied conditional posterior is conditionally agnostic: any conditional posterior density over the given parameterization that is equal across observationally equivalent parameters that satisfy the zero restrictions. Next, we show how the researcher can use a importance sampler to draw from the conditionally agnostic posterior over the chosen parameterization using a proposal motivated by the current numerical algorithms and detailing a conjugate prior over the orthogonal reduced-form parameterization. The importance sampler will depend on the choice of parametrization over which the researcher wants to be conditionally agnostic. Finally, it is also important to remark that by choosing the right conjugate prior over the orthogonal reduced-form parameterization a researcher can also have a conditionally flat prior over the chosen parameterization: any prior density over the chosen parameterization that is flat across parameters that satisfy the zero restrictions. The choice of conjugate prior over the orthogonal reduced-form parameterization would depend on the parameterization over which one wants to have a conditionally flat prior.

The purpose of this paper is to show that the choice of the orthogonal reduced-form parameterization and conjugate priors over such a parameterization facilitates the exercise of drawing from the agnostic posterior, in the case of only using sign restrictions, and from the conditionally agnostic posterior using a importance sampler, in the case of using sign and zero restrictions, but our framework is more general. If the researcher can directly draw from the either the agnostic posterior or the conditionally agnostic posterior over either the structural or the IRF parameterization, she can still use our framework.

When using sign and zero restrictions, the most widely used algorithm is Mountford and Uhlig's (2009) penalty function approach – PFA henceforth – prior. The PFA prior uses the orthogonal reduced-form parameterization but it is not conditionally agnostic over any parameterization. Thus, when using the PFA prior additional restrictions to the proclaimed sign and zero restrictions become part of the identification. We show the consequences of using the PFA prior by first replicating the results in Beaudry, Nam and Wang (2011), and by comparing them with the results that a researcher would obtain if an agnostic prior is used instead instead of the PFA prior. The aim of Beaudry, Nam and Wang (2011) is to provide new evidence on the relevance of optimism shocks as an important driver of macroeconomic fluctuations by means of SVARs identified by imposing a sign restriction on the impact response of stock prices to optimism shocks and a zero restriction on the impact response of TFP to these shocks. Based on the results obtained with PFA prior, Beaudry, Nam and Wang (2011) conclude that optimism shocks are clearly important for explaining standard business cycle type phenomena because they increase consumption and hours. We show that once a conditionally agnostic prior is used, that is absent the additional restrictions imposed by the PFA, the identified optimism shocks do not increase consumption and hours and, hence, there is little evidence supporting the assertion that optimism shocks are important for business cycles.

There is existing literature that criticizes the PFA prior and proposes alternative numerical algorithms that implement sign and zero restrictions to identify SVARs such as the approaches of Baumeister and Benati (2013), Binning (2013), and Gambacorta, Hofmann and Peersman (2014). These algorithms also work in the orthogonal reduced-form parameterization. In the Appendix we show that they all use priors that are not conditionally agnostic. Hence, as it was the case with the PFA prior, researchers using these algorithms are imposing additional restrictions to those acknowledged in the identification strategy.[3] In a very original approach, Giacomini and Kitawaga (2015) are also concerned with the choice of the priors in SVARs identified using sign and zero restrictions. They also work on the orthogonal reduced-form parameterization and propose a method for conducting posterior inference on IRFs that is robust to the choice of priors. Although we see our paper as sympathetic to their concern about priors choice, they do not restrict their priors to be conditionally agnostic, hence some of their prior choices impose additional restrictions to the sign and zero explicitly recognized.

Finally, we have to highlight Baumeister and Hamilton (2015a). This paper uses the structural parameterization. This is a very interesting and novel approach since the rest of the literature (including us) works in the orthogonal reduced-form parameterization. While working in the structural

[3]Caldara and Kamps (2012) also share our concerns about the PFA prior, but providing an alternative algorithm is out of the scope of their paper.

parameterization has clear advantages, mainly being able to detail priors directly on economically interpretable structural parameters, this approach has some drawbacks when compared with ours. First, it considers a smaller class of restrictions. Baumeister and Hamilton (2015a) is limited to impose zero restrictions contemporaneously and they have to choose if the zero restrictions are imposed on either the structural parameters or the IRFs: they cannot impose zero restrictions to both the contemporaneous structural parameters and the contemporaneous IRFs at the same time. Second, it can be shown that Baumeister and Hamilton (2015b), that uses Baumeister and Hamilton's (2015a) approach, defines priors that are not conditionally agnostic.

We wish to state that the aim of this paper is neither to dispute nor to challenge SVARs identified with sign and zero restrictions. In fact, our methodology preserves the virtues of the pure sign restriction approach developed in the work of Faust (1998), Canova and Nicoló (2002), Uhlig (2005), and Rubio-Ramírez, Waggoner and Zha (2010). Our point is that, since the literature has not being using agnostic priors, the reported results do not solely come from the sign and zero restrictions.

2 Implications of the Additional Restrictions

Beaudry, Nam and Wang (2011) analyze the relevance of optimism shocks as a driver of macroeconomic fluctuations using SVARs identified by imposing sign and zero restrictions. More details about their work will be given in Section 7. At this point it suffices to say that in their most basic model, Beaudry, Nam and Wang (2011) use data on total factor productivity (TFP), stock price, consumption, the real federal funds rate, and hours worked. Their identification scheme defines optimism shocks as positively affecting stock prices but being orthogonal to TFP at horizon zero and they use the PFA to implement it. Beaudry, Nam and Wang (2011) also claim to be agnostic in the sense that only these two restrictions identify the model.

Figure 1 replicates the first block of Figure 1 in Beaudry, Nam and Wang (2011). As shown by the narrow 68 percent posterior intervals of their IRFs, Beaudry, Nam and Wang (2011) obtain that consumption and hours worked respond positively and strongly to optimism shocks. If the IRFs shown in Figure 1 were the responses to optimism shocks solely identified using the two restrictions described above, they would clearly endorse the view of those who think that optimism shocks are relevant for business cycle fluctuations. But this is not the case. In Section 7 we will show that the PFA implicitly defines a prior that is not conditionally agnostic and that, therefore, additional restrictions to the two described above are being used in the identification. We will also show that if a conditionally agnostic

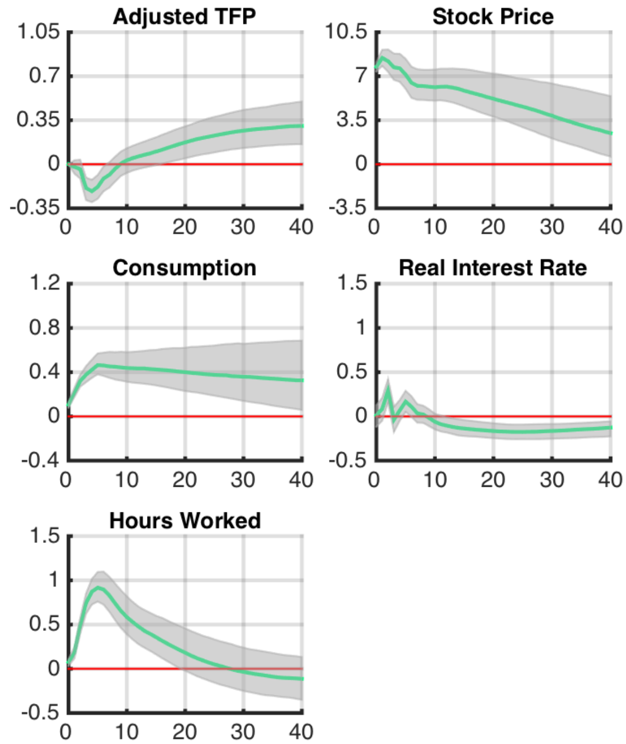


Figure 1: Beaudry, Nam and Wang (2011) Identification 1

prior is used instead the results do not back as strongly the view of those who think of optimism shocks as relevant for business cycle fluctuations. Hence, the results reported in Figure 1 are mostly driven by the additional restrictions imposed by PFA.

3 Our Methodology

This section describes the structural vector autoregression model and the impluse response functions for this model. Several different parameterizations are explored and the transformations between these parameterizations are given. The identification problem in each of the parameterizations is stated.

3.1 The Model

Consider the structural vector autoregression (SVAR) with the general form, as in Rubio-Ramírez, Waggoner and Zha (2010)

$$\mathbf{y}_t'\mathbf{A}_0 = \sum_{\ell=1}^{p} \mathbf{y}_{t-\ell}'\mathbf{A}_\ell + \mathbf{c} + \boldsymbol{\varepsilon}_t' \quad \text{for } 1 \leq t \leq T, \tag{1}$$

where $\mathbf{y}_t$ is an $n\times 1$ vector of endogenous variables, $\boldsymbol{\varepsilon}_t$ is an $n\times 1$ vector of exogenous structural shocks, $\mathbf{A}_\ell$ is an $n\times n$ matrix of parameters for $0 \leq \ell \leq p$ with $\mathbf{A}_0$ invertible, $\mathbf{c}$ is a $1\times n$ vector of parameters, p is the lag length, and T is the sample size. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, ..., \mathbf{y}_{1-p}$, is Gaussian with mean zero and covariance matrix $\mathbf{I}_n$, the $n\times n$ identity matrix. The model described in equation (1) can be written as

$$\mathbf{y}_t'\mathbf{A}_0 = \mathbf{x}_t'\mathbf{A}_+ + \boldsymbol{\varepsilon}_t' \quad \text{for } 1 \leq t \leq T, \tag{2}$$

where $\mathbf{A}_+' = \begin{bmatrix}\mathbf{A}_1' & \cdots & \mathbf{A}_p' & \mathbf{c}'\end{bmatrix}$ and $\mathbf{x}_t' = \begin{bmatrix}\mathbf{y}_{t-1}' & \cdots & \mathbf{y}_{t-p}' & 1\end{bmatrix}$ for $1 \leq t \leq T$. The dimension of $\mathbf{A}_+$ is $m\times n$, where $m = np+1$. The reduced-form representation implied by equation (2) is

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \mathbf{u}_t' \quad \text{for } 1 \leq t \leq T, \tag{3}$$

where $\mathbf{B} = \mathbf{A}_+\mathbf{A}_0^{-1}$, $\mathbf{u}_t' = \boldsymbol{\varepsilon}_t'\mathbf{A}_0^{-1}$, and $\mathbb{E}\left[\mathbf{u}_t\mathbf{u}_t'\right] = \boldsymbol{\Sigma} = \left(\mathbf{A}_0\mathbf{A}_0'\right)^{-1}$. The matrices $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are the reduced-form parameters, while $\mathbf{A}_0$ and $\mathbf{A}_+$ are the structural parameters.

Following Rothenberg (1971), the parameters $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if they imply the same distribution of $\mathbf{y}_t$ for all t. For the linear Gaussian models of the type studied in this paper, this statement is equivalent to saying that $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if they have the same reduced-form representation. This implies the structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{Q}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{Q}$, where $\mathbf{Q}$ is an element of $O(n)$, the set of all orthogonal $n\times n$ matrices.[4] This suggests that in addition to the structural parameterization implicit in equation (2) we can consider an alternative one defined by $\mathbf{B}$, $\boldsymbol{\Sigma}$, and $\mathbf{Q}$ which we label the orthogonal reduced-form parameterization. We now define the mapping that allows us to travel across these two parameterizations.

[4] A square matrix $\mathbf{Q}$ is orthogonal if and only if $\mathbf{Q}^{-1} = \mathbf{Q}'$. Thus, if $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{Q}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{Q}$, then $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ map to the same reduced form parameters. If $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ map to the same reduced form parameters, then $\mathbf{Q} = \tilde{\mathbf{A}}_0^{-1}\mathbf{A}_0$ is orthogonal, $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{Q}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{Q}$.

3.2 The Orthogonal Reduced-Form Parameterization

To make this operational, one must first choose a decomposition of the covariance matrix $\mathbf{\Sigma}$. Let $h(\mathbf{\Sigma})$ be an $n \times n$ matrix which satisfies $h(\mathbf{\Sigma})'h(\mathbf{\Sigma}) = \mathbf{\Sigma}$. For instance, one could choose $h(\mathbf{\Sigma})$ to be the Cholesky decomposition. Alternatively, $h(\mathbf{\Sigma})$ could be the square root of $\mathbf{\Sigma}$, which is the unique symmetric and positive definite matrix whose square is $\mathbf{\Sigma}$. We require h to be differentiable, which both the Cholesky decomposition and the square root function are. Given a decomposition h, we can define the function f_h^{-1} by

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

The orthogonal reduced-form parameterization makes clear how the structural parameters depend on the reduced-form parameters and orthogonal matrices. Given the reduced-form parameters and a decomposition h, one can consider each value of the orthogonal matrix $\mathbf{Q}$ as an particular choice of structural parameters. The function f_h^{-1} is invertible, with inverse defined by f_h

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

3.3 The Impulse Response Function Parameterization

There are many alternative parameterizations of equation (2) that a researcher could consider. Since most of the literature imposes the identifying sign and zero restrictions on the IRFs, it is also natural to consider a third parameterization based on the IRFs.

Definition 1. *Let $(\mathbf{A}_0, \mathbf{A}_+)$ be any value of structural parameters. The IRF of the i^{th} variable to the j^{th} structural shock at horizon k corresponds to the element in row i and column j of the matrix $\mathbf{L}_k(A_0, A_+)$, where $\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+)$ is defined recursively by*

$$\mathbf{L}_0(\mathbf{A}_0, \mathbf{A}_+) = \left(\mathbf{A}_0^{-1}\right)', \tag{4}$$

$$\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+) = \sum_{\ell=1}^{k} \left(\mathbf{A}_\ell\mathbf{A}_0^{-1}\right)' \mathbf{L}_{k-\ell}(\mathbf{A}_0, \mathbf{A}_+), \text{ for } 1 \le k \le p, \tag{5}$$

$$\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+) = \sum_{\ell=1}^{p} \left(\mathbf{A}_\ell\mathbf{A}_0^{-1}\right)' \mathbf{L}_{k-\ell}(\mathbf{A}_0, \mathbf{A}_+), \text{ for } p < k < \infty. \tag{6}$$

If in addition, the SVAR is stationary and the i^{th} variable enters the SVAR in first differences, then the long-run IRF of the i^{th} variable to the j^{th} structural shock corresponds to the element in row i and column j of the matrix $\mathbf{L}_\infty(\mathbf{A}_0, \mathbf{A}_+)$, where

$$\mathbf{L}_\infty(\mathbf{A}_0, \mathbf{A}_+) = \left(\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'\right)^{-1}. \tag{7}$$

An induction argument on k shows that $\mathbf{L}_k(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) = \mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+)\mathbf{Q}$ for $0 \leq k < \infty$ and any $\mathbf{Q} \in O(n)$ and a direct computation shows the same for $k = \infty$. This relationship will be critical in our analysis. We combine the IRFs at horizons one through p and the constant term into a single matrix $\mathbf{L}_+$, so that $\mathbf{L}_+' = \begin{bmatrix}\mathbf{L}_1' & \cdots & \mathbf{L}_p' & \mathbf{c}'\end{bmatrix}$. Note that $\mathbf{L}_0$ is an invertible $n \times n$ matrix while $\mathbf{L}_+$ is an arbitrary $m \times n$ matrix. We call $(\mathbf{L}_0, \mathbf{L}_+)$ the IRF parameterization.

Equations (4) and (5) define a mapping, which we will denote by g, from the structural parameterization to the IRF parameters parameterization. This mapping is invertible and its inverse is given by

$$g^{-1}(\mathbf{L}_0, \mathbf{L}_+) = \Big(\underbrace{(\mathbf{L}_0^{-1})'}_{\mathbf{A}_0}, \underbrace{[\mathbf{A}_1(\mathbf{L}_0, \mathbf{L}_+)' \ \cdots \ \mathbf{A}_p(\mathbf{L}_0, \mathbf{L}_+)' \ \ \mathbf{c}']'}_{\mathbf{A}_+}\Big)$$

where $\mathbf{A}_k(\mathbf{L}_0, \mathbf{L}_+)$ is defined recursively for $1 \leq k \leq p$ by

$$\mathbf{A}_k(\mathbf{L}_0, \mathbf{L}_+) = \left(\mathbf{L}_k\mathbf{L}_0^{-1}\right)' \mathbf{A}_0(\mathbf{L}_0, \mathbf{L}_+) - \sum_{\ell=1}^{k-1} \left(\mathbf{L}_{k-\ell}\mathbf{L}_0^{-1}\right)' \mathbf{A}_\ell(\mathbf{L}_0, \mathbf{L}_+).$$

As with the structural parameterization, the IRFs $(\mathbf{L}_0, \mathbf{L}_+)$ and $(\tilde{\mathbf{L}}_0, \tilde{\mathbf{L}}_+)$ are observationally equivalent if and only if $\mathbf{L}_0 = \tilde{\mathbf{L}}_0\mathbf{Q}$ and $\mathbf{L}_+ = \tilde{\mathbf{L}}_+\mathbf{Q}$, where $\mathbf{Q} \in O(n)$.

3.3.1 The Identification Problem and Sign and Zero Restrictions

We have considered three parameterizations of equation (2), but none of these parameterizations is identified. To solve the identification problem, one often imposes sign and/or zero restrictions on either the structural parameters or some function of the structural parameters, like the IRFs. For instance, a common identification scheme is to assume the matrix $\mathbf{A}_0$ is triangular with positive diagonal. The theory and drawing techniques developed will apply to sign and zero restrictions on any function $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ from the space of structural parameters to the space of $nr \times n$ matrices that satisfies the condition $\mathbf{F}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) = \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{Q}$ for all $\mathbf{Q} \in O(n)$.[5] As we saw in the previous section, the impulse responses $\mathbf{L}_k(\mathbf{A_0}, \mathbf{A}_+)$ satisfy this requirement for $0 \leq k \leq \infty$.

In Rubio-Ramírez, Waggoner and Zha (2010), sufficient conditions for identification are developed. The sufficient condition for identification is that there must be an ordering of the structural shocks so that there are at least $n-j$ zero restrictions on j^{th} shock, for $1 \leq j \leq n$. In addition, there must be at least one sign restriction on the impulse responses to each shock.[6] It is also the case that Rothenberg (1971) order necessary condition for identification is that the number of zero restrictions is greater than or equal to $n(n-1)/2$. In this paper, we will have fewer than $n-j$ zero restrictions on the j^{th} shock. This means that the identification will be a set identification. However, if there are enough sign restrictions, then the identified sets will be small and it will be possible to draw meaningful economic conclusions.

For $1 \leq j \leq n$, let $\mathbf{S}_j$ be a $s_j \times nr$ matrix of full row rank where $0 \leq s_j$ and let $\mathbf{Z}_j$ be a $z_j \times nr$ matrix of full row rank where $0 \leq z_j \leq n-j$. The $\mathbf{S}_j$ will define the sign restrictions on the j^{th} shock and the $\mathbf{Z}_j$ will define the zero restrictions on the j^{th} shock. In particular, we assume for $1 \leq j \leq n$ that

$$\mathbf{S}_j\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j > \mathbf{0} \text{ and } \mathbf{Z}_j\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j = \mathbf{0},$$

where $\mathbf{e}_j$ is the j^{th} column of $\mathbf{I}_n$.

3.4 Agnostic Priors and Posteriors

For any of the three parameterizations considered here, we say that a prior is agnostic with respect to identification if the prior density is equal across observationally equivalent parameters. If general, a researcher wants identification to only come from the sign and zero restrictions. This will be true if one details an agnostic prior and then imposes the sign and zero restrictions. If the prior is not agnostic, then identification will be a influenced by the prior in addition to the sign and zero restrictions. Often, economic arguments are made that each sign and zero restriction must hold. To the extent that these arguments are successful, it implies that if one details an agnostic prior, then the identification will

[5] In addition, a regularity condition on $\mathbf{F}$ is needed. For instance, it suffices to assume that $\mathbf{F}$ is differentiable and that its derivative is of full row rank, though this condition could be weakened somewhat.

[6] Often, one is only interested in partial identification. If there is an ordering such that there are at least $n-j$ zero restrictions and at least one sign restriction on the impluse responses to j^{th} shock for $1 \leq j \leq k$, then the first k shocks under this ordering will be identified.

be justified on economic grounds. This is not to say that one cannot justify detailing a non-agnostic prior, only that it requires more than just arguing that each sign and zero restriction holds because it imposes additional identification restrictions than the sign and zeros.

For a prior $\hat{\pi}$ over the orthogonal form parameterization, an agnostic prior must be flat over $O(n)$. For a prior π over the structural parameterization, an agnostic prior must be that $\pi(\mathbf{A}_0, \mathbf{A}_+) = \pi(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})$ for every $\mathbf{Q} \in O(n)$. Many of the priors over the structural parameterization used in the literature are agnostic, for instance, any prior that can be implemented via dummy observations will be agnostic, which includes the Sims-Zha prior. For a prior $\tilde{\pi}$ over the IRF parameterization, an agnostic prior must be that $\pi(\mathbf{L}_0, \mathbf{L}_+) = \pi(\mathbf{L}_0\mathbf{Q}, \mathbf{L}_+\mathbf{Q})$ for every $\mathbf{Q} \in O(n)$.

In a similar fashion, we say that a posterior is agnostic with respect to identification if the posterior density is equal across observationally equivalent parameters. Thus an agnostic posterior over the orthogonal reduced-form parameterization must be flat over $O(n)$. Because the likelihood is equal across observationally equivalent parameters, it easy to show that conjugate priors are agnostic and that a prior is agnostic if and only if the implied posterior is agnostic.

3.5 Transforming Priors and Posteriors

Let $p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+)$ be the likelihood function over the structural parameterization, where $\mathbf{Y}_T = (\mathbf{y}_1, \cdots, \mathbf{y}_T)$. Then the posterior over the structural parameterization is $p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+)\pi(\mathbf{A}_0, \mathbf{A}+)/p(\mathbf{Y}_t)$, where $p(\mathbf{Y}_T) = \int_{(\mathbf{A}_0,\mathbf{A}+)} p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+)\pi(\mathbf{A}_0, \mathbf{A}+)d(\mathbf{A}_0, \mathbf{A}+)$. Equivalently, we can define $\hat{p}(\mathbf{Y}_T|\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ and $\hat{p}(\mathbf{Y}_T|\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})/\hat{p}(\mathbf{Y}_t)$ to be the likelihood function and the posterior distribution over the orthogonal reduced-form parameterization and $\tilde{p}(\mathbf{Y}_T|\mathbf{L}_0, \mathbf{L}_+)$ and $\tilde{p}(\mathbf{Y}_T|\mathbf{L}_0, \mathbf{L}_+)\tilde{\pi}(\mathbf{L}_0, \mathbf{L}_+)/\tilde{p}(\mathbf{Y}_t)$ to be the likelihood function and the posterior distribution over the IRF parameterization, where $\hat{p}(\mathbf{Y}_T) = \int_{(\mathbf{A}_0,\mathbf{A}+)} \hat{p}(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+)\hat{\pi}(\mathbf{A}_0, \mathbf{A}+)d(\mathbf{A}_0, \mathbf{A}+)$ and $\tilde{p}(\mathbf{Y}_T) = \int_{(\mathbf{A}_0,\mathbf{A}+)} \tilde{p}(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+)\tilde{\pi}(\mathbf{A}_0, \mathbf{A}+)d(\mathbf{A}_0, \mathbf{A}+)$ respectively.

We have defined three different parameterizations for SVARs and gave mappings between these parameterizations. These mappings will allow us to transform priors and posteriors over one parameterization into equivalent priors over another parameterization. Since $p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}+) = \hat{p}(\mathbf{Y}_T|f_h(\mathbf{A}_0, \mathbf{A}+)) = \tilde{p}(\mathbf{Y}_T|g(\mathbf{A}_0, \mathbf{A}+))$, to transform the posteriors across parameterizations we only need to transform the priors. The transformation is done using the change of variable theorem. The usual statement of this theorem is given below.

Theorem 1. *Let U be an open subset of $\mathbb{R}^b$ and let $\alpha : U \to \mathbb{R}^b$ be a differentiable and one-to-one* *function such that* $|\det(D\alpha(\mathbf{u}))| \neq 0$ *for every* $\mathbf{u} \in U$*, where* $D\alpha(\mathbf{u})$ *denotes the derivative of* α *evaluated at* $\mathbf{u}$*. Then for every measurable set* $W \subset U$ *and for every measurable function* $\lambda : U \to \mathbb{R}$

$$\int_W \lambda(\mathbf{u}) d\mathbf{u} = \int_{\alpha(W)} \frac{\lambda(\alpha^{-1}(\mathbf{v}))}{|\det(D\alpha(\alpha^{-1}(\mathbf{v})|} d\mathbf{v}.$$

Proof. See Spivak (1965). □

We can use Theorem 1 to transform priors over the structural parameterization into equivalent priors over the IRF parameterization. For instance, if $\pi(\mathbf{A}_0, \mathbf{A}_+)$ is a prior over the structural parameterization and $(p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}_+)\pi(\mathbf{A}_0, \mathbf{A}_+))\,/p(\mathbf{Y}_t)$ is the implied posterior over the structural parameterization, then $\tilde{\pi}(\mathbf{L}_0, \mathbf{L}_+) = \pi(g^{-1}(\mathbf{L}_0, \mathbf{L}_+))/|\det(Dg(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))|$ is the equivalent prior over the IRF parameterization and

$$\frac{p(\mathbf{Y}_T|g^{-1}(\mathbf{L}_0, \mathbf{L}_+))\pi(g^{-1}(\mathbf{L}_0, \mathbf{L}_+))}{p(\mathbf{Y}_t)|\det(Dg(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))|} \tag{8}$$

is the equivalent posterior over the IRF parameterization, where we have use the fact that $\tilde{p}(\mathbf{Y}_t) = p(\mathbf{Y}_t)$ when $\tilde{\pi}$ is equivalent to π over the IRF parameterization. The following theorem computes $|\det(Dg(\mathbf{A}_0, \mathbf{A}_+))|$.

Proposition 1. *Let* g *be defined as in Section 3.3, then we have*

$$|\det(Dg(\mathbf{A}_0, \mathbf{A}_+))| = |\det(\mathbf{A}_0))|^{-2n(p+1)}.$$

Since $|\det(Dg(\mathbf{A}_0, \mathbf{A}_+))| = |\det(Dg(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}))|$ for every $\mathbf{Q} \in O(n)$, an immediate consequence of Proposition 1 is the following corollary.

Corollary 1. *A prior over the structural parameterization is agnostic if and only if it is equivalent to an agnostic prior over the IRF parameterization.*

We can use the same strategy to transform priors over the structural parameterization into equivalent priors over the orthogonal reduced-form parameterization. But we cannot use Theorem 1 because the domain of f_h is of dimension $n(n+m)$ while the range is of dimension $n(2n+m)$. The following theorem generalizes Theorem 1 to handle this case.

Theorem 2. *Let* U *an open subset of* $\mathbb{R}^b$ *and let* $\alpha : U \to \mathbb{R}^a$ *be a differentiable and one-to-one function such that* $|\det(D\alpha(\mathbf{u})'D\alpha(\mathbf{u}))| \neq 0$ *for every* $\mathbf{u} \in U$ *and* α^{-1} *is continuous. Then for every* *measurable set $W \subset U$ and for every measurable function $\lambda : U \to \mathbb{R}$*

$$\int_W \lambda(\mathbf{u}) d\mathbf{u} = \int_{\alpha(W)} \frac{\lambda(\alpha^{-1}(\mathbf{v}))}{|\det(D\alpha(\alpha^{-1}(\mathbf{v}))' D\alpha(\alpha^{-1}(\mathbf{v})))|^{1/2}} d\mathbf{v}$$

Proof. See Spivak (1965). □

To simplify notation, we denote $v_\alpha(\mathbf{u}) = |\det(D\alpha(\mathbf{u})' D\alpha(\mathbf{u}))|^{1/2}$, where $v_\alpha(\mathbf{u})$ is usually known as the volume element. We can use Theorem 2 to transform priors over the structural parameterization into equivalent priors over the orthogonal reduced-form parameterization. If $\pi(\mathbf{A}_0, \mathbf{A}_+)$ is a prior over the structural parameterization and $\left(p(\mathbf{Y}_T|\mathbf{A}_0, \mathbf{A}_+)\pi(\mathbf{A}_0, \mathbf{A}_+)\right)/p(\mathbf{Y}_t)$ is the implied posterior over the structural parameterization, then $\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = \pi(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))/v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))$ is the equivalent prior over the orthogonal reduced-form parameterization and

$$\frac{p(\mathbf{Y}_T|f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))\pi(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}{p(\mathbf{Y}_t)v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))} \tag{9}$$

is the equivalent posterior over the orthogonal reduced-form parameterization, where we have use the fact that $\hat{p}(\mathbf{Y}_t) = p(\mathbf{Y}_t)$ when $\hat{\pi}$ is equivalent to π over the orthogonal reduced-form parameterization. The following theorem computes $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)$.

Proposition 2. *Let f_h be defined as in Section 3.2, then we have*

$$v_{f_h}(\mathbf{A}_0, \mathbf{A}_+) = 2^{\frac{n(n+1)}{2}} |\det(\mathbf{A}_0)|^{-2(\frac{np}{2}+n+1)}.$$

Since $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+) = v_{f_h}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})$ for every $\mathbf{Q} \in O(n)$, an immediate consequence of Proposition 2 is the following corollary.

Corollary 2. *A prior over the structural parameterization is agnostic if and only if it is equivalent to an agnostic prior over the orthogonal reduced-form parameterization.*

Corollaries 1 and 2 tell us that if a researcher details an agnostic prior over one of the three parameterizations the equivalent prior over any of the other two parameterizations is also agnostic. This means that if one only cares about the prior being agnostic, the choice of parameterization is irrelevant and the transformation of any agnostic prior into the equivalent agnostic prior over the orthogonal reduced-form parameterization has a very particular form.

Corollary 3. *A prior is agnostic if and only if it is equivalent to a prior over the orthogonal reduced-form parameterization that it is flat over $O(n)$.*

Obviously, Corollaries 1, 2, and 3 can be written in terms of posteriors and, therefore, a posterior is agnostic if and only if it is equivalent to a posterior posterior over the orthogonal reduced-form parameterization that it is flat over $O(n)$.

4 Sign Restrictions

As argued in Section 3.3.1, sign restrictions imply set identification. But it also the case that the set of structural parameters satisfying the sign restrictions will be an open set of positive measure in the set of all structural parameters. This is easy to see because $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ is differentiable. Thus, algorithms of the following type are justified to draw from an agnostic posterior over the structural parameterization conditional on the sign restrictions.

Algorithm 1.

1. *Draw $(\mathbf{A}_0, \mathbf{A}_+)$ from the agnostic posterior.*
2. *Keep the draw if the sign restrictions are satisfied.*
3. *Return to Step 1 until the required number of draws from the posterior on the structural parameters conditional on the sign restrictions has been obtained.*

Note that we have written Algorithm 1 in terms of the agnostic posterior over the structural parameterization. But Propositions 1 and 2 show us how to transform priors and posteriors across parameterizations. Hence, if a researcher details the agnostic prior over either the IRF or the orthogonal reduced-form parameterizations instead of detailing it over the orthogonal parameterization, one could either

- consider a variation of Algorithm 1 where one draws from the agnostic posterior over the chosen parameterization and transforms the draws into the structural parameterization using g^{-1} or f_h^{-1} to check whether the sign restrictions are satisfied or
- use the results in Section 3.5 to find the equivalent agnostic prior and posterior over the structural parameterization and use directly Algorithm 1.

As argued by Baumeister and Hamilton (2015a) one should preferable detail priors over either the structural or the IRF parameterizations, since the orthogonal reduced-form parameterization is hard to interpret from an economic point of view. But if one only cares about the prior being agnostic, it may be preferable to detail a conjugate prior over the orthogonal reduced-form parameterization and draw from conjugate posterior over the orthogonal reduced-form parameterization and transform the draws using f_h^{-1}. This approach is convenient because it facilitates drawing from the posterior. If one only cares about the prior being agnostic, this approach is valid because the equivalent prior and posterior over either the structural or the IRF parameterizations are also agnostic.

The conjugate posterior is normal-inverse-Wishart over the reduced-form parameters and flat over the set of orthogonal matrices. It is extremely easy and efficient to draw from the normal-inverse-Wishart posterior over the reduced-form parameters. All that remains is to obtain an efficient algorithm for independently drawing from the flat prior over $O(n)$. Faust (1998), Canova and Nicoló (2002), Uhlig (2005), and Rubio-Ramírez, Waggoner and Zha (2010) propose algorithms to do that. However, Rubio-Ramírez, Waggoner and Zha's (2010) algorithm is the only computationally feasible one for moderately large SVAR systems (e.g., $n > 4$).[7] Rubio-Ramírez, Waggoner and Zha's (2010) results are based on the following theorem.

Theorem 3. *Let* $\mathbf{X}$ *be an* $n \times n$ *random matrix with each element having an independent standard normal distribution. Let* $\mathbf{X} = \mathbf{QR}$ *be the QR decomposition of* $\mathbf{X}$ *with the diagonal of* $\mathbf{R}$ *normalized to be positive. The random matrix* $\mathbf{Q}$ *is orthogonal and is a draw from the flat prior over* $O(n)$.

Proof. The proof follows directly from Stewart (1980). □

The following algorithm uses Theorem 3 together with a variation of Algorithm 1 to independently draw from an agnostic posterior on the structural parameterization conditional on the sign restrictions.

Algorithm 2.

1. *Draw* $(\mathbf{B}, \mathbf{\Sigma})$ *from the posterior distribution of the reduced-form parameters.*

2. *Use Theorem 3 to draw from the flat prior over* $O(n)$.

3. *Set* $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.

4. *Keep the draw if the sign restrictions are satisfied.*

[7]See Rubio-Ramírez, Waggoner and Zha (2010) for details.

5. Return to Step 1 until the required number of draws from the posterior on the structural parameters conditional on the sign restrictions has been obtained.

The results obtained with Algorithm 2 should be identical to the ones that would be obtained with Algorithm 1 when used with the equivalent agnostic prior over either the structural or the IRF parameterizations implied by the particular choice of Normal-Inverse Wishart prior distribution over the reduced-form parameters. We choose the orthogonal reduced-form parameterization out of convenience, since drawing from the conjugate posterior is is extremely easy and efficient. Our algorithm can be easily adapted to draws from the agnostic posterior over either the structural or the IRF parameterizations.

4.1 Flat Priors

Baumeister and Hamilton (2015a) highlights some risks of using Algorithm 2. They point out that a prior over the orthogonal reduced-form parameterization that is flat over $O(n)$ may imply an equivalent non-flat prior over any of the other two parameterizations. This is an important concern because a researcher could be interested in having flat priors over either the structural or the IRF parameterizations. While this is true, it is also the case that any flat prior over either the structural or the IRF parameterizations has to be agnostic and, therefore, it is equivalent to a prior over the orthogonal reduced-form parameterization it is flat over $O(n)$. In other words, if a prior over the structural parameterization implies an equivalent prior over the orthogonal reduced-form parameterization that it is not flat over $O(n)$, then it cannot be flat over the structural parameterization. The same is true for the IRF parameterization. This can be seen in the following corollary.

Corollary 4. *If a prior is flat over either the structural or the IRF parameterizations then the equivalent over the orthogonal reduced-form parameterization is flat over $O(n)$.*

Thus, it is possible to use Algorithm 2 when one wants to work with flat priors over either the structural or the IRF parameterizations. The problem is not the flat over $O(n)$ but the prior over the reduced-form parameters. A researcher can detail a prior over the orthogonal reduced-form parameterization that is flat over $O(n)$ and implies an equivalent prior over the structural parameterization that is flat. One just need to detail the right prior over the reduced-form parameters. The same is true for the IRF parameterization, but the prior over the reduced-form parameters is different. This can be seeing in the following two corollaries that are related to Propositions 1 and 2.

Corollary 5. *A prior over the structural parameterization is flat if and only if the equivalent prior over the orthogonal reduced-form parameterization equals to* $\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{\frac{-n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-(\frac{np}{2}+n+1)}$.

Corollary 6. *A prior over the IRF parameterization is flat if and only if the equivalent prior over the orthogonal reduced-form parameterization equals to* $\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{\frac{-n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{np}{2}-1}$.

In line with the above discussion, Corollaries 5 and 6 imply that in both cases the priors over the orthogonal reduced-form parameterization are flat over $O(n)$. It is important to notice that if one details the prior over the reduced-form parameters so that the equivalent prior over the structural parameterization is flat, the equivalent prior over the IRF parameterization is not flat. It is also the case that if one details the prior over the reduced-form parameters so that the equivalent prior over the IRF parameterization is flat, the equivalent prior over the structural parameterization is not flat. It is also obvious that if the researcher details the prior over the reduced-form parameters so that the prior over the orthogonal reduced-form parameterization is flat, then the equivalent prior over neither structural nor IRF parameterizations is flat. Thus, if the researcher cares about the prior being flat the parameterization does matters. The researcher needs to choose over which parameterization to be flat.

Priors detailed in Corollaries 5 and 6 can both be used together with Algorithm 2 because they imply a Normal-Inverse Wishart posteriors of the reduced-form parameters, hence Step 1 can be easily applied. This is can be seen in the next two propositions.

Proposition 3. *If* $\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{\frac{-n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{-(\frac{np}{2}+n+1)}$*, the Normal-Inverse Wishart for the posterior of the reduced-form parameters is defined by*

$$p\left(\mathbf{B} \mid \mathbf{Y}, \mathbf{X}_{+}, \mathbf{\Sigma}\right) \propto |\mathbf{\Sigma}|^{-\frac{p}{2}} exp\left\{-\frac{1}{2} vec(\mathbf{B}-\hat{\mathbf{B}})' \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{X}_{+}'\mathbf{X}_{+}\right) vec(\mathbf{B}-\hat{\mathbf{B}})\right\} \text{ and}$$

$$p\left(\mathbf{\Sigma} \mid \mathbf{Y}, \mathbf{X}_{+}\right) \propto |\mathbf{\Sigma}|^{-\frac{\nu+n+1}{2}} exp\left\{-\frac{1}{2} tr\left(T-p\right) \mathbf{S}\mathbf{\Sigma}^{-1}\right\}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]'$, $\mathbf{X}_{+} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]'$, $\hat{\mathbf{B}} = \left(\mathbf{X}_{+}'\mathbf{X}_{+}\right)^{-1} \mathbf{X}_{+}'\mathbf{Y}$, $\mathbf{S} = (T-p)^{-1} (\mathbf{Y}-\mathbf{X}_{+}\hat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}_{+}\hat{\mathbf{B}})$*, and* $\nu = T - p(1-n) + n + 1$.

Proof. See DeJong (1992). □

Proposition 4. *If* $\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{\frac{-n(n+1)}{2}} |\det(\mathbf{\Sigma})|^{\frac{np}{2}-1}$*, the Normal-Inverse Wishart for the posterior* *of the reduced-form parameters is*

$$p\left(\mathbf{B} \mid \mathbf{Y}, \mathbf{X}_+, \mathbf{\Sigma}\right) \propto |\mathbf{\Sigma}|^{-\frac{p}{2}}\, exp\left\{-\frac{1}{2}vec(\mathbf{B}-\hat{\mathbf{B}})'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{X}_+'\mathbf{X}_+\right)vec(\mathbf{B}-\hat{\mathbf{B}})\right\} \text{ and}$$

$$p\left(\mathbf{\Sigma} \mid \mathbf{Y}, \mathbf{X}_+\right) \propto |\mathbf{\Sigma}|^{-\frac{\nu+n+1}{2}}\, exp\left\{-\frac{1}{2}tr(T-p)\mathbf{S}\mathbf{\Sigma}^{-1}\right\}$$

where $\nu = T - p(1+n) - n + 1$.

Proof. See DeJong (1992). □

5 Zero Restrictions

We now adapt Algorithm 2 to handle the case of sign and zero restrictions. We will first introduce the concept of smooth manifolds. Then we will show that the zero restrictions define a smooth manifold of dimension lower than $n(n+m)$ in the set of all structural parameters. Because such smooth manifold is of measure zero in the set of all structural parameters, this invalidates the direct use of Algorithm 2 when zero restrictions are considered. While Algorithm 2 implicitly draws from agnostic posterior over the structural parameterization, we will now need to implicitly draw $(\mathbf{A}_0, \mathbf{A}_+)$ from the conditionally agnostic posterior over the structural parameterization.

5.1 Smooth Manifolds

In previous sections we have worked with smooth manifolds, though not by name. For instance, $O(n)$ is a $\frac{n(n-1)}{2}$-dimensional smooth manifold in $\mathbb{R}^{n^2}$ and the space of all symmetric and positive definite $n \times n$ matrices is a $\frac{n(n+1)}{2}$-dimensional smooth manifold in $\mathbb{R}^{n^2}$. In this section, when we impose zero restrictions on the structural parameters, the space of all parameters satisfying the restrictions will be a smooth manifold in $\mathbb{R}^{n(n+m)}$. We now formally define a smooth manifold and state some standard theorems for working with them.

Definition 2. $\mathcal{M} \subset \mathbb{R}^b$ *is an d-dimensional smooth manifold in* $\mathbb{R}^b$ *if and and only if for every* $\mathbf{x} \in \mathcal{M}$ *there exists a differentiable function* $\gamma : U \to \mathbb{R}^b$ *such that*

1. U *is an open subset of* $\mathbb{R}^d$, $\gamma(U)$ *is an open subset of* $\mathcal{M}$, *and* $\mathbf{x} \in \gamma(U)$.

2. $\gamma^{-1} : \gamma(U) \to U$ *exists and is continuous.*

3. The $b \times d$ matrix $D\gamma(\mathbf{u})$ is of rank d for every $\mathbf{u} \in U$.

The function γ is a coordinate system in $\mathcal{M}$ about $\mathbf{x}$.

The following theorem shows that zero restrictions lead to smooth manifolds.[8]

Theorem 4. *Let U be an open subset of $\mathbb{R}^b$ and let $\beta : U \to \mathbb{R}^{b-d}$ be a differentiable function such that $D\beta(\mathbf{u})$ is of rank $b-d$ whenever $\beta(\mathbf{u}) = 0$, then $\mathcal{M} = \{\mathbf{u} \in U | \beta(\mathbf{u}) = 0\}$ will be a d-dimensional smooth manifold in $\mathbb{R}^b$.*

Proof. Follows directly from the implicit function theorem. See Spivak (1965), Theorems 5-1 and 5-2 for details. □

5.2 Zero Restrictions as Smooth Manifolds

This theorem can be used to show that both $O(n)$ and the space of symmetric and positive definite matrices are smooth manifolds. The space $O(n)$ is the set of zeros of the function defined by $\beta(\mathbf{X}) = (\mathbf{e}_i'\mathbf{X}'\mathbf{X}\mathbf{e}_j - \delta_{i,j})_{1 \leq i \leq j \leq n}$, where $\mathbf{X}$ is any $n \times n$ matrix. Here, $\delta_{i,j}$ denotes the function that is one if $i = j$ and zero otherwise. Since there are $\frac{n(n+1)}{2}$ restrictions, the dimension of $O(n)$ will be $\frac{n(n-1)}{2}$.

The space of symmetric $n \times n$ matrices is a linear space and is the set of zeros of the function defined by $\beta(\mathbf{X}) = (\mathbf{e}_i\mathbf{X}\mathbf{e}_j - \mathbf{e}_j\mathbf{X}\mathbf{e}_i)_{1 \leq i < j \leq n}$, where $\mathbf{X}$ is any $n \times n$ matrix. The set of positive definite matrices is open in the space of symmetric matrices. Since a open subset of a smooth manifold is a smooth manifold, the space of symmetric and positive matrices will be a smooth manifold. Because

[8] Smooth manifolds should be thought of as "locally Euclidian" spaces, with the Euclidian structure given by the coordinate systems γ. This local Euclidian structure allows one to extend many of the concepts defined on Euclidian spaces to smooth manifolds. In particular, integration on smooth manifolds can be defined in this manner. If $\mathcal{M}$ is a smooth manifold, $\gamma : U \to \mathbb{R}^b$ is a coordinate system in $\mathcal{M}$, $\lambda : \mathcal{M} \to \mathbb{R}$ is any measurable function, and W is any measurable set in $\gamma(U) \subset \mathcal{M}$, then we can define

$$\int_W \lambda(\mathbf{w}) d\mathbf{w} \equiv \int_{\gamma^{-1}(W)} \lambda(\gamma(\mathbf{u})) v_\gamma(\mathbf{u}) d\mathbf{u},$$

where v_γ is the volume element defined in Theorem 2. A subset W of a d-dimensional manifold is measurable if and only if $\gamma^{-1}(W)$ is measurable in $\mathbb{R}^d$ for every coordinate system γ. A function $\lambda : \mathcal{M} \to \mathbb{R}$ is measurable if and only if $\lambda^{-1}(V)$ is measurable in $\mathcal{M}$ for every V that is measurable in $\mathbb{R}$. It is not hard to show that the definition of the integral is independent of the coordinate system γ. If $\gamma : U \to \mathbb{R}^b$ and $\tilde{\gamma} : \tilde{U} \to \mathbb{R}^b$ are two coordinate systems $\mathcal{M}$ and $\mathbf{y} \in \gamma(U) \cap \tilde{\gamma}(\tilde{U})$, then $\gamma \circ \tilde{\gamma}^{-1}$ is a differentiable function from an open subset of $\mathbb{R}^d$ onto an open subset of $\mathbb{R}^d$. See the discussion and proof of Theorem 5-2 in Spivak (1965). Thus, the chain rule implies that $v_\gamma(\gamma^{-1}(\mathbf{y})) = |\det(D(\gamma \circ \tilde{\gamma}^{-1})(\mathbf{y}))| v_{\tilde{\gamma}}(\tilde{\gamma}^{-1}(\mathbf{y}))$ and Theorem 1 then implies that the definition of integration on a manifold is independent of the choice of coordinate system. Since any measurable set in $\mathcal{M}$ is equal to a disjoint union of countably many measurable sets, each of which is contained in some coordinate system in $\mathcal{M}$, this definition can be extended to arbitrary measurable sets in $\mathcal{M}$. Thus, Theorem 2 is a direct consequence of the definition of integration on smooth manifolds.

there are $\frac{n(n-1)}{2}$ restrictions, the dimension of the set of symmetric and positive definite matrices will be of dimension $\frac{n(n+1)}{2}$.

But it can also be shown that the zero restrictions are also a smooth manifold. The zero restrictions that we will impose are of the form $\beta(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{0}$, where $\beta(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{Z}_i \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j)_{1 \leq j \leq n}$. Note that β is a function from an open subset of $\mathbb{R}^{n(n+m)}$ into $\mathbb{R}^{\sum_{j=1}^{n} z_j}$. As long as the derivative of β is of full row rank, Theorem 4 implies the zeros of β will define a $(n(n+m) - \sum_{j=1}^{n} z_j)$-dimensional smooth manifold in $\mathbb{R}^{n(n+m)}$. In all the applications in this paper, the derivative of $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ is of full row rank, which is sufficient to ensure that the derivative of β is of full row rank.

5.3 The Change of Variables Theorem for Smooth Manifolds

We are now in a position to state the change of variables theorem for smooth manifolds.

Theorem 5. *Let U be an open set in $\mathbb{R}^b$, let $\alpha : U \to \mathbb{R}^a$ be a differentiable function, and let $\beta : U \to \mathbb{R}^{b-d}$ be a differentiable function such that $D\beta(\mathbf{u})$ is of full row rank whenever $\beta(\mathbf{u}) = 0$. Let $\mathcal{M} = \{\mathbf{u} \in U | \beta(\mathbf{u}) = 0\}$, let $\alpha|_{\mathcal{M}}$ denote the function α restricted to $\mathcal{M}$, and define $v_{\alpha|_{\mathcal{M}}}(\mathbf{u})$ by*

$$v_{\alpha|_{\mathcal{M}}}(\mathbf{u}) = \left|\det(\mathbf{N}'_{\mathbf{u}} \cdot D\alpha(\mathbf{u})' \cdot D\alpha(\mathbf{u}) \cdot \mathbf{N}_{\mathbf{u}})\right|^{\frac{1}{2}}, \tag{10}$$

for any $b \times d$ matrix $\mathbf{N}_{\mathbf{u}}$ whose columns form an orthonormal basis for the null space of $D\beta(\mathbf{u})$. If $v_{\alpha|_{\mathcal{M}}}(\mathbf{u}) \neq 0$ for every $\mathbf{u} \in \mathcal{M}$ and $\alpha|_{\mathcal{M}}^{-1}$ exists and is continuous, then for every measurable set $W \subset \mathcal{M}$ and for every measurable function $\lambda : \mathcal{M} \to \mathbb{R}$

$$\int_W \lambda(\mathbf{u}) d\mathbf{u} = \int_{\alpha(W)} \frac{\lambda(\alpha|_{\mathcal{M}}^{-1}(\mathbf{v}))}{v_{\alpha|_{\mathcal{M}}}(\alpha|_{\mathcal{M}}^{-1}(\mathbf{v}))} d\mathbf{v} = . \tag{11}$$

Proof. By Theorem 4, $\mathcal{M}$ is a smooth d-dimensional manifold in $\mathbb{R}^b$. From the definition of $\mathcal{M}$, for every coordinate system $\gamma : V \to \mathbb{R}^b$ in $\mathcal{M}$ we have that $\mathbf{0} = \beta(\gamma(\mathbf{v}))$ for every $\mathbf{v} \in V$. Thus, $\mathbf{0} = D(\beta \circ \gamma)(\mathbf{v}) = D\beta(\gamma(\mathbf{v})) \cdot D\gamma(\mathbf{v})$. So, if $\mathbf{N}_{\gamma(\mathbf{v})}$ is any $b \times d$ matrix whose columns form an orthonormal basis for the null space of $D\beta(\gamma(\mathbf{v}))$, then it must be the case that $D\gamma(\mathbf{v}) = \mathbf{N}_{\gamma(\mathbf{v})}\mathbf{X}$, for some invertiable $d \times d$ matrix $\mathbf{X}$. Because $v_{\alpha|_{\mathcal{M}}}(\mathbf{u}) \neq 0$ for every $\mathbf{u} \in \mathcal{M}$ and $\alpha|_{\mathcal{M}}^{-1}$ is continuous, $\alpha(\mathcal{M})$ will be a smooth d-dimensional manifold and $\alpha \circ \gamma$ will be a coordinate system in $\alpha(\mathcal{M})$. It suffices to prove the result for $W \subset \alpha(\gamma(V)) \subset \alpha(\mathcal{M})$. If $W \subset \alpha(\gamma(V))$, then

$$\int_W \lambda(\mathbf{w})d\mathbf{w} \equiv \int_{(\alpha\circ\gamma)^{-1}(W)} \lambda(\alpha \circ \gamma(\mathbf{v}))v_{\alpha\circ\gamma}(\mathbf{v})d\mathbf{v}$$

and

$$\int_{\alpha^{-1}(W)} \lambda(\alpha(\mathbf{u}))v_{\alpha|_{\mathcal{M}}}(\mathbf{u})d\mathbf{u} \equiv \int_{\gamma^{-1}(\alpha^{-1}(W))} \lambda(\alpha(\gamma(v)))v_{\alpha|_{\mathcal{M}}}(\gamma(\mathbf{v}))v_{\gamma}(\mathbf{v})d\mathbf{v}.$$

The result follows from the fact that $v_{\alpha\circ\gamma}(\mathbf{v}) = v_{\alpha|_{\mathcal{M}}}(\gamma(\mathbf{v}))v_{\gamma}(\mathbf{v})$. □

5.4 Conditionally Agnostic Priors and Posteriors

For any of the three parameterizations considered here, we say that a prior is conditionally agnostic with respect to identification if the prior density is equal across observationally equivalent parameters that satisfy the zero restrictions. In general, a researcher wants identification to only come from the sign and zero restrictions. This will be true if one details a conditionally agnostic prior and then imposes the sign restrictions. If the prior is not conditionally agnostic, then identification will be a influenced by the prior in addition to the sign and zero restrictions. In a similar fashion, we say that a posterior is conditionally agnostic with respect to identification if the posterior density is equal across observationally equivalent parameters that satisfy the zero restrictions. Because the likelihood is equal across observationally equivalent parameters, it easy to show that a prior is conditionally agnostic if and only if the implied posterior is conditionally agnostic.

5.5 Transforming Priors and Posteriors Conditional on Zero Restrictions

Let Z denote the set of all $(\mathbf{A}_0, \mathbf{A}_+)$ satisfying the zero restrictions. If π is a agnostic prior over the structural parametrization, then the prior over the structural parametrization conditional on $(\mathbf{A}_0, \mathbf{A}_+)$ satisfying the zero restrictions is

$$\pi(\mathbf{A}_0, \mathbf{A}_+|Z) = \frac{\pi(\mathbf{A}_0, \mathbf{A}_+)}{\mu_Z}$$

where $\mu_Z = \int_Z \pi(\mathbf{A}_0, \mathbf{A}_+)d(\mathbf{A}_0, \mathbf{A}_+)$. Note that we use the simpler notation $\pi(\mathbf{A}_0, \mathbf{A}_+|Z)$ as opposed to the more cumbersome $\pi(\mathbf{A}_0, \mathbf{A}_+|(\mathbf{A}_0, \mathbf{A}_+) \in Z)$. Since it is the case that $\pi(\mathbf{A}_0, \mathbf{A}_+|Z) = \pi(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}|Z)$ for every $\mathbf{Q} \in O(n)$, $(\mathbf{A}_0, \mathbf{A}_+) \in Z$, and $(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) \in Z$, then $\pi(\mathbf{A}_0, \mathbf{A}_+|Z)$ is conditionally agnostic.

Using Theorem 5, the equivalent prior to $\pi(\mathbf{A}_0, \mathbf{A}_+|Z)$ over the orthogonal reduced-form parameterization conditional on $(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ satisfying the zero restrictions is

$$\frac{\pi(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}{\mu_Z v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}. \tag{12}$$

In general, $v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})) \neq v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}'))$ for every $\mathbf{Q},\mathbf{Q}' \in O(n)$ such that $\mathbf{Q} \neq \mathbf{Q}'$. Let $\hat{Z}$ denote the set of all $(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ satisfying the zero restrictions, hence

$$\frac{\pi(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}{\mu_Z v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))} \neq \frac{\pi(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}'))}{\mu_Z v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}'))}$$

even if $(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}) \in \hat{Z}$ and $(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}') \in \hat{Z}$, but $\mathbf{Q} \neq \mathbf{Q}'$. Therefore, $\frac{\pi(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}{\mu_Z v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}$ is not conditionally agnostic. Thus, while agnostic priors survived the transformation across parameterization, conditionally agnostic priors do not survive the same transformation.

If $\hat{\pi}$ is a prior over the orthogonal reduced-form parametrization, then the prior conditional on $(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$ satisfying the zero restrictions is

$$\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}|\hat{Z}) = \frac{\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})}{\mu_{\hat{Z}}}$$

where $\mu_{\hat{Z}} = \int_{\hat{Z}} \hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})d(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})$. Now, if $\hat{\pi}$ is the equivalent prior to π over the orthogonal reduced-form parameterization, then Theorem 2 implies that

$$\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}|\hat{Z}) = \frac{\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})}{\mu_{\hat{Z}}} = \frac{\pi(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}{\mu_{\hat{Z}} v_{f_h}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))}. \tag{13}$$

Since we have that $v_{f_h}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})) = v_{f_h}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}'))$ for every $\mathbf{Q},\mathbf{Q}' \in O(n)$, observationally equivalent parameter values that satisfy the zero restrictions have the same prior density and $\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}|\hat{Z})$ is conditionally agnostic. Notice that in this case we have transformed the prior π the orthogonal reduced-form parameterization before conditioning on the zero restrictions.

In general, $v_{f_h|Z}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q})) \neq v_{f_h}(f_h^{-1}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}))$, so the priors defined by Equations (12) and (13) are not the equal. Hence, priors that were equivalent before conditioning on the zero restrictions, are not equal equivalent after conditioning. To emphasize on this notice that using Theorem 5, the equivalent prior to $\hat{\pi}(\mathbf{B},\mathbf{\Sigma},\mathbf{Q}|\hat{Z})$ over the structural parameterization conditional on $(\mathbf{A}_0,\mathbf{A}_+)$ satisfying the zero restrictions is

$$\frac{\pi(\mathbf{A}_0,\mathbf{A}_+)}{\mu_{\hat{Z}}} \frac{v_{f_h|Z}(\mathbf{A}_0,\mathbf{A}_+)}{v_{f_h}(\mathbf{A}_0,\mathbf{A}_+)}.$$

In general, $v_{f_h|Z}(\mathbf{A}_0, \mathbf{A}_+) \neq v_{f_h|Z}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})$ for every $\mathbf{Q} \in O(n)$, hence it is now the case that

$$\frac{\pi(\mathbf{A}_0, \mathbf{A}_+)}{\mu_{\hat{Z}}} \frac{v_{f_h|Z}(\mathbf{A}_0, \mathbf{A}_+)}{v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)} \neq \frac{\pi(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})}{\mu_{\hat{Z}}} \frac{v_{f_h|Z}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})}{v_{f_h}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q})}$$

for every $\mathbf{Q} \in O(n)$, $(\mathbf{A}_0, \mathbf{A}_+) \in Z$, and $(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) \in Z$, even if $\pi(\mathbf{A}_0, \mathbf{A}_+|Z) = \pi(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}|Z)$ for every $\mathbf{Q} \in O(n)$, $(\mathbf{A}_0, \mathbf{A}_+) \in Z$, and $(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) \in Z$. Thus, $\frac{\pi(\mathbf{A}_0, \mathbf{A}_+)}{\mu_{\hat{Z}}} \frac{v_{f_h|Z}(\mathbf{A}_0, \mathbf{A}_+)}{v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)}$ is different from $\pi(\mathbf{A}_0, \mathbf{A}_+|Z)$ and it is not conditionally agnostic.

Since, in general, it is also the case that $v_{g|Z}(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)) \neq v_{g|Z}(g^{-1}(\mathbf{L}_0\mathbf{Q}, \mathbf{L}_+\mathbf{Q}))$ for every $\mathbf{Q} \in O(n)$, the same issues appears if we transform $\pi(\mathbf{A}_0, \mathbf{A}_+|Z)$ over the IRF parameterization. Thus, the main implication of these results is that if the researcher wants to use a conditionally agnostic prior, she must choose the parameterization before conditioning on zero restrictions, unlike the case with only sign restrictions. If one wants to be conditionally agnostic over a particular parameterization it is necessary to obtain the agnostic prior over such parameterization before conditioning on zero restrictions. This can be done by either detailing the agnostic prior over the parametrization of interest or detailing the agnostic prior over a more convenient parametrization and obtaining the equivalent, and also agnostic, prior over the parametrization of interest.

Let us assume that we want to be conditionally agnostic over the structural parameterization. As in the case of sign restrictions, it is going to convenient to work on the orthogonal reduced-form parametrization since drawing in that parametrization is very simple. Thus, let us detail $\hat{\pi}$ to be an agnostic prior over the orthogonal reduced-form parametrization. We need use Theorem 2 to find the equivalent agnostic prior over the structural parametrization

$$\pi(\mathbf{A}_0, \mathbf{A}_+) = \hat{\pi}(f_h(\mathbf{A}_0, \mathbf{A}_+))v_{f_h}(\mathbf{A}_0, \mathbf{A}_+).$$

Then the prior over the structural parametrization conditional on $(\mathbf{A}_0, \mathbf{A}_+)$ satisfying the zero restrictions

$$\frac{\hat{\pi}(f_h(\mathbf{A}_0, \mathbf{A}_+))v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)}{\mu_Z}$$

is conditionally agnostic and the implied conditionally agnostic posterior over the structural parameterization implied is

$$\frac{p(Y_T|\mathbf{A}_0, \mathbf{A}_+)\hat{\pi}(f_h(\mathbf{A}_0, \mathbf{A}_+))v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)}{\mu_Z p(Y_T|Z)}. \tag{14}$$

Theorem 5 implies that its equivalent posterior over the orthogonal reduced-form parameterization conditional on $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ satisfying the zero restrictions is

$$\frac{p(Y_T|f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}{\mu_Z p(Y_T|Z)v_{f_h|Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}. \tag{15}$$

Let us assume that we want to be conditionally agnostic over the IRF parameterization. We need use Theorem 1 to find the equivalent agnostic prior over the structural parametrization

$$\frac{\hat{\pi}(f_h(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))v_{f_h}(g^{-1}(\mathbf{L}_0, \mathbf{L}_+))}{|\det(Dg(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))|}.$$

Then the prior over the IRF parametrization conditional on $(\mathbf{L}_0, \mathbf{L}_+)$ satisfying the zero restrictions

$$\frac{\hat{\pi}(f_h(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))v_{f_h}(g^{-1}(\mathbf{L}_0, \mathbf{L}_+))}{\mu_{\tilde{Z}}|\det(Dg(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))|},$$

where $\tilde{Z}$ denote the set of all $(\mathbf{L}_0, \mathbf{L}_+)$ satisfying the zero restrictions, is conditionally agnostic and the implied conditionally agnostic posterior over the structural parameterization implied is

$$\frac{p(Y_T|g^{-1}(\mathbf{L}_0, \mathbf{L}_+))\hat{\pi}(f_h(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))v_{f_h}(g^{-1}(\mathbf{L}_0, \mathbf{L}_+))}{\mu_{\tilde{Z}}\tilde{p}(Y_T|\tilde{Z})|\det(Dg(g^{-1}(\mathbf{L}_0, \mathbf{L}_+)))|}. \tag{16}$$

Theorem 5 implies that its equivalent posterior over the orthogonal reduced-form parameterization conditional on $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ satisfying the zero restrictions is

$$\frac{p(Y_T|f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{g|Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}{\mu_{\tilde{Z}}\tilde{p}(Y_T|\tilde{Z})|\det(Dg(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})))|v_{f_h|Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))}. \tag{17}$$

5.6 Importance Sampler

If a researcher wants to be conditionally agnostic over the structural parameterization, drawing from the conditional posterior over the orthogonal reduced-form parameterization defined by Equation (15) is equivalent to drawing from the conditionally agnostic posterior over the structural parameterization defined by Equation (14). If a researcher wants to be conditionally agnostic over the IRF parameterization, the same is true for Equations (17) and (16). In what follows we show an importance sampler that draws from either the conditional posterior defined by Equation 15 or the conditional posterior defined by Equation 17.

5.6.1 The Proposal Density

As a proposal density, we need to use a density that we know how to draw from. We will construct such a density in two steps. In the first step will construct a proposal marginal density for the reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma})$ and in the second step we will define a proposal conditional density for the orthogonal matrix $\mathbf{Q}$ conditional on $(\mathbf{B}, \mathbf{\Sigma})$.

5.6.1.1 The Proposal Marginal Density

To build the proposal marginal density let us take the agnostic posterior over the orthogonal reduced-form parameterization associated with the agnostic prior $\hat{\pi}$. The implied marginal posterior distribution of the reduced-form parameters is

$$\frac{p(Y_T|f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))\hat{\pi}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})v_{O(n)}}{p(Y_T)}, \tag{18}$$

where $v_{O(n)} = \int_{O(n)} 1dQ$ is the volume of $O(n)$ and we have used the facts that π is agnostic, $p(Y_T|f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = \hat{p}(Y_T|\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, and that $p(Y_T) = \hat{p}(Y_T)$, if $\hat{\pi}$ and π are equivalent. We will use the marginal posterior defined by Equation (18) as our proposal marginal density to obtain our proposals for $(\mathbf{B}, \mathbf{\Sigma})$. This is equivalent to Step 1 of Algorithm 2.

5.6.1.2 The Proposal Conditional Density

We now construct our proposal conditional density for the orthogonal matrix $\mathbf{Q}$. Conditional on $(\mathbf{B}, \mathbf{\Sigma})$, let $\hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$ denotes the set of all $\mathbf{Q} \in O(n)$ such that $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ satisfy the zero restrictions. To draw the orthogonal matrix $\mathbf{Q}$ conditional on belonging to $\hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$, we first draw form a certain $\sum_{i=1}^{n}(n-(i+z_i))$-dimensional smooth manifold in $\mathbb{R}^{\sum_{i=1}^{n}(n-(i-1+z_i))}$ and then map the draw into $\hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$. Define the mapping $g_{\mathbf{B},\mathbf{\Sigma}} : \mathbb{R}^{\sum_{i=1}^{n}(n-(i-1+z_i))} \rightarrow \hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$ by

$$g_{\mathbf{B},\mathbf{\Sigma}}(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \begin{bmatrix} \frac{\mathbf{N}_1\mathbf{x}_1}{\|\mathbf{x}_1\|} & \cdots & \frac{\mathbf{N}_n\mathbf{x}_n}{\|\mathbf{x}_n\|} \end{bmatrix},$$

where $\mathbf{x}_i \in \mathbb{R}^{n-(i-1+z_i)}$ and the columns of the $n \times (n-(i-1+z_i))$ matrix $\mathbf{N}_i$ are orthonormal and lie in the null space of the $(i-1+z_i) \times n$ matrix

$$\mathbf{M}_i = \begin{bmatrix} \frac{\mathbf{N}_1\mathbf{x}_1}{\|\mathbf{x}_1\|} & \cdots & \frac{\mathbf{N}_{i-1}\mathbf{x}_{i-1}}{\|\mathbf{x}_{i-1}\|} & (Z_i\mathbf{F}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{I}_n))' \end{bmatrix}'$$

for $1 \leq i \leq n$. For $1 \leq i \leq n$ the null space of $\mathbf{M}_i$ is of dimension at least $n-(i-1+z_i)$, hence the matrix $\mathbf{N}_i$ will exist, but will not be unique. The mapping $g_{\mathbf{B},\mathbf{\Sigma}}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ takes such a draw and then maps it into $\hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$.

The sphere in $\mathbb{R}^{n-(i-1+z_i)}$ equals to the set of all points $\mathbf{x}_i$ such that $\| \mathbf{x}_i \| = 1$ for $1 \leq i \leq n$. If $\mathbf{x}_i$ is drawn from the standard Gaussian distribution over $\mathbb{R}^{n-(i-1+z_i)}$, then $\frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$ will be a draw from the uniform distribution of the sphere in $\mathbb{R}^{n-(i-1+z_i)}$ for $1 \leq i \leq n$. Hence, by construction, applying $g_{\mathbf{B},\mathbf{\Sigma}}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ to $\mathbf{x}_i$ for $1 \leq i \leq n$, where $\mathbf{x}_i$ for $1 \leq i \leq n$ is drawn from the standard Gaussian distribution over $\mathbb{R}^{n-(i-1+z_i)}$, is a draw of an orthogonal matrix in $\hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$.

If $\mathcal{M}$ is the set of all points $(\mathbf{x}_1, \cdots, \mathbf{x}_n) \in \mathbb{R}^{\sum_{i=1}^{n}(n-(i-1+z_i))}$ such that $\| \mathbf{x}_i \| = 1$ for $1 \leq i \leq n$, Theorem 5 tells us that the density of $\mathbf{Q} \in \hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$ is

$$\frac{1}{s v_{g_{\mathbf{B},\mathbf{\Sigma}}|_{\mathcal{M}}}(g_{\mathbf{B},\mathbf{\Sigma}}|_{\mathcal{M}}^{-1}(\mathbf{Q}))}, \tag{19}$$

where s is the product of the volume of the spheres in $\mathbb{R}^{n-(i-1+z_i)}$ for $1 \leq i \leq n$.[9] This is equivalent to Step 2 of Algorithm 2.

5.7 The Weights and the Algorithm

If a researcher wants to be conditionally agnostic over the structural parameterization, the reader has to remember that this researcher wants to draw from the conditionally agnostic posterior over the structural parameterization

$$\frac{p(Y_T | (\mathbf{A}_0, \mathbf{A}_+))\hat{\pi}(f_h(\mathbf{A}_0, \mathbf{A}_+))v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)}{\mu_Z p(Y_T|Z)}.$$

But as argued above this equivalent to draw from the posterior distribution over the orthogonal reduced-form parametrization conditional on $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ satisfying the zero restrictions defined by Equation (15). In order to do that the resulting draws $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ obtained in the two steps defined in Sections 5.6.1.1 and 5.6.1.2 have to be corrected using importance sampling weighting. Then, the unnormalized importance weight is given by Equation (15) divided by the product of Equations (18) and (19), which equals to

$$\frac{v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{g_{\mathbf{B},\mathbf{\Sigma}}|_{\mathcal{M}}}(g_{\mathbf{B},\mathbf{\Sigma}}|_{\mathcal{M}}^{-1}(\mathbf{Q}))}{v_{f_h|_Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))} \left\{ \frac{s p(Y_T)}{v_{O(n)}\mu_Z p(Y_T|Z)} \right\}. \tag{20}$$

[9] While it is not possible to choose the $\mathbf{N}_i$ in such a way that $g_{\mathbf{B},\mathbf{\Sigma}}$ will be differentiable over all of $\mathbb{R}^{\sum_{i=1}^{n}(n-(i-1+z_i))}$, it is possible to define them so that $\mathbb{R}^{\sum_{i=1}^{n}(n-(i-1+z_i))}$ is differentiable except on a closed set of measure zero.

The following algorithm uses the above results to independently draw from an posterior over the structural parameterization on $(\mathbf{A}_0, \mathbf{A}_+)$ satisfying the zero restrictions.

Algorithm 3.

1. *Draw* $(\mathbf{B}, \mathbf{\Sigma})$ *from the marginal posterior of the reduced-form parameters defined by Equation* (18).
2. *Use* $g_{\mathbf{B},\mathbf{\Sigma}}$ *to draw from* $\mathbf{Q} \in \hat{Z}_{\mathbf{B},\mathbf{\Sigma}}$.
3. *Calculate weights using Equation* (20).
4. *Set* $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.
5. *Keep the draw if the sign restrictions are satisfied.*
6. *Return to Step 1 until the required number of draws from the posterior on the structural parameters conditional on the sign restrictions has been obtained.*

Now any function of the structural parameters, included the IRFs, can be calculated using the $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ obtained in Steps 1 and 2 and the weights obtained in Step 3.

If we want to be conditionally agnostic over the IRF parameterization, we also use Algorithm 3, but the weights will calculated using the following equation instead

$$\frac{v_{f_h}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{g|Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))v_{g_{\mathbf{B},\mathbf{\Sigma}}|\mathcal{M}}(g_{\mathbf{B},\mathbf{\Sigma}}|_{\mathcal{M}}^{-1}(\mathbf{Q}))}{|\det(Dg(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})))|v_{f_h|Z}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))} \left\{ \frac{sp(Y_T)}{v_{O(n)}\mu_{\tilde{Z}}\tilde{p}(Y_T|\tilde{Z})} \right\}. \tag{21}$$

5.8 Conditionally Flat Prior

We say that a prior is conditionally flat if the prior density is equal across all parameters that satisfy the zero restrictions. Following the same arguments as in the case of sign restrictions, it easy to show that one can choose $\hat{\pi}$ so the prior is conditionally flat over either the structural or the IRF parameterizations. In particular, if one wants to be conditionally flat over the structural parameterization Corollary 5 describes the $\hat{\pi}$ to be used in conjunction with weights described by Equation (20). If one wants to be conditionally flat over the IRF parameterization Corollary 6 describes the $\hat{\pi}$ to be used in conjunction with weights described by Equation (21). Clearly, if one is conditionally flat over the structural parameterization, then it is conditionally agnostic over the structural parameterization. The same is true for the IRF parameterization.

6 Mountford and Uhlig's (2009) PFA

In this section, we discuss the PFA developed by Mountford and Uhlig (2009). First, we describe the Algorithm and highlight how it selects a particular orthogonal matrix $\mathbf{Q}$. Third, we mention some of the drawbacks of the PFA as reported in the literature. Finally, we argue that all these drawbacks can be summarized as the PFA approach using a prior on the orthogonal matrices that it is not conditionally agnostic.

6.1 PFA Algorithm

Let us assume that we want to use the PFA to identify a single structural shock: the j-th structural shock.[10] Let $(\mathbf{B}, \boldsymbol{\Sigma})$ be any draw of the reduced-form parameters. Consider a case where the identification of the j-th structural shock restricts the IRF of a set of variables indexed by $I_{j,+}$ to be positive and the IRF of a set of variables indexed by $I_{j,-}$ to be negative, where $I_{j,+}$ and $I_{j,-} \subset \{0, 1, \ldots, n\}$. Furthermore, assume that the restrictions on variable $i \in I_{j,+}$ are enforced during $H_{i,j,+}$ periods and the restrictions on variable $i \in I_{j,-}$ are enforced during $H_{i,j,-}$ periods. In addition to the sign restrictions, assume that the researcher imposes zero restrictions on the IRFs to identify the j-th structural shock. The PFA finds a vector, $\bar{\mathbf{q}}_j^*$, in the sphere in $\mathbb{R}^n$ such that the IRFs come close to satisfying the sign restrictions, conditional on the zero restrictions being satisfied, according to a loss function.[11] In particular, it solves the following optimization problem

$$\bar{\mathbf{q}}_j^* = \operatorname{argmin}_{\bar{\mathbf{q}}_j \in \mathcal{S}^0} \quad \Psi\left(\bar{\mathbf{q}}_j\right)$$

subject to

$$\mathbf{Z}_j \mathbf{F}\left(h(\boldsymbol{\Sigma})^{-1}, \mathbf{B}h(\boldsymbol{\Sigma})^{-1}\right) \bar{\mathbf{q}}_j = \mathbf{0}$$

where

$$\Psi\left(\bar{\mathbf{q}}_j\right) = \sum_{i \in I_+} \sum_{h=0}^{H_{i,+}} g\left(-\frac{\mathbf{e}_i' \mathbf{L}_h\left(h(\boldsymbol{\Sigma})^{-1}, \mathbf{B}h(\boldsymbol{\Sigma})^{-1}\right) \bar{\mathbf{q}}_j}{\sigma_i}\right) + \sum_{i \in I_-} \sum_{h=0}^{H_{i,-}} g\left(\frac{\mathbf{e}_i' \mathbf{L}_h\left(h(\boldsymbol{\Sigma})^{-1}, \mathbf{B}h(\boldsymbol{\Sigma})^{-1}\right) \bar{\mathbf{q}}_j}{\sigma_i}\right),$$

[10] The PFA can be used to identify as many shocks as variables there are in the system, but assuming that we only identify the j-th structural shock makes the exposition easier.

[11] See Mountford and Uhlig (2009) for details.

$g(\omega) = 100\omega$ if $\omega \geq 0$ and $g(\omega) = \omega$ if $\omega \leq 0$, σ_i is the standard error of variable i, and $\bar{\mathbf{Q}}^*_{j-1} = \begin{bmatrix} \bar{\mathbf{q}}^*_1 & \cdots & \bar{\mathbf{q}}^*_{j-1} \end{bmatrix}$ for $1 \leq j \leq n$.[12] Notice that $\mathbf{L}_h\left(h(\mathbf{\Sigma})^{-1}, \mathbf{B}h(\mathbf{\Sigma})^{-1}\right)\bar{\mathbf{q}}_j$ is the j-th column of the IRFs at horizon h implied by $\left(\mathbf{B}, \mathbf{\Sigma}, \bar{\mathbf{Q}}\right)$.

6.2 Drawbacks of the PFA

The literature has mentioned several possible issues with the PFA approach. First, by choosing a single orthogonal matrix the PFA does not consider the structural parameters robustness attached to sign restrictions. This drawback is important because robustness is one of the most appealing motivations for using sign restrictions. Second, the optimal orthogonal matrix that solves the system of equations may be such that the sign restrictions do not hold.[13] Third, as acknowledged by Uhlig (2005), the PFA rewards orthogonal matrices that imply responses strongly satisfying the sign restrictions. This rewarding scheme may result on imposing additional sign restrictions on seemingly unrestricted variables. For example, Caldara and Kamps (2012) use a bivariate SVAR to show how the penalty function restricts the output response to a tax increase to be negative. Finally, Binning (2013) points out that in those cases in which several shocks are identified recursively, the ordering of the shocks determines their importance, and that the penalty function relies on numerical optimization methods that are slow to implement. For the reasons mentioned above, Uhlig (2005) and Caldara and Kamps (2012) conclude that the penalty function approach should be interpreted as an additional identifying assumption. In the next section we formalize this argument as the PFA using a prior that it is not conditionally agnostic over any parameterization and, thus, on the orthogonal matrices that imposes additional restrictions to the ones embedded in the identification scheme.

6.3 The PFA Prior

This paper highlights that all these issues can be summarized as the PFA using a prior that it is not conditionally agnostic over any parameterization. Hence, the PFA imposes additional restrictions to the ones embedded in the identification scheme. In fact, the PFA prior is such that, for every value of the reduced-form parameters, a single value of the structural parameters has positive prior probability. This implies that many structural parameters that satisfy the restrictions are discarded by the PFA prior. In the next section, we will use Beaudry, Nam and Wang (2011) empirical application to illustrate

[12]To obtain σ_i, we compute the standard deviation of the OLS residuals associated with the i-th variable.
[13]This is true even in the extreme case in which no orthogonal matrix satisfaces the sign restrictions.

the implications of the PFA prior relative to an agnostic prior over the structural parameterization.

7 Application to Optimism Shocks

In this section, we illustrate our methodology revisiting one application related to optimism shocks previously analyzed in the literature by Beaudry, Nam and Wang (2011) using the PFA prior. The aim of Beaudry, Nam and Wang (2011) is to contribute to the debate regarding the source and nature of business cycles. The authors claim to provide new evidence on the relevance of optimism shocks as the main driver of macroeconomic fluctuations using a SVAR identified with sign and zero restrictions.

For illustrative purposes it suffices to focus on Beaudry, Nam and Wang's (2011) less restrictive identification scheme, Identification 1. While using the PFA prior once could conclude that optimism shocks are associated with standard business cycle type phenomena because they generate a simultaneous boom in consumption, and hours worked, we show that this conclusion relays on additional restrictions implied by the PFA prior and not in the identification restrictions explicitly stated by the authors.

7.1 Data and Identification Strategy

Beaudry, Nam and Wang (2011) consider a SVAR with five variables: TFP, stock price, consumption, the real federal funds rate, and hours worked.[14] In their less restrictive identification strategy, shown in Table 1, optimism shocks are identified as positively affecting stock prices and as being orthogonal to TFP at horizon zero. The remaining variables are unrestricted. Appendix A gives details on the priors and the data.

Adjusted TFP	Stock Price	Consumption	Real Interest Rate	Hours Worked
0	Positive	Unrestricted	Unrestricted	Unrestricted

Table 1: Restrictions on Impulse Response Functions at Horizon 0.

It is straightforward to map this identification scheme to the function $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ and the matrices $\mathbf{S}$s and $\mathbf{Z}$s necessary to implement our methodology. We choose to call the first structural shock

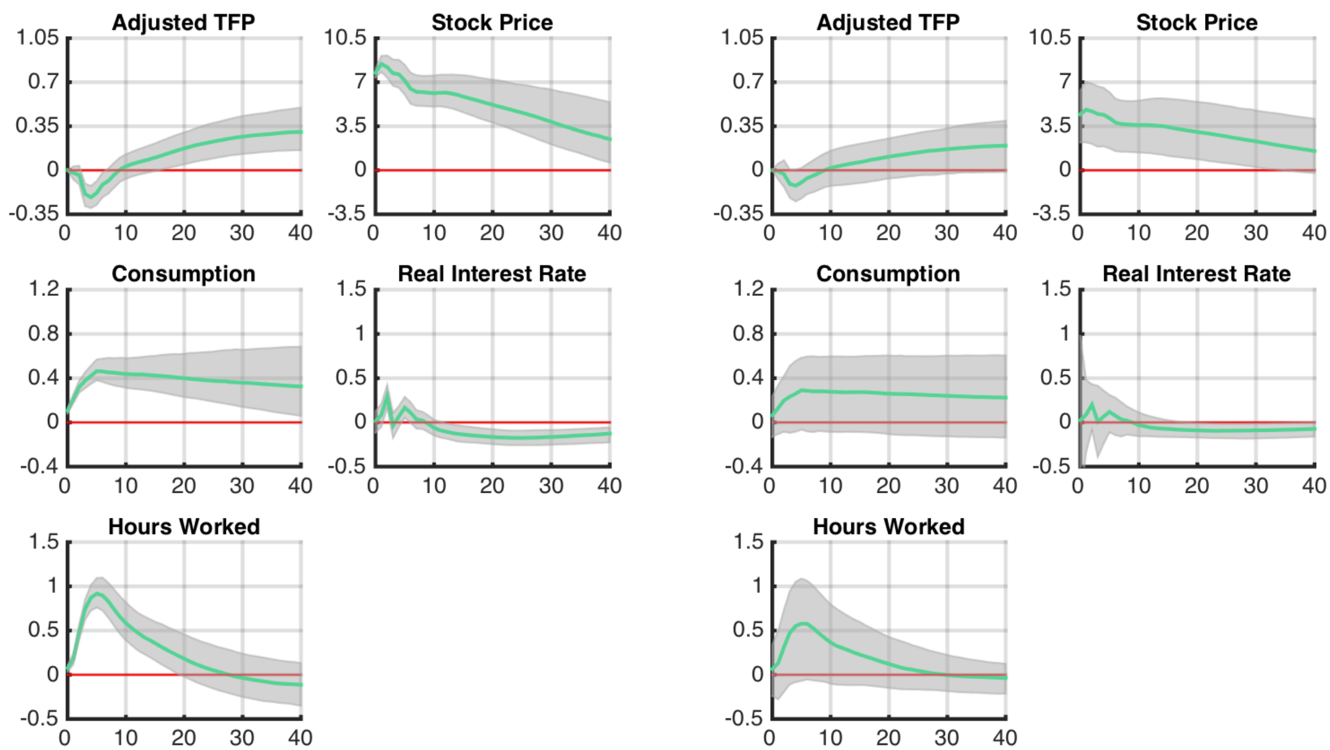
[14] Beaudry, Nam and Wang (2011) also use an extended version of their model including investment and output as additional variables. The issues illustrated here are also present when using the 7 variables model.

the optimism shock. Since the sign and zero restrictions are imposed at horizon zero, we have that $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{L}_0(\mathbf{A}_0, \mathbf{A}_+)$. The matrices **S**s and **Z**s associated with the optimism shocks are

$$\mathbf{S}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{Z}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

7.2 IRFs

Let's begin by comparing the IRFs obtained using the PFA prior with the IRFs obtained using the agnostic prior. This comparison is helpful to illustrate how Beaudry, Nam and Wang's (2011) results relay on the additional restrictions introduced by the PFA prior.



(a) PFA prior

(b) Agnostic Prior over $(\mathbf{A}_0, \mathbf{A}_+)$

Figure 2: IRFs to an Optimism Shock

Panel (a) in Figure 2 shows the pointwise median as well as the 68 percent pointwise credible sets for the IRFs of TFP, stock price, consumption, the federal funds rate, and hours worked using the PFA prior.[15] The key message from this panel is that optimism shocks generate a boom in consumption

[15]There are alternative methods of summarizing the outcome of set identified SVARs models, see Inoue and Kilian

and hours worked: The pointwise credible sets associated with the IRFs do not contain zero for, at least, 20 quarters. Thus, a researcher looking at these results would conclude that optimism shocks generate standard business cycle type phenomena.

Indeed, these IRFs are highlighted by Beaudry, Nam and Wang (2011). If these IRFs were the result of only imposing the two restrictions described in Table 1, the findings reported in Panel (a) of Figure 2 would strongly support the view that optimism shocks are relevant for business cycle fluctuations. The question is how much of the results reported in Panel (a) of Figure 2 are due to the restrictions described in Table 1 and how much are due to the additional restrictions imposed by PFA prior.

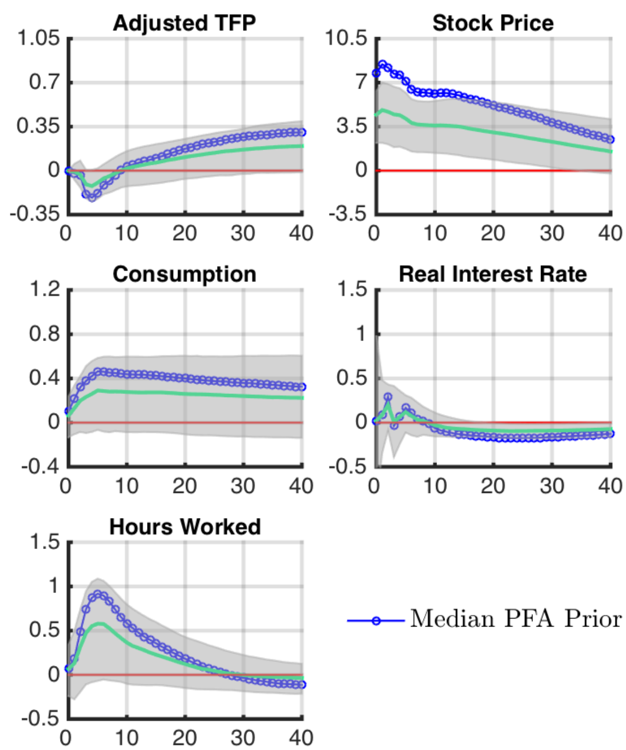


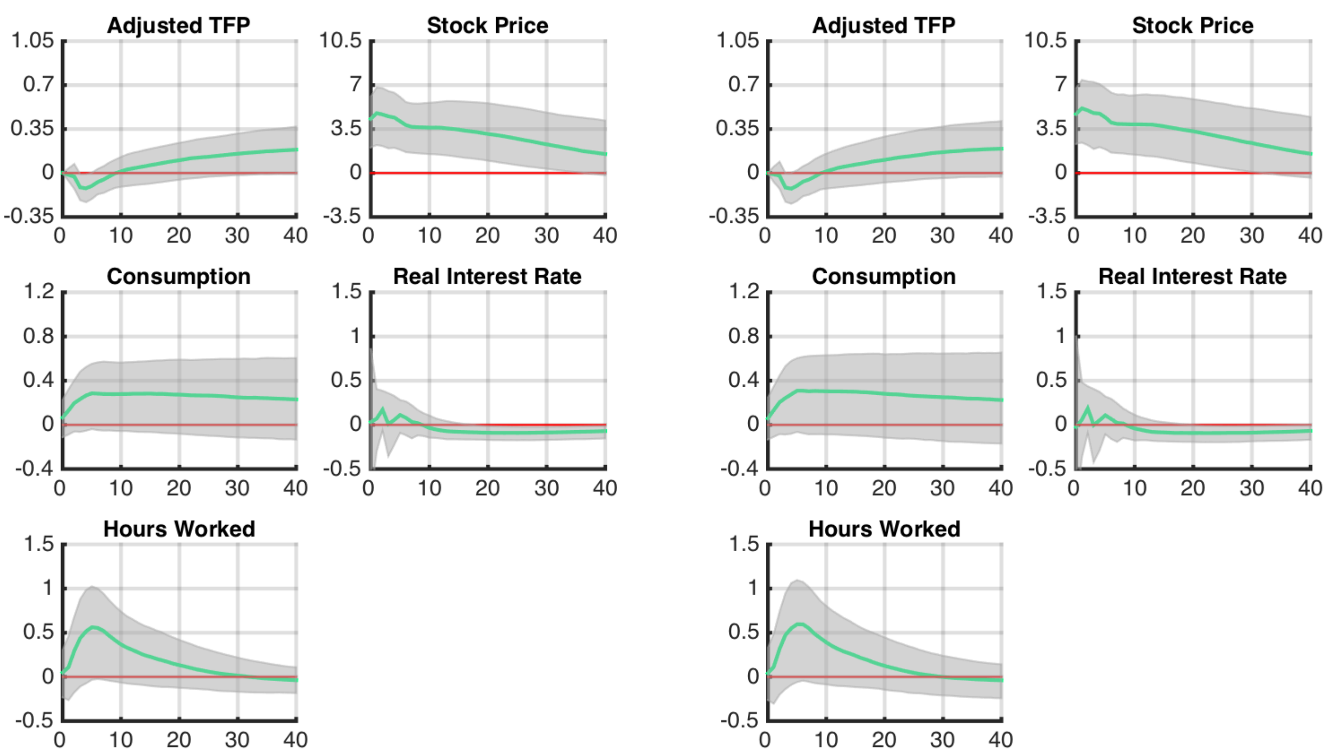
Figure 3: Agnostic Prior over Structural Parameterization Relative to PFA Prior

Panel (b) in Figure 2 shows that once we use a conditionally agnostic prior over the structural parameterization to compute the implied posterior IRFs, the results highlighted by Beaudry, Nam and Wang (2011) disappear. There are three important differences with the results reported when working with the PFA prior. First, the PFA prior chooses a very large median response of stock prices in order to

(2013). We report 68 percent pointwise credible sets in order to facilitate the comparison with the results reported by Beaudry, Nam and Wang (2011).

minimize the loss function used to impose the sign restriction on stock prices. Second, the pointwise median IRFs for consumption and hours worked are closer to zero when we use the conditionally agnostic prior over the structural parameterization. Third, the pointwise credible sets associated with the conditionally agnostic prior over the structural parameterization are much larger than the ones obtained with the PFA prior. Figure 3 compares the IRFs obtained with the conditionally agnostic prior over the structural parameterization with the median IRFs computed with the PFA prior. As the reader can see the additional restrictions implied by the PFA prior boosts the effects of optimism shocks on stock prices, consumption, and hours.

We conclude this section by discussing the IRFs obtained using a conditionally flat prior over the structural parameterization, shown Panel (a) of Figure 4, and IRFs obtained using a conditionally flat prior over the IRF parameterization, shown in Panel (b) of Figure 4. Both the conditionally flat priors imply similar IRF to those obtained using the conditionally agnostic prior over the structural parameterization.



(a) Flat Prior over $(\mathbf{A}_0, \mathbf{A}_+)$

(b) Flat Prior over $(\mathbf{L}_0, \mathbf{L}_+)$

Figure 4: IRFs to an Optimism Shock

7.3 Forecast Error Variance Decomposition (FEVD)

Next, we examine the contribution of optimism shocks to the FEVD obtained using the conditionally agnostic prior over structural parameterization and the PFA prior. For ease of exposition, in Table 2 we only focus on the contributions to the FEV at horizon 40.

	PFA Prior	Agnostic Prior on $(\mathbf{A}_0, \mathbf{A}_+)$
Adjusted TFP	0.18 [0.08 , 0.31]	0.10 [0.03 , 0.25]
Stock Price	0.73 [0.55 , 0.85]	0.26 [0.07 , 0.56]
Consumption	0.26 [0.13 , 0.42]	0.16 [0.02 , 0.50]
Real Interest Rate	0.14 [0.07 , 0.22]	0.19 [0.09 , 0.38]
Hours Worked	0.31 [0.21 , 0.44]	0.18 [0.05 , 0.51]

Table 2: Share of FEV Attributable to Optimism Shocks at Horizon 40.

	Flat Prior on $(\mathbf{A}_0, \mathbf{A}_+)$	Flat Prior on $(\mathbf{L}_0, \mathbf{L}_+)$
Adjusted TFP	0.10 [0.03 , 0.26]	0.10 [0.03 , 0.27]
Stock Price	0.29 [0.07 , 0.61]	0.27 [0.06 , 0.58]
Consumption	0.18 [0.03 , 0.54]	0.16 [0.03 , 0.51]
Real Interest Rate	0.19 [0.09 , 0.36]	0.19 [0.08 , 0.39]
Hours Worked	0.19 [0.05 , 0.53]	0.18 [0.05 , 0.51]

Table 3: Share of FEV Attributable to Optimism Shocks at Horizon 40.

Using the PFA prior, the median contribution of optimism shocks to the FEV of consumption and hours worked is 26 and 31 percent, respectively. In contrast, using the unrestricted prior the median contributions are 16 and 18 percent, respectively. Table 2 also reports the 68 percent pointwise credible sets. As was the case with IRFs, the confidence intervals are much wider under the conditionally agnostic prior over structural parameterization. Table 3 replicates Table 2 using conditionally flat priors on the structural parameterization and on the IRF parameterization respectively.

7.4 Understanding the Additional Restrictions of the PFA Prior

We conclude the empirical application by characterizing analytically the PFA prior, and by analyzing its implications for the application studied here. Let $(\mathbf{B}, \mathbf{\Sigma})$ be any value of the reduced-form parameters. If we want to identify the first shock, the optimal $\bar{\mathbf{q}}_1^*$ has to solve

$$\mathbf{e}_1' \mathbf{L}_0 \left(\mathbf{T}^{-1}, \mathbf{B}\mathbf{T}^{-1} \right) \bar{\mathbf{q}}_1^* = 0. \tag{22}$$

where we have assumed the notational convention that $\mathbf{T} = h(\mathbf{\Sigma})$. Equation (22) implies that the optimal $\bar{\mathbf{q}}_1^*$ has to be such that $\mathbf{e}_1' \mathbf{L}_0 \left(\mathbf{T}^{-1}, \mathbf{B}\mathbf{T}^{-1} \right) \bar{\mathbf{q}}_1^* = \mathbf{e}_1' \mathbf{T}' \bar{\mathbf{q}}_1^* = \mathbf{t}_{1,1} \bar{\mathbf{q}}_{1,1}^* = 0$, where the next to last equality follows because $\mathbf{T}'$ is lower triangular. Thus, $\bar{\mathbf{q}}_{1,1}^* = 0$. To find the remaining entries of $\bar{\mathbf{q}}_1^*$, it is convenient to write $\mathbf{e}_2' \mathbf{L}_0 \left(\mathbf{T}^{-1}, \mathbf{B}\mathbf{T}^{-1} \right) \bar{\mathbf{q}}_1^* = \mathbf{e}_2' \mathbf{T}' \bar{\mathbf{q}}_1^* = \sum_{s=1}^{2} \mathbf{t}_{s,2} \bar{\mathbf{q}}_{s,1}^*$, where the last equality follows because $\mathbf{T}'$ is lower triangular. Substituting $\bar{\mathbf{q}}_{1,1}^* = 0$ into $\mathbf{e}_2' \mathbf{L}_0 \left(\mathbf{T}^{-1}, \mathbf{B}\mathbf{T}^{-1} \right) \bar{\mathbf{q}}_1^*$ yields $\mathbf{t}_{2,2} \bar{\mathbf{q}}_{2,1}^*$. Since $\bar{\mathbf{q}}_{1,1}^* = 0$, it is straightforward to verify that the criterion function is minimized at $\bar{\mathbf{q}}_1^* = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}'$. Since no other structural shocks are identified, one may think that the PFA does not impose any constraint on the rest of the columns of the orthogonal matrix $\bar{\mathbf{Q}}^*$. In fact this is not the case. Since the rest of the columns have to be orthogonal to $\bar{\mathbf{q}}_1^*$, it is the case that the second entry of every other column of $\bar{\mathbf{Q}}^*$ is restricted to be zero. This has implications on the rest of the structural shocks in the model.

Let us now analyze the implications of the above results to the prior implied by PFA in the structural parameters space. Because $\bar{\mathbf{q}}_1^* = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}'$, the PFA imposes restrictions on the first column of the structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$. To see this notice that the PFA approach only gives positive prior probability to structural matrices $(\mathbf{A}_0, \mathbf{A}_+)$ such that the first column of $\mathbf{A}_0$ equals $\mathbf{T}^{-1} \bar{\mathbf{q}}_1^*$, and the first column of $\mathbf{A}_+$ equals $\mathbf{B}\mathbf{T}^{-1} \bar{\mathbf{q}}_1^*$. Hence, the PFA prior in the structural parameters space only gives positive probability to structural matrices $(\mathbf{A}_0, \mathbf{A}_+)$ such that their first columns equal

$$\begin{bmatrix} -\mathbf{t}_{1,2} \left(\mathbf{t}_{1,1} \mathbf{t}_{2,2} \right)^{-1} & \mathbf{t}_{2,2}^{-1} & 0 & \cdots & 0 \end{bmatrix}', \text{ and } \left(\mathbf{t}_{1,1} \mathbf{t}_{2,2} \right)^{-1} \begin{bmatrix} (\mathbf{b}_{1,2} \mathbf{t}_{1,1} - \mathbf{b}_{1,1} \mathbf{t}_{1,2}) & (\mathbf{b}_{2,2} \mathbf{t}_{1,1} - \mathbf{b}_{2,1} \mathbf{t}_{1,2}) & \dots \end{bmatrix}'$$

respectively. As one can see, some of these restrictions are zero restrictions on $\mathbf{A}_0$. The results presented in this Section can be generalized because by choosing an optimal orthogonal matrix $\mathbf{Q}$ the PFA approach is always going to impose a Dirac prior on some (or all) of the columns of matrix $\mathbf{Q}$ and thus impose restrictions on some (or all) of the columns of matrix $(\mathbf{A}_0, \mathbf{A}_+)$, although these may be on the affine type.

8 Conclusion

We developed an efficient algorithm for inference based on SVARs identified with sign and zero restrictions that properly draws from the agnostic posterior distribution on the structural parameters conditional on the sign and zero restrictions. Critically, our theoretical grounded algorithms guarantee that identification is coming only from the sign and zero restrictions proclaimed by the researcher. We extend the sign restrictions methodology developed by Rubio-Ramírez, Waggoner and Zha (2010) to allow for zero restrictions. As was the case in Rubio-Ramírez, Waggoner and Zha (2010), we obtain most of our results by imposing sign and zero restrictions on the impulse response functions, but our algorithm allows for a larger class of restrictions.

Appendices

A Empirical Application: Estimation and Inference

Following Beaudry, Nam and Wang (2011) we estimate equation (3) with four lags using Bayesian methods with a Normal-Wishart prior as in Uhlig (2005). We use the data set created by Beaudry, Nam and Wang (2011). This data set contains quarterly U.S. data for the sample period 1955Q1-2010Q4 and includes the following variables: TFP, stock price, consumption, real federal funds rate, hours worked, investment, and output. TFP is the factor-utilization-adjusted TFP series from John Fernald's website. Stock price is the Standard and Poor's 500 composite index divided by the CPI of all items from the Bureau of Labor Statistics (BLS). Consumption is real consumption spending on non-durable goods and services from the Bureau of Economic Analysis (BEA). The real federal funds rate corresponds to the effective federal funds rate minus the inflation rate as measured by the growth rate of the CPI all items from the BLS. Hours worked is the hours of all persons in the non-farm business sector from the BLS. The series corresponding to stock price, consumption, and hours worked are normalized by the civilian non-institutional population of 16 years and over from the BLS. All variables are logarithmic levels except for the real interest rate, which is in levels but not logged.

References

Baumeister, C. and L. Benati (2013). Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound. *International Journal of Central Banking 9*(2), 165–212.

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

Baumeister, C. and J. D. Hamilton (2015b). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *Working Paper*.

Beaudry, P., D. Nam, and J. Wang (2011). Do Mood Swings Drive Business Cycles and is it Rational? *NBER Working Papers*.

Binning, A. (2013). Underidentified SVAR Models: A Framework for Combining Short and Long-run Restrictions with Sign-restrictions. *Norges Bank Working Papers*.

Caldara, D. and C. Kamps (2012). The Analytics of SVAR: A Unified Framework to Measure Fiscal Multipliers. *Finance and Economics Discussion Series 2012-20*.

Canova, F. and G. D. Nicoló (2002). Monetary Disturbances Matter for Business Fluctuations in the G-7. *Journal of Monetary Economics 49*(6), 1131–1159.

DeJong, D. N. (1992). Co-integration and Trend-stationarity in Macroeconomic Time Series: Evidence from the Likelihood Function. *Journal of Econometrics 52*(3), 347–370.

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

Gafarov, B. and J. L. Montiel Olea (2015). On the Maximum and Minimum Response to an Impulse in SVARs. *Working Paper*.

Gambacorta, L., B. Hofmann, and G. Peersman (2014). The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis. *Journal of Money, Credit and Banking 46*(4), 615–642.

Giacomini, R. and T. Kitawaga (2015). Robust Inference about Partially Identified SVARs.

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

Moon, H. R. and F. Schorfheide (2012). Bayesian and frequentist inference in partially identified models. *Econometrica 80*(2), 755–782. ¡p¿The previous version is available as NBER Working Paper 14882. The first (incomplete) version of the paper circulated under the title "A Bayesian Look at Partially-Identified Models."¡/p¿.

Moon, H. R., F. Schorfheide, and E. Granziera (2013). Inference for Vars Identified with Sign Restrictions.

Mountford, A. and H. Uhlig (2009). What are the Effects of Fiscal Policy Shocks? *Journal of Applied Econometrics 24*(6), 960–992.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica 39*, 577–591.

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

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

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

Uhlig, H. (2005). What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics 52*(2), 381–419.