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

Jonas E. Arias, Juan F. Rubio-Ramírez, and Daniel F. Waggoner

Abstract: In this paper, we develop algorithms to independently draw from a family of conjugate posterior distributions over the structural parameterization when sign and zero restrictions are used to identify SVARs. We call this family of conjugate posterior distributions normal-generalized-normal. Our algorithms draw from a conjugate uniform-normal-inverse-Wishart posterior over the orthogonal reduced-form parameterization and transform the draws into the structural parameterization; this transformation induces a normal-generalized-normal posterior distribution over the structural parameterization. The uniform-normal-inverse-Wishart posterior over the orthogonal reduced-form parameterization has been prominent after the work of Uhlig (2005). We use Beaudry, Nam, and Wang's (2011) work on the relevance of optimism shocks to show the dangers of using alternative approaches to implement sign and zero restrictions to identify SVARs like the penalty function approach. In particular, we analytically show that the penalty function approach adds restrictions to the ones described in the identification scheme.

JEL classification: C11, C32, E50

Key words: identification, sign restrictions, simulation

---

1   Introduction

Structural vector autoregressions (SVARs) identified with sign and zero restrictions have become prominent. The fact that identification generally comes from fewer restrictions than in traditional identification schemes and that any conclusions are robust across the set of SVARs consistent with the restrictions has made the approach attractive to researchers. Most papers using this approach work in the Bayesian paradigm.[1] In this paper we develop algorithms to independently draw from a family of conjugate posterior distributions over the structural parameterization conditional on the sign and zero restrictions. We call this family of conjugate posterior distributions normal-generalized-normal and we show that it is commonly used in the literature, mainly after the work of Sims and Zha (1998).

We focus on two different parameterizations of SVARs. In addition to the typical structural parameterization, SVARs can also be written as the product of the reduced-form parameters and the set of orthogonal matrices, which we call the orthogonal reduced-form parameterization. Our algorithms will draw from a conjugate posterior distribution over the orthogonal reduced-form parameterization and then transform the draws into the structural parameterization. We follow the literature in our choice of the family of conjugate posterior distributions over the reduced parameters and use the normal-inverse-Wishart density.[2] This choice is common because it is a conjugate family and it is extremely easy to independently draw from it. Our choice of conjugate posterior over the set of orthogonal matrices conditional on the reduced-form parameters is uniform. This uniform-normal-inverse-Wishart density over the orthogonal reduced-form parameterization is a recurrent choice after the work of Uhlig (2005). We then develop a change of variable theory that allows us to characterize the induced family of posterior densities over the structural parameterization. This theory shows that a uniform-normal-inverse-Wishart posterior density over the orthogonal reduced-form parameterization induces a normal-generalized-normal distributions posterior distribution over the structural parameterization. The family of normal-generalized-normal densities over the structural parameterization is also conjugate and it is often used in the literature. In any case, our algorithms can be easily modified to consider a more general family of posterior distributions, both conjugate and non-conjugate.

Using our change of variable theory we first show that current algorithms for SVARs identified only

---

[1]Exceptions are Moon and Schorfheide (2012), Moon, Schorfheide and Granziera (2013), and Gafarov, Meier and Montiel Olea (2016a,b). Moon and Schorfheide (2012) analyze the differences between Bayesian probability bands and frequentist confidence sets in partially identified models. Moon, Schorfheide and Granziera (2013) and Gafarov, Meier and Montiel Olea (2016a,b) develop methods of constructing error bands for impulse response functions of sign-restricted SVARs that are valid from a frequentist perspective.

[2]Alternatively, one could use a normal-Wishart density.

by sign restrictions, as described by Rubio-Ramírez, Waggoner and Zha (2010), are in fact making independent draws from the normal-generalized-normal distribution over the structural parameterization conditional on the sign restrictions. These algorithms independently draw from the uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization, only accepting draws such that the sign restrictions hold and then transforming the accepted draws into the structural parameterization. Next, we adapt these algorithms to consider zero restrictions. While the set of all structural parameters satisfying the sign restrictions will be open in the set of all structural parameters, the set of all structural parameters satisfying the signs and zero restrictions is of measure zero in the set of all structural parameters. This invalidates the direct use of current algorithms when zero restrictions are considered. But the set of all structural parameters satisfying both the sign and zero restrictions is of positive measure in the set of all structural parameters satisfying the zero restrictions. Hence, we describe an algorithm that makes independent draws from the set of all structural parameters satisfying the zero restrictions. The key to this algorithm is that the class of zero restrictions on the structural parameters maps to linear restrictions on the orthogonal matrices, conditional on the reduced-form parameters. This algorithm independently draws from normal-inverse-Wishart over the reduced-form parameters and from the set of orthogonal matrices such that the zero restrictions hold. Because the zero restrictions define a lower dimensional smooth manifold in the set of all structural parameters, our change of variable theory allows us to do two things. First, we show that this algorithm does not induce a posterior distribution over the structural parameterization from the family of normal-generalized-normal distributions conditional on the sign and zero restrictions. Second, we calculate the induced density and write an importance sampler that independently draws from normal-generalized-normal distributions over the structural parameterization conditional on the sign and zero restrictions.

When using sign and zero restrictions, a commonly used algorithm is Mountford and Uhlig's (2009) penalty function approach − PFA henceforth. We show that the PFA adds restrictions; hence, identification does not solely come from the sign and zero restrictions considered in the identification scheme. We show the consequences of using the PFA by first replicating the results in Beaudry, Nam and Wang (2011), and by comparing them with the results that a researcher would obtain if our importance sampler were to be used instead. 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 an SVAR identified by imposing a sign restriction on the impact response of stock prices to optimism shocks and a zero restriction on the contemporaneous response of TFP to these shocks. Based on

the results obtained with the PFA, 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. Once our importance sampler is used, 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. The results reported in Beaudry, Nam and Wang (2011) are significantly affected by the additional restrictions imposed by the PFA.

We are not the first ones to criticize the PFA. There is an existing literature that already does exactly that. For example, Caldara and Kamps (2012) and Binning (2013) share some of our concerns, while adding others, about the PFA. In related and very original work, Giacomini and Kitawaga (2015) are also concerned with the choice of the priors densities 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 densities. We see our paper as sympathetic to their concern about the choice of priors densities.

Finally, we have to highlight Baumeister and Hamilton (2015). This paper directly draws in 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 parameterization has clear advantages, mainly being able to define priors densities directly on economically interpretable structural parameters, this approach uses a Metropolis-Hastings algorithm to make the draws. Hence this approach is inefficient compared with ours and harder to implement in larger models.

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).

2   Pitfalls of the Penalty Function Approach

Beaudry, Nam and Wang (2011) analyze the relevance of optimism shocks as a driver of macroeconomic fluctuations using SVARs identified with sign and zero restrictions. More details about their work will be given in Section 8. At this point it suffices to say that in their most basic SVAR, Beaudry, Nam and Wang (2011) use data on total factor productivity (TFP), stock prices, consumption, the real federal funds rate, and hours worked. Their identification scheme defines optimism shocks as

positively affecting stock prices but not affecting TFP contemporaneously and they use the PFA to implement it. Beaudry, Nam and Wang (2011) also claim that identification solely comes from these two restrictions.

Figure 1: IRFs to a one standard deviation optimism shock. The solid curves represent the point-wise posterior medians and the shaded areas represent the 68 percent point-wise probability bands. The figure is based on 10,000 independent draws obtained using the PFA.

Figure 1 replicates the main result in Beaudry, Nam and Wang (2011). As shown by the narrow 68 percent point-wise probability bands of their IRFs, Beaudry, Nam and Wang (2011) obtain the result that consumption and hours worked respond positively and strongly to optimism shocks. If the IRFs shown in Figure 1 were the IRFs 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. When the PFA is used, identification does not solely come from the identifying restrictions; as we show below, the PFA introduces restrictions in addition to the ones specified in the identification scheme. In Section 8 we will show that if our importance sampler is used instead, the results do not back as strongly the view that optimism shocks are relevant for business cycle fluctuations. Hence, we can conclude that the results reported in Figure 1 are mostly driven by the additional restrictions imposed by the PFA.

3 Our Methodology

This section first describes the SVAR. It then discusses the identification problem and the class of sign and zero restrictions considered in this paper. Next, it introduces the structural and the orthogonal reduced-form parameterization. Our algorithms will draw from the orthogonal reduced-form parameterization and then transform the draws to the structural parameterization. Hence, we must be able to transform the prior and posterior distributions from one parameterization to another. The necessary theory to accomplish this is also described in this section. The usual change of variable theorem is not sufficient because we are not transforming between open subsets of Euclidean spaces of the same dimension, but instead are transforming between smooth manifolds of the same dimension. We will briefly outline the volume measure on smooth manifolds and state the appropriate generalizations of the change of variable theorem. Finally, we explicitly specify the conjugate prior distributions that will be used. While the algorithms developed here can be applied to other non-conjugate prior distributions, they are most efficient in the conjugate case.

3.1 The Model

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

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

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 compactly written as

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

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

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

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

3.2 The Identification Problem and Sign and Zero Restrictions

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 that the structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are observationally equivalent if and only if $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{Q}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{Q}$ for some $\mathbf{Q} \in \mathcal{O}(n)$, which is the set of all $n \times n$ orthogonal matrices.

To solve the identification problem, one often imposes sign and/or zero restrictions on either the structural parameters or some function of the structural parameters, like the IRFs. For instance, the element in row i and column j of $(\mathbf{A}_0^{-1})'$ is the contemporaneous response of the ith variable to the jth shock.[3] Restricting this element to be zero would imply that the ith variable does not respond contemporaneously to the jth shock. Restricting this element to be positive would imply that the initial response of the ith variable to the jth shock is positive. The theory and simulation techniques that we develop apply to sign and zero restrictions on any function $\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ from the structural parameters to the space of $r \times n$ 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 every $\mathbf{Q} \in \mathcal{O}(n)$, which is true for IRFs.[4]

To set the notation, let $\mathbf{S}_j$ be a $s_j \times r$ matrix of full row rank, where $0 \leq s_j$, and let $\mathbf{Z}_j$ be a $z_j \times r$ matrix of full row rank, where $0 \leq z_j \leq n - j$ for $1 \leq j \leq n$. The $\mathbf{S}_j$ will define the sign restrictions on the jth structural shock and the $\mathbf{Z}_j$ will define the zero restrictions on the jth structural shock for $1 \leq j \leq n$. In particular, we assume that $\mathbf{S}_j\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j > \mathbf{0}$ and $\mathbf{Z}_j\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j = \mathbf{0}$ for $1 \leq j \leq n$, where $\mathbf{e}_j$ is the jth column of $\mathbf{I}_n$.

In Rubio-Ramírez, Waggoner and Zha (2010), sufficient conditions for identification are established. 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 the jth structural shock, for $1 \leq j \leq n$. In addition,

---

[3]More generally, the IRF of the ith variable to the jth structural shock at horizon k is the element in row i and column j of the matrix $\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+)$, where $\mathbf{L}_0(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{A}_0^{-1})'$ and $\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+) = \sum_{\ell=1}^{\min\{k,p\}} (\mathbf{A}_\ell\mathbf{A}_0^{-1})'\mathbf{L}_{k-\ell}(\mathbf{A}_0, \mathbf{A}_+)$, for $k > 0$. 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 all k and $\mathbf{Q} \in \mathcal{O}(n)$.

[4]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.

there must be at least one sign restriction on the impulse responses to each structural shock.[5] The necessary order condition for identification, Rothenberg (1971), 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 jth structural shock. This means that no matter how many sign restrictions are imposed, identification will only 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.

3.3   The Orthogonal Reduced-Form Parameterization

Equation (2) represents the SVAR in terms of the structural parameterization, which is characterized by $\mathbf{A}_0$ and $\mathbf{A}_+$. Given the discussion in Section 3.2, the SVAR can alternatively be written in what we call the orthogonal reduced-form parameterization. This parameterization is characterized by the reduced-form parameters $\mathbf{B}$ and $\boldsymbol{\Sigma}$ together with an orthogonal matrix $\mathbf{Q}$ and is given by the following equation

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \varepsilon_t'\mathbf{Q}'h(\boldsymbol{\Sigma}) \text{ for } 1 \leq t \leq T, \tag{4}$$

where the $n \times n$ matrix $h(\boldsymbol{\Sigma})$ is any decomposition of the covariance matrix $\boldsymbol{\Sigma}$ satisfying $h(\boldsymbol{\Sigma})'h(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}$. We will take h to be the Cholesky decomposition, though any differentiable decomposition would do.

As we will see, the orthogonal reduced-form parameterization is convenient for drawing. However, the researcher will be interested in making draws from the structural parameterization; thus, we will need to transform $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ into $(\mathbf{A}_0, \mathbf{A}_+)$. Given Equation (2), Equation (4), and the decomposition h, we can define a mapping between $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ by

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

By a direct computation, it is easy to see that $h((\mathbf{A}_0\mathbf{A}_0')^{-1})\mathbf{A}_0$ is an orthogonal matrix. The function f_h is invertible, with inverse defined by

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

The orthogonal reduced-form parameterization makes clear how the structural parameters depend

---

[5]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 impulse responses to the jth structural shock for $1 \leq j \leq k$, then the first k structural shocks under this ordering will be identified.

on the reduced-form parameters and orthogonal matrices. Given the reduced-form parameters and a decomposition h, one can consider each value of $\mathbf{Q} \in \mathcal{O}(n)$ as a particular choice of structural parameters.

3.4 Change of Variable Theorems

As mentioned above, the researcher will be interested in making draws from the structural parameterization but it is simpler to make draws from the reduced-form parameterization and then transform them into the structural parameterization using f_h^{-1}. Hence, it is crucial to understand how to transform densities between these two representations. In this section, we discuss the change of variable theorems that will allow us do exactly that. While we will apply these theorems to the structural parameterization, they can be used with any parameterization as long as the mapping between the orthogonal reduced-form parameterization and the desired parameterization can be explicitly computed.[6] As we will see, there are differences between transforming densities when there are only sign restrictions as opposed to when there are also zero restrictions. Details and proofs, or references to proofs, will be relegated to Appendix A. The usual change of variable theorem can be stated as follows.

Theorem 1. *Let $U \subset \mathbb{R}^b$ be an open set and let $\gamma : U \to \mathbb{R}^b$ be a one-to-one and continuously differentiable function. If $A \subset \gamma(U)$ and $\lambda : A \to \mathbb{R}$ is an integrable function, then*

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \int_{\gamma^{-1}(A)} \lambda(\gamma(\mathbf{u}))|\det(D\gamma(\mathbf{u}))|d\mathbf{u}. \tag{5}$$

Proof. See Appendix A.1. $\square$

Note that the integral on each side of Equation (5) is with respect to Lebesgue measure over $\mathbb{R}^b$. The term $v_\gamma(\mathbf{u}) = |\det(D\gamma(\mathbf{u}))|$ is the volume element of γ at $\mathbf{u}$, where $D\gamma(\mathbf{u})$ denotes the derivative of γ evaluated at $\mathbf{u}$. Note that $D\gamma(\mathbf{u})$ is a $b \times b$ matrix. When the range of γ is not $\mathbb{R}^b$, but is instead a b-dimensional smooth manifold in $\mathbb{R}^a$, one has the following change of variable theorem.[7]

Theorem 2. *Let $U \subset \mathbb{R}^b$ be an open set, let $\mathcal{V} \subset \mathbb{R}^a$ be a b-dimensional smooth manifold, and let $\gamma : U \to \mathcal{V}$ be a one-to-one and continuously differentiable function. If $A \subset \gamma(U)$ and $\lambda : A \to \mathbb{R}$ is an*

---

[6]In particular, it is easy to replicate the theory and algorithms of this paper for the IRF parameterization. This parameterization is characterized by the IRFs of the SVAR and the details are in Appendix B.

[7]A b-dimensional smooth manifold in $\mathbb{R}^a$ is a subset $\mathcal{V}$ of $\mathbb{R}^a$ that admits a local b-dimensional coordinate system in $\mathcal{V}$. This means that for each $\mathbf{v} \in \mathcal{V}$, there is an open set $U \subset \mathbb{R}^b$ and a continuously differentiable function $\gamma : U \to \mathcal{V}$ such that $\gamma(U)$ is open in $\mathcal{V}$, $D\gamma(\mathbf{u})$ is of rank b for every $\mathbf{u} \in U$, the inverse of γ exists and is continuous, and $\mathbf{v} \in \gamma(U)$. The function $\gamma : U \to \mathcal{V}$ is a coordinate system in $\mathcal{V}$ about $\mathbf{v}$ and $\gamma(U)$ is a coordinate patch in $\mathcal{V}$ about $\mathbf{v}$.

integrable function, then

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \int_{\gamma^{-1}(A)} \lambda(\gamma(\mathbf{u}))|\det(D\gamma(\mathbf{u})'D\gamma(\mathbf{u}))|^{\frac{1}{2}}d\mathbf{u}. \tag{6}$$

Proof. See Appendix A.1. □

The integral on the right hand side of Equation (6) is with respect to Lebesgue measure over $\mathbb{R}^b$, but the integral on the left hand side cannot be with respect to Lebesgue measure in $\mathbb{R}^a$ if $a > b$ because $\mathcal{V}$ is of measure zero in $\mathbb{R}^a$. However, the smooth manifold structure of $\mathcal{V}$, together with Lebesgue measure over $\mathbb{R}^b$, uniquely defines the volume of a set in $\mathcal{V}$, which determines a well-defined measure over $\mathcal{V}$ that we call the volume measure. More formally, if $\gamma : U \to \mathcal{V}$ is a coordinate system, $A \subset \gamma(U)$, $\gamma^{-1}(A)$ is Lebesgue measurable over $\mathcal{R}^b$, and $\lambda(\mathbf{v}) = 1$ for every $\mathbf{v} \in A$, then Equation (6) can be taken as the definition of the volume of A. It can be shown that this definition is independent of the choice of coordinate system and can be extended to sets not contained in a single coordinate patch. The integral on the left hand side of Equation (6) is with respect to the volume measure over $\mathcal{V}$.

The matrix $D\gamma(\mathbf{u})$ is $a \times b$, so that $D\gamma(\mathbf{u})'D\gamma(\mathbf{u})$ is a $b \times b$ matrix. As before, the term $v_\gamma(\mathbf{u}) = |\det(D\gamma(\mathbf{u})'D\gamma(\mathbf{u}))|^{\frac{1}{2}}$ is the volume element of γ at $\mathbf{u}$. When $a = b$, Theorem 2 reduces to Theorem 1. As we will see below, Theorem 2 will be used to transform densities when only sign restrictions are considered. The final generalization of the change of variable theorem is given below, which will be of use when there are zero restrictions.

Theorem 3. *Let $U \subset \mathbb{R}^b$ be an open set, let $\mathcal{V} \subset \mathbb{R}^a$ be a d-dimensional smooth manifold, and let the functions $\gamma : U \to \mathbb{R}^a$ and $\beta : U \to \mathbb{R}^{b-d}$ be continuously differentiable with $D\beta(\mathbf{u})$ of rank b-d whenever $\beta(\mathbf{u}) = \mathbf{0}$. Define $\mathcal{U} = \beta^{-1}(\{\mathbf{0}\})$ and suppose that $\gamma(\mathcal{U}) \subset \mathcal{V}$ and γ is one-to-one on $\mathcal{U}$. If $A \subset \gamma(\mathcal{U})$ and $\lambda : A \to \mathbb{R}$ is an integrable function, then*

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \int_{\gamma^{-1}(A)\cap\mathcal{U}} \lambda(\gamma(\mathbf{u}))\,|\det(\mathbf{N}_{\mathbf{u}}' \cdot D\gamma(\mathbf{u})' \cdot D\gamma(\mathbf{u}) \cdot \mathbf{N}_{\mathbf{u}})|^{\frac{1}{2}}\,d\mathbf{u}, \tag{7}$$

where $\mathbf{N}_{\mathbf{u}}$ is any $b \times d$ matrix whose columns form an orthonormal basis for the null space of $D\beta(\mathbf{u})$.

Proof. See Appendix A.1. □

The conditions on the function β, which will be used to describe the zero restrictions, imply that $\mathcal{U}$ is a d-dimensional smooth manifold in $\mathbb{R}^b$ and the integral on the right hand side of Equation (7) is with

respect to the volume measure over $\mathcal{U}$. By assumption, $\gamma(\mathcal{U})$ is contained in the d-dimensional smooth manifold $\mathcal{V}$ and the integral on the left hand side of Equation (7) is with respect to the volume measure over $\mathcal{V}$. The matrix $D\gamma(\mathbf{u})$ is $a \times b$, so that $\mathbf{N'_u} \cdot D\gamma(\mathbf{u})' \cdot D\gamma(\mathbf{u}) \cdot \mathbf{N_u}$ is $d \times d$. The matrix $\mathbf{N_u}$ is not unique. If $\mathbf{N_u}$ and $\tilde{\mathbf{N}}_\mathbf{u}$ are two matrices whose columns form an orthonormal basis for the null space of $D\beta(\mathbf{u})$, then there exists a $d \times d$ orthogonal matrix $\mathbf{X}$ such that $\mathbf{N_u} = \tilde{\mathbf{N}}_\mathbf{u}\mathbf{X}$. Because the determinant of a product of square matrices is equal to the product of the determinants and the determinant of a orthogonal matrix is plus or minus one, the value of the expression $|\det(\mathbf{N'_u} \cdot D\gamma(\mathbf{u})' \cdot D\gamma(\mathbf{u}) \cdot \mathbf{N_u})|$ is independent of the choice of $\mathbf{N_u}$. As before, the term $|\det(\mathbf{N'_u} \cdot D\gamma(\mathbf{u})' \cdot D\gamma(\mathbf{u}) \cdot \mathbf{N_u})|^{\frac{1}{2}}$ is the volume element of γ restricted to $\mathcal{U}$ at $\mathbf{u}$. To emphasize the importance of the restriction, we denote this volume element by $v_{\gamma|_\mathcal{U}}(\mathbf{u})$.

Both Theorems 2 and 3 will be used in subsequent sections in the following way. We will have a distribution over $\mathcal{V}$ whose density with respect to the volume measure over $\mathcal{V}$ evaluated at $\mathbf{v}$ is $p(\mathbf{v})$. A draw $\mathbf{v}$ from this distribution can be uniquely transformed to $\mathbf{u} = \gamma^{-1}(\mathbf{v})$ and we will want to compute the density over the $\mathbf{u}$.[8] When there are only sign restrictions, Theorem 2 will be applicable and the density of $\mathbf{u}$ will be $p(\gamma(\mathbf{u}))v_\gamma(\mathbf{u})$ with respect to Lebesgue measure over $\mathbb{R}^b$. When there are zero restrictions given by β, Theorem 3 will be applicable and the density of $\mathbf{u}$ will be $p(\gamma(\mathbf{u}))v_{\gamma|_\mathcal{U}}(\mathbf{u})$ with respect to the volume measure over $\mathcal{U}$. When applying these theorems, $\mathbf{v}$ will be an orthogonal reduced-form parameter and $\mathbf{u}$ will be a structural parameter. If the researcher is using a parameterization other than the structural parameterization, $\mathbf{u}$ will belong to it instead.

In this section we have discussed change of variable formulas for integration over smooth manifolds with respect to the volume measure. In order to fix ideas, it is useful to relate the volume measure over commonly used smooth manifolds to other measures that could be defined over the same smooth manifolds. Some of these examples will be used later in the paper. First, an open subset $U \subset \mathbb{R}^b$ is a b-dimensional smooth manifold in $\mathbb{R}^b$. The volume measure over U is identical to Lebesgue measure over $\mathbb{R}^b$ in this case.

Second, an open subset of a b-dimensional linear subspace of $\mathbb{R}^a$ is a b-dimensional smooth manifold in $\mathbb{R}^a$ and so there is a well-defined volume measure over these sets. For instance, the set of all $n \times n$ symmetric and positive definite matrices is a $\frac{n(n+1)}{2}$-dimensional smooth manifold in $\mathbb{R}^{n^2}$. If $\mathcal{V} \subset \mathbb{R}^a$ is a b-dimensional linear subspace, we know that there exists a linear mapping γ from $\mathbb{R}^b$ onto $\mathcal{V}$. Because γ is linear, the volume element is constant. Thus, by Theorem 2, the volume of $A \subset \mathcal{V}$ will be the

---

[8]It must be the case that the support of $p(\mathbf{v})$ is contained in $\gamma(U)$ when applying Theorem 2 and contained in $\gamma(\mathcal{U})$ when applying Theorem 3. This ensures that $\mathbf{u} = \gamma^{-1}(\mathbf{v})$ exists and is unique.

Lebesgue measure of $\gamma^{-1}(A) \subset \mathbb{R}^b$ times the constant value of the volume element. Often the constant volume element is ignored, in which case the measure will not be the volume measure. Because of the simple nature of linear subspaces this generally causes no problems; however, in this paper we will always use the volume measure and thus the constant volume element will always be explicitly taken into account.

Third, the set of all $n \times n$ orthogonal matrices, $\mathcal{O}(n)$, is a $\frac{n(n-1)}{2}$-dimensional smooth manifold in $\mathbb{R}^{n^2}$. In addition to the volume measure over $\mathcal{O}(n)$, there are Haar measures defined over $\mathcal{O}(n)$, which is any measure that is invariant to multiplication by orthogonal matrices. Any two Haar measures differ only by a constant scale factor. Because volume is invariant to rigid transformations, which multiplication by an orthogonal matrix is, the volume measure over $\mathcal{O}(n)$ is a Haar measure.

Throughout the rest of the paper all densities will be with respect to the volume measure, even though we will not explicitly state it. Sometimes, for instance when there are no zero restrictions and we are working with $\mathbf{A}_0$, $\mathbf{A}_+$, or $\mathbf{B}$, the volume measure will be Lebesgue measure. However, when we are working with symmetric and positive definite matrices, orthogonal matrices, or when there are zero restrictions, the volume measure will not be Lebesgue.

3.5 Conjugate Priors and Posteriors

While the techniques developed here will work with any prior distribution, they are most efficient when used with prior distributions that belong to a certain family of conjugate distributions.[9] For the reduced-form representation in Equation (3), the normal-inverse-Wishart family of distributions is conjugate.[10] If the prior distribution over the reduced-form parameters is $NIW(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})$, then

---

[9]A family of distributions is conjugate if the prior distribution being a member of this family implies that the posterior distribution is a member of the family. Some authors also require the likelihood to be a member of the family.

[10]A normal-inverse-Wishart distribution over the reduced-form parameters is characterized by four parameters: a scalar $\nu \geq n$, an $n \times n$ symmetric and positive definite matrix $\mathbf{\Phi}$, an $m \times n$ matrix $\mathbf{\Psi}$, and an $m \times m$ symmetric and positive definite matrix $\mathbf{\Omega}$. We denote this distribution by $NIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$ and its density by $NIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma})$. Furthermore,

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

the posterior distribution over the reduced-form parameters is $NIW(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$, where

$$\tilde{\nu} = T + \bar{\nu},$$

$$\tilde{\mathbf{\Omega}} = (\mathbf{X}'\mathbf{X} + \bar{\mathbf{\Omega}}^{-1})^{-1},$$

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

$$\tilde{\mathbf{\Phi}} = \mathbf{Y}'\mathbf{Y} + \bar{\mathbf{\Phi}} + \bar{\mathbf{\Psi}}'\bar{\mathbf{\Omega}}^{-1}\bar{\mathbf{\Psi}} - \tilde{\mathbf{\Psi}}'\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{\Psi}},$$

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

If $\pi(\mathbf{Q}|\mathbf{B}, \mathbf{\Sigma})$ is any conditional density over $\mathcal{O}(n)$, then prior densities of the form $NIW_{(\bar{\nu}, \bar{\mathbf{\Phi}}, \bar{\mathbf{\Psi}}, \bar{\mathbf{\Omega}})}(\mathbf{B}, \mathbf{\Sigma})\pi(\mathbf{Q}|\mathbf{B}, \mathbf{\Sigma})$ over the orthogonal reduced-form parameterization will be conjugate. We will take $\pi(\mathbf{Q}|\mathbf{B}, \mathbf{\Sigma})$ to be the uniform density. We make this choice for three reasons. First, as we will see below priors densities over the orthogonal reduced-form parameterization of this form induce standard prior densities over the structural parameterization. Second, prior and posterior densities over the orthogonal reduced-form parameterization of this form will be very easy to independently draw from. Third, the likelihood is of this form and so this family of densities will be conjugate in even the stronger sense. We call this the uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization; denote it by $UNIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$, and denote its density over the orthogonal reduced-form parameterization by $UNIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$.[11]

Densities over the orthogonal reduced-form parameterization induce densities over the structural parameterization via the function f_h. If $\pi(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ is any density over the orthogonal reduced-form parameterization, then by Theorem 2, the induced density over the structural parameterization will be

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

It is easy to verify that the hypotheses of Theorem 2 are satisfied and so Theorem 2 is applicable. The volume element could be computed numerically, but for any function f_h it can be computed analytically using Proposition 1, described below. The reader should notice that the volume element does not depend on the choice of h.

---

[11]It is the case that $UNIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = NIW_{(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma})/\int_{\mathcal{O}(n)} 1 d\mathbf{Q}$. Because $\mathcal{O}(n)$ is compact, $\int_{\mathcal{O}(n)} 1 d\mathbf{Q}$ is finite.

Proposition 1. *The volume element of f_h at $(\mathbf{A}_0, \mathbf{A}_+)$ is*

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

Proof. See Appendix A.2. □

Using Theorem 2, Proposition 1, and the definition of the normal-inverse-Wishart distribution, the density over the structural parameterization induced by the uniform-normal-inverse-Wishart density over the orthogonal reduced-form parameterization is

$$NGN_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{A}_0, \mathbf{A}_+) = UNIW_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(f_h(\mathbf{A}_0,\mathbf{A}_+))v_{f_h}(\mathbf{A}_0,\mathbf{A}_+)$$

$$\propto \underbrace{|\det(\mathbf{A}_0)|^{\nu-n}e^{-\frac{1}{2}\,\mathrm{vec}(\mathbf{A}_0)'(\mathbf{I}_n\otimes\boldsymbol{\Phi})\,\mathrm{vec}(\mathbf{A}_0)}}_{\text{generalized-normal}}\underbrace{e^{-\frac{1}{2}\,\mathrm{vec}(\mathbf{A}_+-\boldsymbol{\Psi}\mathbf{A}_0)'(\mathbf{I}_n\otimes\boldsymbol{\Omega})^{-1}\,\mathrm{vec}(\mathbf{A}_+-\boldsymbol{\Psi}\mathbf{A}_0)}}_{\text{conditionally normal}}. \quad (8)$$

Thus, if we independently draw $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ from a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization with parameters $(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ and then transform the draws to $(\mathbf{A}_0, \mathbf{A}_+)$ using f_h^{-1} we are in fact independently drawing from the density over the structural parameterization represented in Equation (8). We call this a normal-generalized-normal distribution over the structural parameterization; denote it by $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$, and denote its density over the structural parameterization by $NGN_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{A}_0, \mathbf{A}_+)$. When $\nu = n$ the marginal distribution of $\mathrm{vec}(\mathbf{A}_0)$ is normal with mean zero and variance $\mathbf{I}_n \otimes \boldsymbol{\Phi}^{-1}$. In general we call it a generalized-normal distribution. The distribution of $\mathrm{vec}(\mathbf{A}_+)$, conditional on $\mathbf{A}_0$, is normal with mean $\mathrm{vec}(\boldsymbol{\Psi}\mathbf{A}_0)$ and variance $\mathbf{I}_n \otimes \boldsymbol{\Omega}$. Because the uniform-normal-inverse-Wishart family of distributions is conjugate over the orthogonal reduced-form parameterization, the normal-generalized-normal family of distributions over the structural parameterization is conjugate. This is because if the prior and posterior densities have the same functional form in one parameterization, then, because the volume element will be the same for the prior and posterior densities, the induced prior and posterior densities in the other parameterization will also have the same functional form. Normal-generalized-normal prior distributions over the structural parameterization are often used in the literature, particularly with $\nu = n$. For instance, any prior distribution over the structural parameterization that can be implemented through dummy observations will be of this form. Thus, the Sims-Zha prior distribution over the structural parameterization is also of this form.

If the researcher needs to work with a parameterization other than the structural parameteriza-

tion, an analog to Equation (8) can easily be obtained as long as the mapping between the orthogonal reduced-form parameterization and the desired parameterization can be explicitly computed. However, it may not be possible to derive an analytical expression for the volume element as in Proposition 1, but Theorem 2 can always be used to numerically compute the density over the desired parameterization induced by the uniform-normal-inverse-Wishart density over the orthogonal reduced-form parameterization.

It is very easy to independently draw from the uniform-normal-inverse-Wishart distribution. Matlab, Mathematica, and R have routines for making independent draws from both the inverse-Wishart distribution and the normal distribution. There are efficient algorithms for making independent draws from the uniform distribution over $\mathcal{O}(n)$. Faust (1998), Canova and Nicoló (2002), Uhlig (2005), and Rubio-Ramírez, Waggoner and Zha (2010) all propose algorithms to do this. The algorithm of Rubio-Ramírez, Waggoner and Zha (2010) is the most efficient, particularly for larger SVAR systems (e.g., $n > 4$).[12] Rubio-Ramírez, Waggoner and Zha's (2010) results are based on the following theorem.

Theorem 4. *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 uniform distribution over $\mathcal{O}(n)$.*

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

In this section we have explicitly derived expressions for prior densities over the structural parameterization that are conjugate and are induced by a uniform-normal-inverse-Wishart prior density over the orthogonal reduced-form parameterization. Furthermore, they are a family of prior densities often used in the literature.

4   Sign Restrictions

Because $\mathbf{F}$ is continuous, the set of all structural parameters satisfying the sign restrictions will be open in the set of all structural parameters. An important point to make here is that the condition $\mathbf{F}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) = \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{Q}$ for every $\mathbf{Q} \in \mathcal{O}(n)$ and the regularity condition are only needed to implement the algorithms for zero restrictions to be presented later. When only sign restrictions

---

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

are considered, it is enough to assume that $\mathbf{F}$ is continuous. So, if the sign restrictions are non-degenerate, so that there is at least one parameter value satisfying the sign restrictions, then the set of all structural parameters satisfying the sign restrictions will be of positive measure in the set of all structural parameters. This justifies algorithms of the following type.

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

1. *Draw $(\mathbf{B}, \mathbf{\Sigma})$ independently from the $NIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$ distribution.*

2. *Draw $\mathbf{Q}$ independently from the uniform distribution over $\mathcal{O}(n)$ using Theorem 4.*

3. *Keep $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ if the sign restrictions are satisfied.*

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

Algorithm 1 follows the steps highlighted in Section 3.4: it draws from a distribution over the orthogonal reduced-form parameterization conditional on the sign restrictions and then transforms the draws into the structural parameterization using f_h^{-1}. It follows from the discussion in Section 3.5 that the independent draws of $(\mathbf{A}_0, \mathbf{A}_+)$ produced by Algorithm 1 will be from the $NGN(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$ distribution over the structural parameterization conditional on the sign restrictions.

As argued by Baumeister and Hamilton (2015), one should choose the parameterization so that at least some, if not all, of the parameters have an economic interpretation and the prior distribution over the selected parameterization should be chosen to reflect what economic theory has to say about those parameters. Usually, this parameterization will not be the orthogonal reduced-form parameterization since those parameters, particularly the orthogonal matrix, $\mathbf{Q}$, are hard to interpret from an economic point of view. Although we agree with Baumeister and Hamilton (2015), we will draw from the orthogonal reduced-form parameterization because Algorithm 1 is relatively efficient. We will then transform the orthogonal reduced-form draws back to the desired parameterization. While Algorithm 1 is stated in terms of the structural parameterization, it will work for any parameterization as long as one can explicitly compute the transformation between the orthogonal reduced-form and the desired parameterization and the draws produced by this algorithm will be from a conjugate distribution over the desired parameterization. Clearly, if the desired parameterization is the orthogonal reduced-form parameterization, the transformation is the identity and Algorithm 1 can be used to produce independent draws of $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ from the $UNIW(\nu, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega})$ distribution over the orthogonal reduced-form parameterization conditional on the sign restrictions.

Given the desired parameterization, conjugate prior distributions have many useful properties and if one wants to use a conjugate prior distribution that is induced by a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization, then Algorithm 1 is an efficient technique for making independent draws from the associated posterior distribution. If one wants to use a different prior distribution, then Algorithm 1 can still be used as the proposal density in an importance sampler. Of course the efficiency of this algorithm depends heavily on how close such a prior distribution is to one that can be induced by a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization. In this paper we work only with conjugate priors distributions over the structural parameterization whose density can be described by Equation (8) as they are commonly used in the literature, as mentioned in Section 3.5.

5   Zero Restrictions

We now adapt Algorithm 1 to handle the case of sign and zero restrictions. When there are only sign restrictions, the set of all structural parameters satisfying the restrictions is of positive measure in the set of all structural parameters. However, when there are both sign and zero restrictions, the set of all structural parameters satisfying the restrictions is of measure zero in the set of all structural parameters. This invalidates the direct use of Algorithm 1. But since the set of all structural parameters satisfying both the sign and zero restrictions is of positive measure in the set of all structural parameters satisfying the zero restrictions, if we could make independent draws from the set of all structural parameters satisfying the zero restrictions, then we could apply a variant of Algorithm 1 to obtain independent draws from the set of all structural parameters satisfying both the sign and zero restrictions. Algorithm 3, described in this section, does precisely that.

5.1   Zero Restrictions in the Orthogonal Reduced-Form Parameterization

The zero restrictions in the structural parameterization are $\mathbf{Z}_j\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_j = \mathbf{0}$ for $1 \leq j \leq n$. From the definition of f_h and the fact that $\mathbf{F}(\mathbf{A}_0\mathbf{Q}, \mathbf{A}_+\mathbf{Q}) = \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{Q}$, the zero restrictions in the orthogonal reduced-form parameterization are

$$\mathbf{Z}_j\mathbf{F}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}))\mathbf{e}_j = \mathbf{Z}_j\mathbf{F}(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{I}_n))\mathbf{Q}\mathbf{e}_j = \mathbf{0} \text{ for } 1 \leq j \leq n.$$

This means that the zero restrictions in the orthogonal reduced-form parameterization are really just linear restrictions on each column of the orthogonal matrix $\mathbf{Q}$, conditional on the reduced-form parameters $(\mathbf{B}, \boldsymbol{\Sigma})$. It is this observation that is key to being able to make independent draws from the set of all structural parameters satisfying the zero restrictions.

Algorithm 2. *The following makes independent draws from a distribution over the structural parameterization conditional on the zero restrictions.*

1. *Draw* $(\mathbf{B}, \boldsymbol{\Sigma})$ *independently from the* $NIW(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ *distribution.*

2. *For* $1 \leq j \leq n$, *draw* $\mathbf{x}_j \in \mathbb{R}^{n+1-j-z_j}$ *independently from a standard normal distribution and set* $\mathbf{w}_j = \mathbf{x}_j / \parallel \mathbf{x}_j \parallel$.

3. *Define* $\mathbf{Q} = [\mathbf{q}_1 \cdots \mathbf{q}_n]$ *recursively by* $\mathbf{q}_j = \mathbf{K}_j \mathbf{w}_j$ *for any matrix* $\mathbf{K}_j$ *whose columns form an orthonormal basis for the null space of the* $(j - 1 + z_j) \times n$ *matrix*

$$\mathbf{M}_j = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{j-1} & (\mathbf{Z}_j \mathbf{F}(f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{I}_n)))' \end{bmatrix}'.$$

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

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

The null space of $\mathbf{M}_j$ will be of dimension $n + 1 - j - z_j$ if and only if $\mathbf{M}_j$ is of full row rank. Because of the regularity condition on the function $\mathbf{F}$, this will always be case. The details of this argument appear in Appendix A.3. This is crucial because otherwise the product $\mathbf{K}_j \mathbf{w}_j$ is not defined. It is also the case that matrix $\mathbf{K}_j$ is not unique. If the columns of $\mathbf{K}_j$ form an orthonormal basis for the null space of $\mathbf{M}_j$, then so will the columns of $\mathbf{K}_j \mathbf{X}$ for any $\mathbf{X} \in \mathcal{O}(n + 1 - j - z_j)$. The particular choice of $\mathbf{K}_j$ does not make a material difference in the output of Algorithm 2, but in the next section, when we compute the density over the structural parameterization conditional on the zero restrictions implied by Algorithm 2, we will need the function $\mathbf{K}_j = \mathbf{K}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ to be differentiable almost everywhere.[13] In Appendix A.3 we define $\mathbf{K}_j$ so that it is differentiable almost everywhere.

---

[13]The function $\mathbf{K}_j$ depends on $f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{I}_n)$, and so implicitly requires $\boldsymbol{\Sigma}$ to be symmetric and positive definite. Thus the domain of $\mathbf{K}_j$ is not an open set in $\mathbb{R}^{n(m+n+j-1)}$. In Appendix A.2, we extend the definition of f_h^{-1} so that the domain of $\mathbf{K}_j$ is an open set and the derivative can be defined. Also, in general, it is not possible to define $\mathbf{K}_j$ so that it is differentiable everywhere. For instance, if there are no restrictions, then the existence of everywhere continuous $\mathbf{K}_j$ would imply that $\mathcal{O}(n)$ is topologically equivalent to a product of spheres, which is not true if $n \geq 3$.

Our implementation is quite straight forward and the computation of each $\mathbf{K}_j$ requires only a single QR-decomposition of an $n \times n$ invertible matrix.[14]

By construction, the vector $\mathbf{q}_j$ is perpendicular to the rows of $\mathbf{M}_j$ and $\| \mathbf{q}_j \| = \| \mathbf{w}_j \| = 1$. Thus the matrix $\mathbf{Q}$ obtained in Steps 2 and 3 is orthogonal and $\mathbf{Z}_j \mathbf{F}(f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{I}_n)) \mathbf{Q} \mathbf{e}_j = \mathbf{0}$ for $1 \leq j \leq n$. So, the algorithm produces independent draws from a distribution over the structural parameterization conditional on the zero restrictions. In the next subsection, we show how to numerically compute the density of this distribution using Theorem 3. Unlike the sign restriction only case, the distribution over the structural parameterization conditional on the zero restrictions implied by Algorithm 2 is not equal to the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution conditional on the zero restrictions. However, once we know how to numerically compute its density, we can use Algorithm 2 as a proposal distribution for an importance sampler to draw from the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution over the structural parameterization conditional on the zero restrictions.

Because Algorithm 2 will be used as a proposal distribution for an importance sampler for a distribution whose support is the set of all structural parameters satisfying the zero restrictions, it must be the case that the support of the distribution implied by Algorithm 2 is also the set of all structural parameters that satisfy the zero restrictions. To see that this is the case, suppose that $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ satisfies the zero restrictions. To show that these parameters are in the support of the distribution implied by Algorithm 2, it suffices to show that there exist $\mathbf{w}_j \in \mathbb{R}^{n+1-j-z_j}$, for $1 \leq j \leq n$, such that Step 3 of Algorithm 2 maps to the orthogonal matrix $\mathbf{Q}$. Let $\mathbf{w}_j = \mathbf{K}_j' \mathbf{q}_j$. Since $f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ satisfies the zero restrictions and the matrix $\mathbf{Q}$ is orthogonal, the jth column of $\mathbf{Q}$ is in the null space of $\mathbf{M}_j$. Thus $\mathbf{K}_j \mathbf{w}_j = \mathbf{K}_j \mathbf{K}_j' \mathbf{q}_j = \mathbf{q}_j$, because multiplication by $\mathbf{K}_j \mathbf{K}_j'$ is projection onto the null space of $\mathbf{M}_j$.

Algorithm 2 also follows the steps highlighted in Section 3.4: it draws from a distribution over the orthogonal reduced-form parameterization conditional on the zero restrictions and then transforms the draws into the structural parameterization using f_h^{-1}. As mentioned, in this case the independent draws of $(\mathbf{A}_0, \mathbf{A}_+)$ produced by Algorithm 2 will not be from the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution over the structural parameterization conditional on the zero restrictions. The density implied by Algorithm 2 will be analyzed below.

---

[14]In Matlab, an obvious choice would be to define $\mathbf{K}_j = \text{null}(\mathbf{M}_j)$. While it is surely the case that this choice would be differentiable almost everywhere, to prove this would require details of the Matlab implementation of this function.

5.2 The Density Implied by Algorithm 2

In this section we use Theorem 3 to show how to numerically compute the distribution over the structural parameterization conditional on the zero restrictions implied by Algorithm 2. In order to do that we need to carefully characterize the mapping implied by the steps in the algorithm. Step 1 of Algorithm 2 independently draws $\mathbf{B}$ and $\boldsymbol{\Sigma}$ from the $NIW(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution. Step 2 draws $\mathbf{w}_j$ from the uniform distribution on the unit sphere in $\mathbb{R}^{n+1-j-z_j}$. This implies that the density over $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{w}_1, \cdots, \mathbf{w}_n)$ will be proportional to $NIW_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{B}, \boldsymbol{\Sigma})$. Step 3 maps $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{w}_1, \cdots, \mathbf{w}_n)$ to $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ and Step 4 maps $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ to $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. It is this composite mapping, together with Theorem 3, that we will use to compute the density.

It will be shown in Appendix A.3 that there exists an open set $V \subset \mathbb{R}^{nm+n^2+n^2}$ such that the functions $\mathbf{K}_j = \mathbf{K}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ can be defined so that they are differentiable for all $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) \in V$.[15] Thus, we can define a differentiable function $g : V \to \mathbb{R}^{nm+n^2+\sum_{j=1}^{n}(n+1-j-z_j)}$ by

$$g(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = (\mathbf{B}, \boldsymbol{\Sigma}, (\mathbf{K}_1(\mathbf{B}, \boldsymbol{\Sigma})'\mathbf{q}_1, \cdots, \mathbf{K}_n(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{n-1})'\mathbf{q}_n)).$$

On the set of all $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) \in V$ such that $\boldsymbol{\Sigma}$ is symmetric and positive definite, $\mathbf{Q}$ is orthogonal, and $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ satisfies the zero restrictions, the function g will be one-to-one. The easiest way to see this is that the function defined by Step 3 of Algorithm 2 is the inverse of g on this restricted set. The argument is identical to that used to show that the support of the distribution implied by Algorithm 2 is the set of all structural parameters satisfying the zero restrictions. Let $U = f_h^{-1}(V)$, which will be an open set in the set of all structural parameters. The composite function $g \circ f_h$, when restricted to the $(\mathbf{A}_0, \mathbf{A}_+) \in U$ that satisfy the zero restrictions, will be the inverse of the function defined by Steps 3 and 4 of Algorithm 2. If $\mathcal{Z}$ denotes the set of all structural parameters that satisfy the zero restrictions, then by Theorem 3 the density over the structural parameterization conditional on the zero restrictions implied by Algorithm 2 is proportional to $NIW_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{B}, \boldsymbol{\Sigma})v_{(g \circ f_h)|_{\mathcal{Z}}}(\mathbf{A}_0, \mathbf{A}_+)$, where $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = f_h(\mathbf{A}_0, \mathbf{A}_+)$. It is easy to verify that the hypotheses of Theorem 3 are satisfied. The only difficulty is to check whether the derivative of the function describing the zero restrictions, which is given by $\beta(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{Z}_j \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+) \mathbf{e}_j)_{j=1}^n$, has full row rank. This will follow from the regularity conditions on the function $\mathbf{F}$, the details of which are in Appendix A.3.

Finally, it must be the case that any structural parameter satisfying the zero restrictions must

---

[15]Here, we are implicitly considering $\mathbf{K}_j$ to be a function of $\mathbf{Q}$, even though it actually only depends on the first $j-1$ columns of $\mathbf{Q}$.

almost surely be in the set U. If this were not the case, then there would be a set of positive measure for which the techniques of the section would not apply. As with the other details in this section, this will be shown in Appendix A.3.

5.3 An Importance Sampler

The results of Sections 5.1 and 5.2 show that, first, Algorithm 2 generates independent draws from a distribution over the structural parameterization conditional on the zero restrictions that is not equal to the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution conditional on the zero restrictions and, second, we know how to to numerically compute this density. Thus, they justify the following importance sampler algorithm to independently draw from the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution over the structural parameterization conditional on the sign and zero restrictions.

Algorithm 3. *The following algorithm independently draws from the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution over the structural parameterization conditional on the sign and zero restrictions.*

1. *Use Algorithm 2 to independently draw $(\mathbf{A}_0, \mathbf{A}_+)$.*

2. *If $(\mathbf{A}_0, \mathbf{A}_+)$ satisfies the sign restrictions, then set its importance weight to*

$$\frac{NGN_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{A}_0, \mathbf{A}_+)}{NIW_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\mathbf{B}, \boldsymbol{\Sigma}) v_{(g \circ f_h)|\mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)} \propto \frac{|\det(\mathbf{A}_0)|^{-(2n+m+1)}}{v_{(g \circ f_h)|\mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)},$$

   where $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = f_h(\mathbf{A}_0, \mathbf{A}_+)$ and $\mathcal{Z}$ denotes the set of all structural parameters that satisfy the zero restrictions. Otherwise, set its importance weight to zero.

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

As was the case with Algorithms 1 and 2, Algorithm 3 follows the steps highlighted in Section 3.4: it draws from a distribution over the orthogonal reduced-form parameterization conditional on the sign and zero restrictions and then transforms the draws into the structural parameterization. It follows from the discussion in Section 5.2 that the independent draws of $(\mathbf{A}_0, \mathbf{A}_+)$ produced by Algorithm 3 will be from the $NGN(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ distribution over the structural parameterization conditional on the sign and zero restrictions.

Algorithm 3 inherits the key features of Algorithm 1. First, being able to independently draw $(\mathbf{A}_0, \mathbf{A}_+)$ from the normal-generalized-normal family of distributions over the structural parameterization conditional on the sign and zero restrictions means that we can use Algorithm 3 to independently

20

draw from the posterior distribution over the structural parameterization conditional on the sign and zero restrictions for any conjugate prior distribution from this family.

Second, Algorithm 3 is stated in terms of the structural parameterization, but it will work for any parameterization as long as one can explicitly compute the transformation between the orthogonal reduced-form parameterization and the desired parameterization. Because Algorithm 3 first draws from the uniform-normal-inverse-Wishart distribution, which is conjugate over the orthogonal reduced-form parameterization, the draws produced by this algorithm will be from a conjugate distribution over the desired parameterization.

Finally, Algorithm 3 is very efficient if one uses a conjugate prior distribution over the desired parameterization that can be induced by a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization. If one wants to use a prior distribution other than one that can be induced by a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization, then Algorithm 3 should be modified accordingly. As before, the efficiency of this algorithm depends heavily on how close the used prior distribution is to a conjugate prior distribution that can be induced by a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization.

It is also important to notice that computing the volume element $v_{(g \circ f_h)|\mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)$ in Step 2 is the most expensive part in implementing Algorithm 3. The rest of Algorithm 3 is quite fast. But the reader should note that we do not need to compute $v_{(g \circ f_h)|\mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)$ for all the draws, only for those that satisfy the sign restrictions. Numerical and timing issues associated with the computation of the volume elements will be analyzed in Section 6.

If there are no zero restrictions, Algorithm 1 and Algorithm 3 both produce draws from the same distribution over the structural parameterization conditional on the sign restrictions. It would seem that Algorithm 1 would be much more efficient since it is not an importance sampler and no weights would have to be computed. However, it is not hard to numerically verify that when there are no zero restrictions, the weights are constant and so Algorithm 3 is not really an importance sampler in this case either. Still, computing a QR-decomposition is more efficient than the recursive procedure in Step 3 of Algorithm 2, so Algorithm 1 should certainly be used in this case.

5.4 The Density Implied by Algorithm 2 Revisited

Algorithm 3 draws from the density $NGN_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{A}_0, \mathbf{A}_+)$ over the structural parameterization conditional on the zero restrictions, while Algorithm 2 draws from the density $NIW_{(\nu,\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Omega})}(\mathbf{B}, \mathbf{\Sigma})v_{(g \circ f_h)|\mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)$ over the structural parameterization conditional on the zero restrictions, where $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = f_h(\mathbf{A}_0, \mathbf{A}_+)$ and $\mathcal{Z}$ denotes the set of all structural parameters that satisfy the zero restrictions.[16] Given that the main expense of Algorithm 3 is computing the importance weights, one might be tempted to dispense with Algorithm 3 and simply use Algorithm 2, particularly if one was not wedded to using the normal-generalized-normal distribution over the structural parameterization for the prior and posterior. While this is permissible, the researcher should be aware of two unpleasant features of the distribution over the structural parameterization conditional on the zero restrictions produced by Algorithm 2. First, this distribution is not invariant to a reordering of the shocks.[17] Second, it does not respect the imposition of additional zero restrictions. Algorithm 3 does not suffer from either of these problems. We expand on each of these below.

It is easiest to see each of the above features in the context of a simple example. Consider an SVAR with 3 variables and one lag, without a constant. Suppose further that there were two zero restrictions on the contemporaneous IRF given by

$$\mathbf{Z}_1 \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_1 = \mathbf{0} \text{ and } \mathbf{Z}_2 \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_2 = \mathbf{0} \tag{9}$$

where

$$\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+) = \left(\mathbf{A}_0^{-1}\right)', \ \mathbf{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \text{ and } \mathbf{Z}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}. \tag{10}$$

Let $\nu = n$, $\mathbf{\Phi} = \mathbf{I}_n$, $\mathbf{\Psi} = \mathbf{0}_n$, and $\mathbf{\Omega} = \mathbf{I}_n$. Here, we will interpret this density as a prior density over the structural parameterization conditional on the zero restrictions, but the posterior would have the same functional form in this case as well.

We first consider what happens if the same restrictions were imposed in a different order. So, instead of the zero restrictions defined by Equation (9), suppose one imposed the zero restrictions given by

$$\mathbf{Z}_2 \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_1 = \mathbf{0} \text{ and } \mathbf{Z}_1 \mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_2. \tag{11}$$

---

[16]In this section we are going to compare the outcomes of Algorithm 2 and Algorithm 3. To simplify the presentation we are going to assume that there are no sign restrictions.

[17]We thank Tom Doan for pointing out this fact in an earlier version of the paper.

Clearly, this simply interchanges the roles of the first and second shocks. Let $\mathcal{Z}_{\mathcal{R}_1}$ denote the set of all structural parameters satisfying the restrictions defined by Equation (9) and let $\mathcal{Z}_{\mathcal{R}_2}$ denote the set of all structural parameters satisfying the restrictions defined by Equation (11). If $(\mathbf{A}_0, \mathbf{A}_+) \in \mathcal{Z}_{R_1}$, then swapping the first and second columns of both $\mathbf{A}_0$ and $\mathbf{A}_+$ will produce an element of $\mathcal{Z}_{R_2}$. To illustrate that the order makes a difference, we make ten draws from $\mathcal{Z}_{\mathcal{R}_1}$ using Algorithm 2 and compute their densities using $NIW_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{B},\boldsymbol{\Sigma})v_{(g\circ f_h)|_{\mathcal{Z}_{R_1}}}(\mathbf{A}_0,\mathbf{A}_+)$, where $(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q}) = f_h(\mathbf{A}_0,\mathbf{A}_+)$. We then swap the first and second columns of both $\mathbf{A}_0$ and $\mathbf{A}_+$ and compute their densities using $NIW_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{B},\boldsymbol{\Sigma})v_{(g\circ f_h)|_{\mathcal{Z}_{R_2}}}(\mathbf{A}_0,\mathbf{A}_+)$, where $(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q}) = f_h(\mathbf{A}_0,\mathbf{A}_+)$. Their ratios are reported in Table 1.[18] If the order did not matter, then all the entries would be one. Even though only ten draws were analyzed, we see that the ordering changes the density of all the draws, with some changing by quite a lot.

Draw	1	2	3	4	5	6	7	8	9	10
	1.17	1.20	0.98	0.66	0.65	1.02	8.45	0.96	1.22	1.06

Table 1: Ratio of densities for ten draws of the structural parameters using differing shock orderings. The ratio is computed with the density in $\mathcal{Z}_{R_1}$ on the numerator and the density in $\mathcal{Z}_{R_2}$ on the denominator.

Next, we show what happens if we impose additional zero restrictions. In the context of our example, let $\mathcal{Z}_{R_3}$ denote the set of all structural parameters satisfying the following zero restriction

$$\mathbf{Z}_1\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_1 = \mathbf{0}.$$

Of course, if $(\mathbf{A}_0, \mathbf{A}_+) \in \mathcal{Z}_{R_1}$, then $(\mathbf{A}_0, \mathbf{A}_+) \in \mathcal{Z}_{R_3}$. Hence, for the ten draws analyzed in Table 1, we can compute the ratio of $NIW_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{B},\boldsymbol{\Sigma})v_{(g\circ f_h)|_{\mathcal{Z}_{R_1}}}(\mathbf{A}_0,\mathbf{A}_+)$ to $NIW_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\mathbf{B},\boldsymbol{\Sigma})v_{(g\circ f_h)|_{\mathcal{Z}_{R_3}}}(\mathbf{A}_0,\mathbf{A}_+)$, where $(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q}) = f_h(\mathbf{A}_0,\mathbf{A}_+)$. The ratios are reported in Table 2. If adding new restrictions did not matter, then all the entries in the table would be one. Even though only ten draws were analyzed, we see that adding restrictions changes the density of all the draws, with some changing by quite a lot.

Draw	1	2	3	4	5	6	7	8	9	10
	0.74	0.62	0.52	2.13	1.58	1.11	1.37	0.55	1.21	0.65

Table 2: Ratio of densities for ten draws of the structural parameters after adding zero restrictions. The ratio is computed with the density in $\mathcal{Z}_{R_1}$ on the numerator and the density in $\mathcal{Z}_{R_3}$ on the denominator.

This means that when using distributions over the structural parameterization conditional on the

---

[18]In this section, the densities are only computed up to a constant, but in this case the constant is the same across both densities. So, the ratios are correctly reported.

zero restrictions implied by Algorithm 2 as priors, including additional restrictions not only adds the restrictions, restricting the support of the distribution, but changes the prior distribution over the structural parameterization as well. So, if we saw the marginal likelihood increase with the addition of restrictions, it is hard to disentangle whether the data favored the new restrictions or favored the new prior distribution.[19]

6 Numerical Issues

While Algorithm 3 is fairly straightforward to apply, as with any numerical algorithm, there are different implementations and these differences can affect both the efficiency and the accuracy of the algorithm. In this section we highlight some of these issues. We first describe how to compute the numerical derivatives associated with the computation of the volume element $v_{(g \circ f_h)|_{\mathcal{Z}}}(\mathbf{A}_0, \mathbf{A}_+)$. Second, we will highlight how computing these volume elements is expensive, so we need to be judicious when doing so. Finally, we will talk about the effective sample size associated with the importance sampling embedded in Algorithm 3.

6.1 Numeric Derivatives

As mentioned, computing the volume element $v_{(g \circ f_h)|_{\mathcal{Z}}}(\mathbf{A}_0, \mathbf{A}_+)$ in Step 2 is the most expensive part in implementing Algorithm 3. Both of these can be computed using Theorem 3, which requires the computation of two derivatives. Generically, if $f : \mathbb{R}^b \to \mathbb{R}^a$, then Df is an $a \times b$ matrix and the jth column of Df can be approximated by $(f(\mathbf{x} + \varepsilon \mathbf{e}_j) - f(\mathbf{x}))/\varepsilon$, where $\mathbf{e}_j$ is the jth column of the $b \times b$ identity matrix. This is called the one-sided approximation and requires $b + 1$ function evaluations. Alternatively, one could approximate using $(f(\mathbf{x} + \varepsilon \mathbf{e}_j) - f(\mathbf{x} - \varepsilon \mathbf{e}_j))/(2\varepsilon)$. This is the two-sided approximation and requires $2b$ function evaluations. In general, the two-sided approximation is more accurate but almost twice as expensive (see Section 6.2). In most applications, we find that the one-sided approximation is good enough, but one must keep in mind that this might not always be the case. Before choosing which approach to use, we recommend that in preliminary runs, one try both and only use the two-sided approximation if the results vary between the two techniques. In Section

---

[19]The problems with Algorithm 2 described here are bound to appear in any algorithm that makes draws in the orthogonal reduced-form parameterization and then transforms them into the desired parameterization abstracting from computing the relevant volume elements. For example, Gambacorta, Hofmann and Peersman (2014), Baumeister and Benati (2013), and Binning (2013) would suffer from the issues described.

8.4 we show that, for our application, using one-sided derivatives affects the accuracy of neither the computation of the volume elements nor the IRFs.

After the choice of technique has been made, we also need to choose the size of the tolerance, which is ε in the previous paragraph. We recommend values between 10^{-4} and 10^{-7}, with 10^{-6} as a good starting point. Given the complexities of the function $g \circ f_h$ we do not recommend values smaller than square root machine epsilon, which is approximately 10^{-7} when using double precision.

6.2   Judicious Evaluations of Volume Elements

Since Step 2 is the most expensive part of Algorithm 3, it is important to compute the volume elements only if the sign restrictions hold. For example, assume we want to identify a seven-variables system with 12 lags using three sign and three zero restrictions. If we apply Algorithm 3 naively and we compute the volume elements for all the iterations and not only if the sign restrictions hold, it takes 4,806 seconds (80.1 minutes) to do 10,000 iterations of Algorithm 3.[20] This exercise is represented by Timing 2 in the last column of Table 3. Of this time, 4,795 seconds (79.9 minutes) is devoted to computing the volume elements, which is Step 2, and, as represented by Timing 1 in the same column of the table, just 11 seconds are used by all the other steps combined. For this reason, it is critical to keep the volume element evaluation to a minimum. In this example, about 20 percent of the iterations satisfy the sign restrictions so if we evaluate the volume elements only for independent draws that satisfy the sign restrictions, the run time drops to under 1040 seconds (17.3 minutes), a significant savings. This exercise is represented by Timing 4 in the last column of Table 3. Efficiency becomes crucial as the number of sign restrictions increases, so that the percentage of draws satisfying the sign restrictions decreases. For instance, if only 1 percent of the iterations were to satisfy the sign restrictions, then it would require approximately 1,000,000 iterations of Algorithm 3 to produce 10,000 independent draws satisfying the sign and zero restrictions. If the volume elements were computed for all the iterations, this would take more than five full days to complete. However, if the volume elements were evaluated only for those draws satisfying the sign restrictions, then 10,000 independent draws satisfying the sign and zero restrictions could be produced in less than two hours.

All the timings mentioned above were for a two-sided derivative technique. Using a one-sided derivative decreases the computation time by more than 40 percent. For instance, instead of taking 1040 seconds to run 10,000 iterations of Algorithm 3, it only took 544 seconds. This exercise is

---

[20]This was programmed in Matlab and run on a MacBook Pro with a 2.3GHz Intel i7 processor and 16GB of RAM.

	4 lags + 3 signs + 1 zero			12 lags + 3 signs + 3 zeros		
	$n = 5$	$n = 6$	$n = 7$	$n = 5$	$n = 6$	$n = 7$
Timing 1	7	8	9	8	9	11
Timing 2	707	1123	1661	1954	3162	4806
Timing 3	367	576	853	1002	1646	2521
Timing 4	115	186	265	440	676	1040
Timing 5	62	98	140	231	351	544
Draws satisfying signs	1542	1592	1546	2231	2103	2116
Effective sample size*	0.79	0.84	0.82	1.00	1.00	1.00

Table 3: Time is reported in seconds. Timing 1 computes how long it takes to run Algorithm 3 when Step 5 is not computed. Timing 2 evaluates the same procedure as Timing 1 but computing the weights for every draw, even if they do not satisfy the sign restrictions. Timing 3 evaluates the same procedure as Timing 2 but using a one-sided derivative. Timing 4 computes the timing of Algorithm 3, i.e., computing the weights only if the sign restrictions are satisfied. Timing 5 evaluates the same procedure as Timing 4 but using a one-sided derivative. Draws satisfying signs refers to the number of iterations satisfying the sign restrictions. The letter n denotes the total number of variables in the system. *Effective sample size as a share of the draws satisfying the sign and zero restrictions.

represented by Timing 5 in the last column of Table 3. Timing 3 represents the time for the naive implementation of the algorithm when a one-sided derivative technique is used. As we will see in Section 8.4, the effects of using a one-sided derivative instead of a two-sided derivative are not reflected in the IRFs.

The results just analyzed correspond to a seven-variables SVAR with 12 lags using three sign and three zero restrictions. Table 3 also reflects the times for five additional systems: two other SVARs with 12 lags using three sign and three zero restrictions (one with six variables and another with five) and three SVARs with four lags using three sign restrictions and one zero restriction (in seven-, six-, and five-variable systems). As the reader can see, the conclusions are pretty robust to changes in the size of the SVAR.

6.3 Effective Sample Size

When using importance sampling, one should always compute the effective sample size to guard against having only a few draws dominate the sample. If w_i is the weight associated with the ith draw, then the effective sample size is $\left(\left(\sum_{i=1}^{N} w_i \right)^2 \right) / \left(\sum_{i=1}^{N} w_i^2 \right)$, where N is the total number of draws.

The effective sample size as a share of the draws satisfying the sign and zero restrictions is always between 0 and 1. As its formula implies, the effective sample size should be interpreted as the actual number of independent draws produced by the importance sampler. If the weights were all equal, then

the effective sample size as a share of the draws satisfying the sign and zero restrictions would be 1.

Because the weight of a draw that does not satisfy the sign restrictions is zero, for Algorithm 3 the maximum effective sample size will be the number of draws satisfying the sign restrictions. In most of the applications that we have run, the effective sample size was close to the number of draws satisfying the sign restrictions, but one should always compute this diagnostic. In particular, for the systems represented in Table 3 the effective sample size as a share of draws satisfying the sign and zero restrictions was always greater than or equal to 0.79.

7   The Penalty Function Approach

In this section, we discuss the PFA developed by Mountford and Uhlig (2009). We briefly describe the PFA and mention its drawbacks as reported in the literature. We choose the PFA to compare our results because the PFA is commonly used to implement sign and zero restrictions. The PFA consists of using a loss function to find an orthogonal matrix that satisfies the zero restrictions and that satisfies or comes close to satisfying the sign restrictions; see Mountford and Uhlig (2009) for details. In the PFA context, we call this the optimal orthogonal matrix. The literature has mentioned several issues with this approach. First, by choosing a single orthogonal matrix the PFA is not set identifying the SVAR. This drawback is important because the robustness associated with set identification is one of the most appealing motivations for using sign and zero restrictions instead of more traditional approaches. Second, the optimal orthogonal matrix may be such that the sign restrictions do not hold.[21] Third, as acknowledged by Uhlig (2005), the PFA rewards orthogonal matrices that imply responses strongly satisfying the sign restrictions. This rewarding scheme can result in imposing additional sign restrictions. For example, Caldara and Kamps (2012) use a bivariate SVAR to show how the PFA restricts the output response to a tax increase to be negative. Finally, Binning (2013) points out that in those cases in which several structural shocks are identified using the PFA, the ordering on which the structural shocks are identified determines their importance. For the reasons mentioned above, Uhlig (2005) and Caldara and Kamps (2012) conclude that the PFA should be interpreted as incorporating additional identifying restrictions.

All these issues can be summarized by saying that in the case of the PFA, the identification does not solely come from the sign and zero restrictions. The prior density over the structural parameterization implied by the PFA is such that, for every value of the reduced-form parameters, a single value

---

[21]This is true even in the extreme case in which no orthogonal matrix satisfies the sign restrictions.

of structural parameters has positive prior probability. This implies that many structural parameters that satisfy the sign and zero restrictions are discarded by the prior density over the structural parameterization implied by the PFA. In the next section, we will use Beaudry, Nam and Wang's (2011) empirical application to illustrate the implications of the PFA relative to our importance sampler. As we will see, this example is insightful because it would allow us to analytically highlight how the PFA adds identification restrictions.

8   Application to Optimism Shocks

In this section, we illustrate our importance sampler by revisiting the application about optimism shocks previously analyzed in the literature by Beaudry, Nam and Wang (2011) using the PFA. 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 an 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. Using the PFA one could conclude that optimism shocks are associated with standard business cycle type phenomena because they generate a simultaneous boom in consumption and hours worked. But this conclusion relies on the fact that the identification does not solely come from the sign and zero restrictions because the PFA is used. Once we use our importance sampler, the described conclusion is harder to maintain. The benchmark SVAR in Beaudry, Nam and Wang (2011) considers five variables: TFP, stock prices, consumption, the real federal funds rate, and hours worked.[22]   In their less restrictive identification strategy, shown in Table 4, optimism shocks are identified as positively affecting stock prices and not affecting TFP at horizon zero. Appendix C gives details on their reduced-form prior density and the data.

In what follows we first compare the results obtained, for both the IRFs and the forecast error variance decomposition (FEVD), using the PFA and our importance sampler. Next, we parse how the implicit PFA prior density over the structural parameterization affects identification by unveiling zero restrictions not explicitly recognized as being part of the identification strategy. Finally, we conclude the empirical application by analyzing how using one- or two-sided derivatives affects the computation of the volume elements and the IRFs.

---

[22]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 seven-variable model.

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

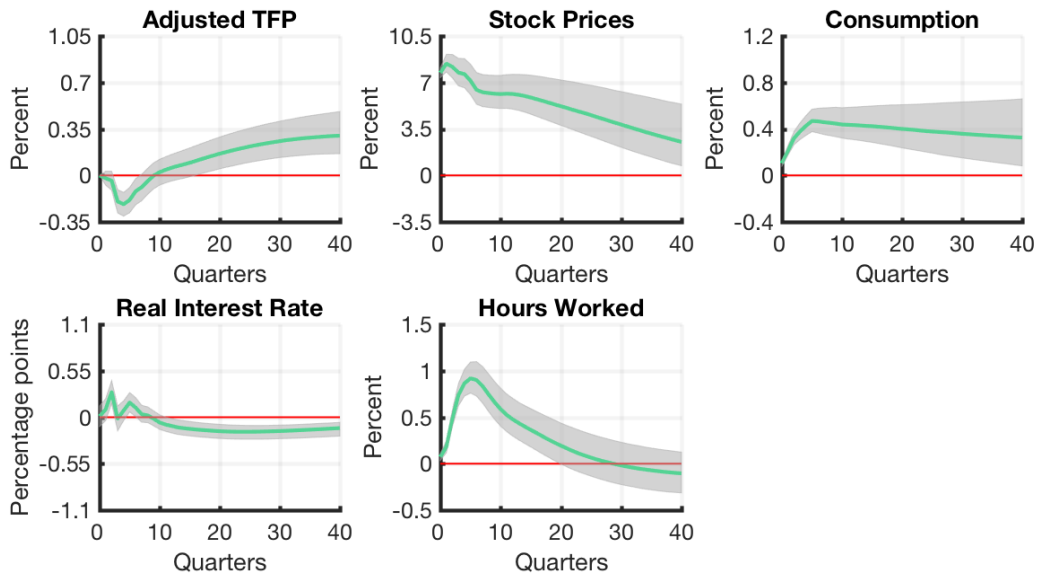Table 4: Restrictions on IRFs at horizon 0.

8.1 IRFs

Let's begin by comparing the IRFs obtained using the PFA with the ones obtained with our importance sampler. This comparison is helpful to illustrate how Beaudry, Nam and Wang (2011) results do not solely come from the sign and zero restrictions described in Table 4.

Panel (a) in Figure 2 shows the point-wise median as well as the 68 percent point-wise probability bands for the IRFs of TFP, stock prices, consumption, the federal funds rate, and hours worked using the PFA.[23] The key message from this panel is that optimism shocks generate a boom in consumption and hours worked. The point-wise probability bands 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.
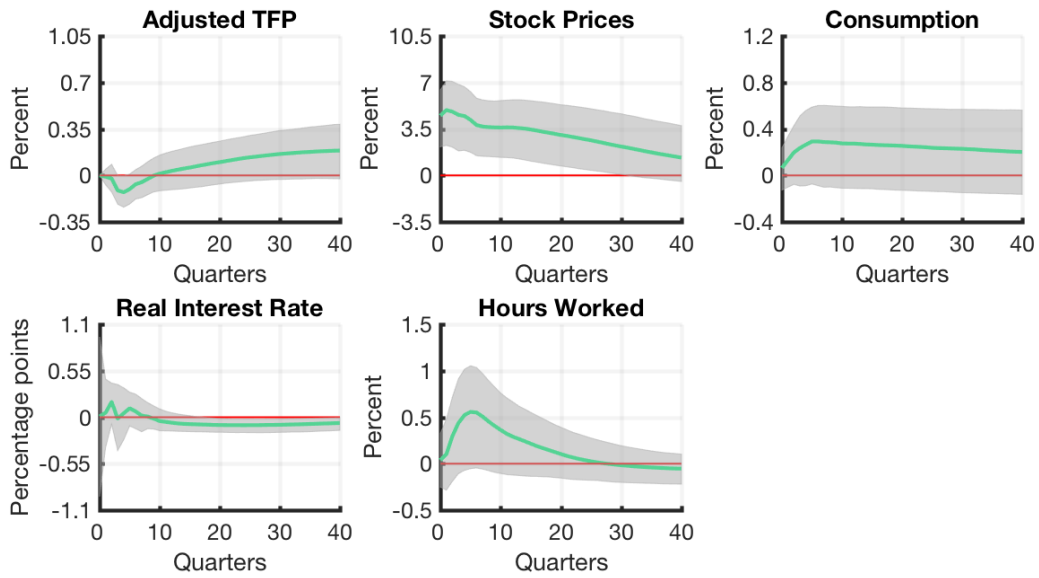
Indeed, these IRFs are highlighted by Beaudry, Nam and Wang (2011). If these IRFs were the result of only imposing the restrictions described in Table 4, the findings reported in Panel (a) of Figure 2 would strongly support the view that optimism shocks are relevant for business cycle fluctuations. The question is how much of the results reported in Panel (a) of Figure 2 are due to the restrictions described in Table 4 and how much are due to the restrictions that the PFA adds.

Panel (b) in Figure 2 shows that, once we use our importance sampler, the results highlighted by Beaudry, Nam and Wang (2011) disappear. There are three important differences with the results reported when working with the PFA. First, the PFA chooses a very large median response of stock prices in order to minimize the loss function used to impose the sign restriction on stock prices. Second, the point-wise median IRFs for consumption and hours worked are closer to zero when we use our importance sampler. Third, the point-wise probability bands associated with our importance sampler are much larger than the ones obtained with the PFA. To highlight that the PFA not only implies artificially narrow point-wise probability bands but also distorts the point-wise median IRFs, Figure 3 compares the IRFs obtained with our importance sampler with the median IRFs computed with the PFA. As the reader can see, the PFA boosts the effects of optimism shocks on stock prices,

---

[23]There are alternative methods of summarizing the outcome of set identified SVARs models; see Inoue and Kilian (2013). We report 68 percent point-wise probability bands in order to facilitate the comparison with the results reported by Beaudry, Nam and Wang (2011).

(a) PFA

(b) Importance Sampler

Figure 2: IRFs to a one standard deviation optimism shock. PFA and importance sampler comparison. The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent point-wise probability bands. The figure is based on 10,000 independent draws obtained using the PFA and Algorithm 3, respectively.
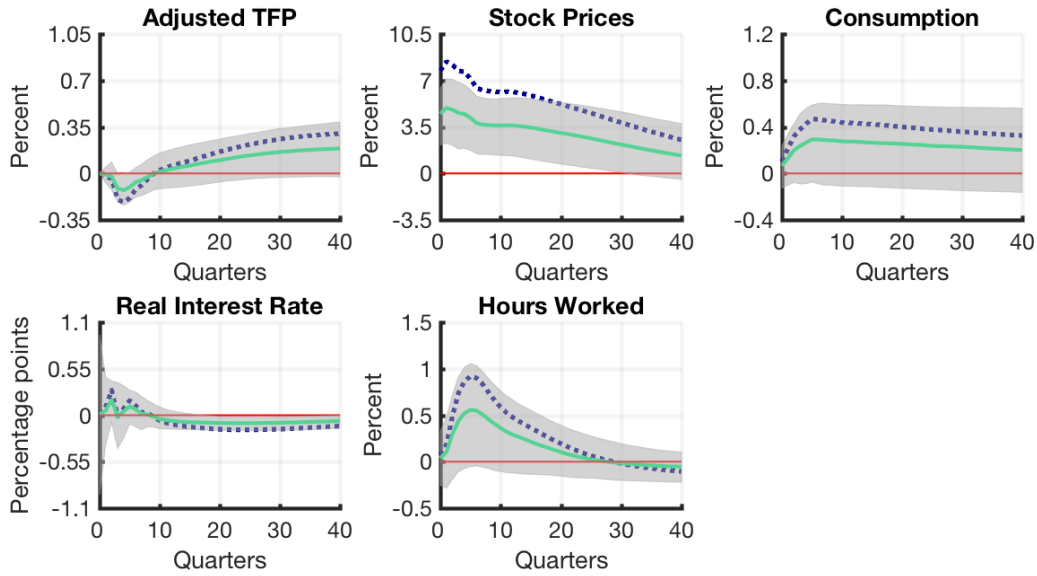
consumption, and hours.

Figure 3: IRFs to a one standard deviation optimism shock. The dotted curves depict the point-wise posterior medians when using the PFA. The solid curves and the shaded areas represent the point-wise posterior medians and 68 percent point-wise probability bands, respectively, when using our importance sampler. The figure is based on 10,000 independent draws obtained using the PFA and Algorithm 3, respectively.

	PFA	**Importance Sampler**
Adjusted TFP	0.17 [0.08,0.30]	0.10 [0.03,0.25]
Stock Prices	0.73 [0.57,0.85]	0.26 [0.06,0.58]
Consumption	0.27 [0.14,0.43]	0.16 [0.03,0.49]
Real Interest Rate	0.14 [0.07,0.22]	0.19 [0.08,0.38]
Hours Worked	0.32 [0.21,0.44]	0.17 [0.05,0.47]

Table 5: Share of FEVD attributable to optimism shocks at horizon 40. The column PFA (Importance Sampler) reports for each variable the posterior median and the 68 percent probability intervals of the FEVD attributable to optimism shocks at horizon 40 when using the PFA (Importance Sampler). The table is based on 10,000 independent draws obtained using the PFA and Algorithm 3, respectively.

8.2 Forecast Error Variance Decomposition

Table 5 compares the contribution of optimism shocks to the FEVD obtained under the PFA and our importance sampler. For ease of exposition, we only focus on the contributions to the FEVD at

horizon 40. Using the PFA, the median contribution of optimism shocks to the FEVD of consumption and hours worked is 27 and 32 percent, respectively. In contrast, using our importance sampler, the median contributions are 16 and 17 percent, respectively. Table 5 also reports the 68 percent point-wise probability intervals in brackets. Similarly to the case with the IRFs, using our importance sampler results in smaller contributions of optimism shocks to the FEVD of most variables and probability intervals for the FEVD that are much wider than the ones obtained using the PFA.

8.3   Additional Restrictions

As mentioned, the PFA introduces additional restrictions. For the application studied here, we now analytically characterize the optimal orthogonal matrix that the PFA chooses. This will allow us to analytically show how the PFA incorporates additional identifying restrictions (zero restrictions in this case) to the sign and the zero restrictions described in Table 4. Because $h(\boldsymbol{\Sigma})$ is the Cholesky decomposition with $h(\boldsymbol{\Sigma})$ upper triangular and positive diagonal, direct computations show that, for any value of the reduced-form parameters, the optimal orthogonal matrix that the PFA chooses has the form $\mathbf{Q}^* = [\mathbf{q}_1^* \ \dots \ \mathbf{q}_5]$ where the first column is equal to $\mathbf{q}_1^* = [0 \ 1 \ 0 \ 0 \ 0]'$.[24] For any value of the reduced-form parameters $(\mathbf{B}, \boldsymbol{\Sigma})$, let $(\mathbf{A}_0^*, \mathbf{A}_+^*)$ be the value of the structural parameters implied by $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}^*)$, i.e. $(\mathbf{A}_0^*, \mathbf{A}_+^*) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}^*)$. Since the IRFs at horizon zero 0 equal $\left((\mathbf{A}_0^*)^{-1}\right)' = h(\boldsymbol{\Sigma})'\mathbf{Q}^*$, this value of the structural parameters satisfies both the sign and the zero restrictions described in Table 4.

Is $(\mathbf{A}_0^*, \mathbf{A}_+^*)$ the only value of the structural parameters that satisfies the sign and zero restrictions? The answer is no. To see this, consider now the set

$$\mathbb{O}(5) = \left\{ [\mathbf{q}_1 \ \dots \ \mathbf{q}_5] \in \mathcal{O}(5) \mid \mathbf{q}_1 = [0 \ q_{2,1} \ q_{3,1} \ q_{4,1} \ q_{5,1}]' \text{ and } q_{2,1} > 0 \right\}.$$

For any value of the reduced-form parameters $(\mathbf{B}, \boldsymbol{\Sigma})$ and $\mathbf{Q} \in \mathbb{O}(5)$, let $(\mathbf{A}_0, \mathbf{A}_+)$ be the value of the structural parameters implied by $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, i.e., $(\mathbf{A}_0, \mathbf{A}_+) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. This value of the structural parameters also satisfies both the sign and the zero restrictions described in Table 4. Clearly, the PFA disregards most of the elements of $\mathbb{O}(5)$. What are the consequences of using the PFA? Since $(\mathbf{A}_0^*, \mathbf{A}_+^*) = f_h^{-1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}^*)$, the first column of $\mathbf{A}_0^* = h(\boldsymbol{\Sigma})^{-1}\mathbf{Q}^*$ is of the form $[t_{1,2} \ t_{2,2} \ 0 \ 0 \ 0]'$, where $t_{i,j}$ is the (i,j) entry of $h(\boldsymbol{\Sigma})^{-1}$ for any value of the reduced-form parameters. Hence, the PFA imposes

---

[24]Similar computations could be done for other choices of $h(\boldsymbol{\Sigma})$, but they would not be as direct. In fact, we use this decomposition because it is the one that Mountford and Uhlig (2009) use when describing the PFA.

the condition that the first column of the matrix of contemporaneous structural parameters has three zeros in the positions just described. This clearly means that the PFA introduces zero restrictions on the structural parameters that are not explicitly acknowledged in the identification strategy described in Table 4.

8.4 One-Sided Versus Two-Sided Derivatives

In Section 6.2, we showed several cases where using one-sided derivatives decreases computing time considerably. We now show that, at least for our application, this can be done without compromising the numerical accuracy of either the volume elements or the IRFs.
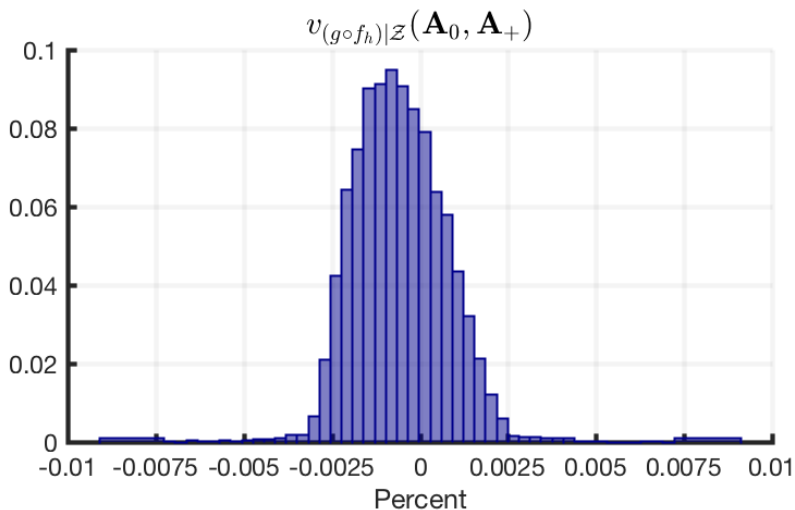

Figure 4: Volume elements comparison. The histogram shows the percent difference between the volume element $v_{(g \circ f_h) | \mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)$ computed using one-sided relative to two-sided derivatives. The histogram includes 99 percent of the support of the distribution and it is based on 10,000 independent draws obtained using Algorithm 3.

Consider Figure 4. This figure shows the histogram of the percentage difference between the volume element associated with the mapping $v_{(g \circ f_h) | \mathcal{Z}}(\mathbf{A}_0, \mathbf{A}_+)$, when computed using one-sided derivatives relative to when the same volume element is computed using two-sided derivatives. The percentage difference is expressed in terms of the volume elements computed using two-sided derivatives. The figure shows that most of the percent differences are smaller than 0.1. These results suggest that it is unlikely to find cases in which the one- and two-sided derivatives substantially differ. Next, we show that the percent differences in the volume element reported above do not affect the IRFs. Figure 5 plots the percentage point difference between the median and the bounds of the probability bands of the IRFs obtained using one-sided derivatives and the ones obtained using two-sided derivatives. In all
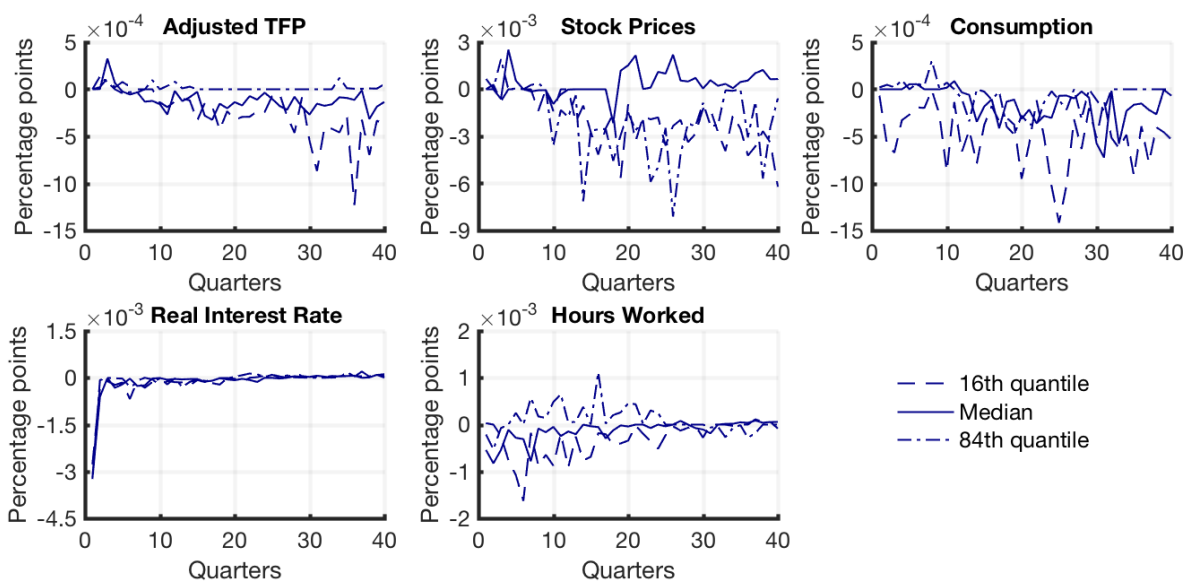
33

Figure 5: IRFs comparison. The solid lines show the percentage point difference between the point-wise median IRFs computed using one-sided relative to two-sided derivatives. The dashed lines and the dashed-dotted lines show the percentage point difference in point-wise 16th and 84th quantile IRFs computed using one-sided relative to two-sided derivatives, respectively. The figure is based on 10,000 independent draws obtained using Algorithm 3.

the cases, the percentage point difference is smaller than 0.1, and the conclusions a researcher would obtain are unchanged.

9   Conclusion

We developed efficient algorithms for Bayesian inference based on SVARs identified with sign and zero restrictions. Critically, our 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.

Appendix

A   Proofs

In this section of the appendix, we provide all the proofs mentioned in the main text. We first provide details and proofs, or references to proofs, for Section 3.4. Then, we prove Proposition 1. Finally, we prove the three claims made in Sections 5.1 and 5.2.

A.1   Integration on Manifolds and Change of Variables

In this section we fill in the details of Section 3.4. We follow the treatment in Spivak (1965). Theorem 1 of Section 3.4 is Theorem 3-13 of Spivak (1965), with two differences. First, in Theorem 3-13 there is an additional assumption that $\det(D\gamma(\mathbf{u})) \neq 0$, which can be relaxed using Sard's Theorem.[25] Second, in Spivak (1965) integration is defined in the spirit of Riemann as opposed to Lebesgue. Thus the notion of an integrable function is much more restrictive.[26] However, since Theorem 1 holds for all Riemann integrable functions, it also holds for all Lebesgue integrable functions. When we use the term integrable, we will mean Lebesgue integrable.

While Spivak (1965) develops all the necessary machinery to prove Theorems 2 and 3, they are not explicitly treated because the goal of that book is the generalization of Stokes' Theorem. The same is true of other texts on multivariate calculus. For this reason, we will very briefly review the necessary machinery and then prove Theorems 2 and 3 using Theorem 1.

Let $\mathcal{V}$ be a b-dimensional smooth manifold in $\mathbb{R}^a$. We can define a σ-algebra $\mathfrak{S}$ on $\mathcal{V}$ by $A \in \mathfrak{S}$ if and only if $\gamma^{-1}(A)$ is Lebesgue measurable for any coordinate system $\gamma : U \to \mathcal{V}$. Because there exists a countable basis of open balls in $\mathbb{R}^a$, for any $A \in \mathfrak{S}$ there exists a decomposition $(A_i, \gamma_i, U_i)_{i=1}^{\infty}$ such that $\gamma_i : U_i \to \mathcal{V}$ is a coordinate system with $\int_{U_i} |\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}} d\mathbf{u}_i < \infty$, $A_i \in \mathfrak{S}$ with $A_i \subset \gamma_i(U_i)$, $A_i \cap A_j = \varnothing$ if $i \neq j$, and $A = \bigcup_{i=1}^{\infty} A_i$. These decompositions can be used to define a measure μ on $\mathcal{V}$ by

$$\mu(A) = \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(A_i)} |\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}} d\mathbf{u}_i.$$

Since the decompositions are not unique, in order for μ to be well-defined, it must be independent of the decomposition of A, which the following proposition implies.

---

[25] See Theorem 3-14 of Spivak (1965) for a statement and proof of Sard's Theorem.

[26] See the third section of Chapter 3 in Spivak (1965) and Section 3 of Chapter 11 in Royden (1963) for a discussion of these two notions of an integrable function. We will refer to these as Riemann integrable and Lebesgue integrable.

Proposition 2. *Let $\mathcal{V} \subset \mathbb{R}^a$ be a b-dimensional smooth manifold, let $\gamma_1 : U_1 \to \mathcal{V}$ be a coordinate system in $\mathcal{V}$, let $U_2 \subset \mathbb{R}^b$ be open, and let $\gamma_2 : U_2 \to \mathcal{V}$ be one-to-one and continuously differentiable. If $A \subset \gamma(U_1) \cap \gamma(U_2)$, $\lambda : A \to \mathbb{R}$, and the function $\lambda(\gamma_1(\mathbf{u}_1))|\det(D\gamma_1(\mathbf{u}_1)'D\gamma_1(\mathbf{u}_1))|^{\frac{1}{2}}$ is integrable over $\gamma_1^{-1}(A)$, then*

$$\int_{\gamma_1^{-1}(A)} \lambda(\gamma_1(\mathbf{u}_1))|\det(D\gamma_1(\mathbf{u}_1)'D\gamma_1(\mathbf{u}_1))|^{\frac{1}{2}}d\mathbf{u}_1 = \int_{\gamma_2^{-1}(A)} \lambda(\gamma_2(\mathbf{u}_2))|\det(D\gamma_2(\mathbf{u}_2)'D\gamma_2(\mathbf{u}_2))|^{\frac{1}{2}}d\mathbf{u}_2.$$

Proof. It follows from the proof of Theorem 5-2 in Spivak (1965) that $\gamma_2^{-1}(\gamma_1(U_1))$ is an open set in $\mathbb{R}^b$ and $\gamma_1^{-1} \circ \gamma_2 : \gamma_2^{-1}(\gamma_1(U_1)) \to \mathbb{R}^b$ is continuously differentiable. Since $\gamma_1^{-1} \circ \gamma_2$ is one-to one, it follows from Theorem 1 and the chain rule that

$$\int_{\gamma_1^{-1}(A)} \lambda(\gamma_1(\mathbf{u}_1))|\det(D\gamma_1(\mathbf{u}_1)'D\gamma_1(\mathbf{u}_1))|^{\frac{1}{2}}d\mathbf{u}_1$$

$$= \int_{\gamma_2^{-1}(A)} \lambda(\gamma_1(\gamma_1^{-1} \circ \gamma_2(\mathbf{u_2})))|\det(D\gamma_1(\gamma_1^{-1} \circ \gamma_2(\mathbf{u}_2))'D\gamma_1(\gamma_1^{-1} \circ \gamma_2(\mathbf{u}_2)))|^{\frac{1}{2}}|\det(D(\gamma_1^{-1} \circ \gamma_2)(\mathbf{u}_2))|d\mathbf{u}_2$$

$$= \int_{\gamma_2^{-1}(A)} \lambda(\gamma_2(\mathbf{u_2}))|\det(D(\gamma_1^{-1} \circ \gamma_2)(\mathbf{u}_2)'D\gamma_1(\gamma_1^{-1} \circ \gamma_2(\mathbf{u}_2))'D\gamma_1(\gamma_1^{-1} \circ \gamma_2(\mathbf{u}_2))D(\gamma_1^{-1} \circ \gamma_2)(\mathbf{u}_2))|^{\frac{1}{2}}d\mathbf{u}_2$$

$$= \int_{\gamma_2^{-1}(A)} \lambda(\gamma_2(\mathbf{u_2}))|\det(D(\gamma_1 \circ \gamma_1^{-1} \circ \gamma_2)(\mathbf{u}_2)'D(\gamma_1 \circ \gamma_1^{-1} \circ \gamma_2)(\mathbf{u}_2))|^{\frac{1}{2}}d\mathbf{u}_2$$

$$= \int_{\gamma_2^{-1}(A)} \lambda(\gamma_2(\mathbf{u_2}))|\det(D\gamma_2(\mathbf{u}_2)'D\gamma_2(\mathbf{u}_2))|^{\frac{1}{2}}d\mathbf{u}_2$$

$\square$

So, if $(A_i, \gamma_i, U_i)_{i=1}^{\infty}$ and $(\tilde{A}_j, \tilde{\gamma}_j, \tilde{U}_j)_{j=1}^{\infty}$ are decompositions of $A \in \mathfrak{S}$, then applying Proposition 2, with $\lambda(\mathbf{v}) = 1$ for every $\mathbf{v} \in A$, gives

$$\sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(A_i)} |\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \int_{\gamma_i^{-1}(A_i \cap \tilde{A}_j)} |\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \int_{\tilde{\gamma}_j^{-1}(A_i \cap \tilde{A}_j)} |\det(D\tilde{\gamma}_j(\tilde{\mathbf{u}}_j)'D\tilde{\gamma}_j(\tilde{\mathbf{u}}_j))|^{\frac{1}{2}}d\tilde{\mathbf{u}}_j = \sum_{j=1}^{\infty} \int_{\tilde{\gamma}_j^{-1}(\tilde{A}_j)} |\det(D\tilde{\gamma}_j(\tilde{\mathbf{u}}_j)'D\tilde{\gamma}_j(\tilde{\mathbf{u}}_j))|^{\frac{1}{2}}d\tilde{\mathbf{u}}_j.$$

So the measure is independent of the decomposition.

This measure is called the volume measure on $\mathcal{V}$. Comparing this definition of "volume" for the case that b is one or two with the formulas for arc length or surface area from elementary calculus, one sees that they are identical. More generally, $|\det(D\gamma(\mathbf{u})'D\gamma(\mathbf{u}))|^{\frac{1}{2}}$ is equal to the volume of the

parallelepiped spanned by the columns of $D\gamma(\mathbf{u})$, which justifies calling this measure volume.[27]

This measure on the smooth manifold $\mathcal{V}$ can be used to define integration on $\mathcal{V}$ and if $A \in \mathfrak{G}$, $\lambda : A \to \mathbb{R}$ is integrable, and (A_i, γ_i, U_i) is a decomposition of A, then

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \sum_{i=1}^{\infty} \int_{A_i} \lambda(\mathbf{v})d\mathbf{v} = \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(A_i)} \lambda(\gamma_i(\mathbf{u}_i))|\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i.$$

Here, $d\mathbf{v}$ denotes integration with respect to the volume measure, while $d\mathbf{u}_i$ is integration with respect to Lebesgue measure.

Theorem 2 also follows directly from Proposition 2. To see this, let U, γ, A, and λ be as in Theorem 2 and let $(A_i, \gamma_i, U_i)_{i=1}^{\infty}$ be a decomposition of A. By the definition of integration with respect to the volume measure and Proposition 2, with λ_i playing the role of λ_1 and λ playing the role of λ_2,

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(A_i)} \lambda(\gamma_i(\mathbf{u}_i))|\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i$$

$$= \sum_{i=1}^{\infty} \int_{\gamma^{-1}(A_i)} \lambda(\gamma(\mathbf{u}))|\det(D\gamma(\mathbf{u})'D\gamma(\mathbf{u}))|^{\frac{1}{2}}d\mathbf{u} = \int_{\gamma^{-1}(A)} \lambda(\gamma(\mathbf{u}))|\det(D\gamma(\mathbf{u})'D\gamma(\mathbf{u}))|^{\frac{1}{2}}d\mathbf{u}.$$

We now prove Theorem 3. Let U, γ, β, $\mathcal{U}$, $\mathcal{V}$, A, and λ be as in Theorem 3. First, Theorem 5-1 in Spivak (1965) guarantees that $\mathcal{U}$ is a d-dimensional smooth manifold in $\mathbb{R}^b$. Let $(A_i, \gamma_i, U_i)_{i=1}^{\infty}$ be a decomposition of A. Using Theorem 2 and the chain rule, we have

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \sum_{i=1}^{\infty} \int_{A_i} \lambda(\mathbf{v})d\mathbf{v}$$

$$= \sum_{i=1}^{\infty} \int_{(\gamma \circ \gamma_i)^{-1}(A_i)} \lambda(\gamma \circ \gamma_i(\mathbf{u}_i))|\det(D(\gamma \circ \gamma_i)(\mathbf{u}_i)'D(\gamma \circ \gamma_i)(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i$$

$$= \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(\gamma^{-1}(A_i) \cap U_i)} \lambda(\gamma(\gamma_i(\mathbf{u}_i)))|\det(D\gamma_i(\mathbf{u}_i)'D\gamma(\gamma_i(\mathbf{u}_i))'D\gamma(\gamma_i(\mathbf{u}_i))D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i.$$

The last integral is over $\gamma_i^{-1}(\gamma^{-1}(A_i) \cap U_i)$ because γ is only guaranteed to be one-to-one over $\beta^{-1}(\{\mathbf{0}\})$ and so $\gamma^{-1}(A_i)$ may not be contained in $\gamma_i(U_i)$. Because $\beta(\gamma_i(\mathbf{u}_i)) = \mathbf{0}$ for every $\mathbf{u}_i \in U_i$, $D\beta(\gamma_i(\mathbf{u}_i))D\gamma_i(\mathbf{u}_i) = D(\beta \circ \gamma_i)(\mathbf{u}_i) = \mathbf{0}$ for every $\mathbf{u}_i \in U_i$. So, if the columns of the $b \times d$ matrix $\mathbf{N}_{\gamma_i(\mathbf{u}_i)}$ are an orthonormal basis for the null space of $D\beta(\gamma_i(\mathbf{u}_i))$, then $D\gamma_i(\mathbf{u}_i) = \mathbf{N}_{\gamma_i(\mathbf{u}_i)}\mathbf{X}$ for some $d \times d$

---

[27]See Chapter 5 of Munkres (1991).

matrix $\mathbf{X}$. Furthermore, $D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i) = \mathbf{X}'\mathbf{X}$. Again, applying Theorem 2 we have that

$$\int_A \lambda(\mathbf{v})d\mathbf{v} = \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(\gamma^{-1}(A_i) \cap U_i)} \lambda(\gamma(\gamma_i(\mathbf{u}_i)))|\det(D\gamma_i(\mathbf{u}_i)'D\gamma(\gamma_i(\mathbf{u}_i))'D\gamma(\gamma_i(\mathbf{u}_i))D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i$$

$$= \sum_{i=1}^{\infty} \int_{\gamma_i^{-1}(\gamma^{-1}(A_i) \cap U_i)} \lambda(\gamma(\gamma_i(\mathbf{u}_i)))|\det(\mathbf{N}'_{\gamma_i(\mathbf{u}_i)}D\gamma(\gamma_i(\mathbf{u}_i))'D\gamma(\gamma_i(\mathbf{u}_i))\mathbf{N}_{\gamma_i(\mathbf{u}_i)})|^{\frac{1}{2}}|\det(D\gamma_i(\mathbf{u}_i)'D\gamma_i(\mathbf{u}_i))|^{\frac{1}{2}}d\mathbf{u}_i$$

$$= \sum_{i=1}^{\infty} \int_{\gamma^{-1}A_i \cap U_i} \lambda(\gamma(\mathbf{u}))|\det(\mathbf{N}'_{\mathbf{u}}D\gamma(\mathbf{u})'D\gamma(\mathbf{u})\mathbf{N}_{\mathbf{u}})|^{\frac{1}{2}}d\mathbf{u} = \int_{\gamma^{-1}(A) \cap \mathcal{U}} \lambda(\gamma(\mathbf{u}))|\det(\mathbf{N}'_{\mathbf{u}}D\gamma(\mathbf{u})'D\gamma(\mathbf{u})\mathbf{N}_{\mathbf{u}})|^{\frac{1}{2}}d\mathbf{u}.$$

This completes the proof of Theorem 3.

A.2 Proof of Proposition 1

We will compute v_{f_h} indirectly using f_h^{-1} and Theorem 3. The function f_h^{-1} is not defined over an open set, so to apply Theorem 3 we must extend the definition of f_h^{-1}. Let $U \subset \mathbb{R}^{n(m+2n)}$ be the set of all $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ such that $\mathbf{B}$ is $m \times n$, $\mathbf{Q}$ is $n \times n$, and $\boldsymbol{\Sigma}$ is $n \times n$ and positive definite.[28] Define $g_h : U \to \mathbb{R}^{n(m+n)}$ by $g_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = (\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q}, \mathbf{B}\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q})$, where $\tilde{h}(\boldsymbol{\Sigma}) = h\left(\frac{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'}{2}\right)$. When $\boldsymbol{\Sigma}$ is restricted to be symmetric and $\mathbf{Q}$ is restricted to be orthogonal, g_h is equal to f_h^{-1}. Let $\beta : U \to \mathbb{R}^{n^2}$ be any differentiable function such that its derivative is of full row rank and $\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \mathbf{0}$ if and only if $\boldsymbol{\Sigma}$ is symmetric and $\mathbf{Q}$ is orthogonal. Let $\mathcal{U} = \beta^{-1}(\{\mathbf{0}\})$. Theorem 3 implies that $v_{g_h|_\mathcal{U}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = |\det(\mathbf{N}'_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})'Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})})|^{\frac{1}{2}}$, where the columns of $\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$ form an orthonormal basis for the null space of $D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$.[29] In this case, because $Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$ is square, $v_{g_h|_\mathcal{U}}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ will be equal to $|\det(Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})})|$. Finally, because f_h and $g_h|_\mathcal{U}$ are inverses, $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)$ will be equal to the reciprocal of $v_{g_h|_\mathcal{U}}(f_h(\mathbf{A}_0, \mathbf{A}_+))$. In what follows, we will first compute $Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, then define $\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, compute $D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$, and then find $\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$. The final, and most involved, step then be to compute $|\det(Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})})|$ to obtain $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+)$.

The derivative of g_h can be computed using formulas for the chain rule, the derivative of a matrix product, and the derivative of a matrix inverse, all of which are given in Appendix A.13 of Lütkephol

---

[28]A matrix $\boldsymbol{\Sigma}$ is positive definite if and only if $\min_{\|\mathbf{x}\|=1} \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x} > 0$. The set of all positive definite matrices is open in the set of all $n \times n$ matrices and if $\boldsymbol{\Sigma}$ is positive definite, then so is $\frac{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'}{2}$.

[29]Derivatives are defined for vector-valued functions of a vector argument. The function g_h can be written in this form as $g_h(\text{vec}(\mathbf{B}), \text{vec}(\boldsymbol{\Sigma}), \text{vec}(\mathbf{Q})) = (\text{vec}(\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q}), \text{vec}(\mathbf{B}\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q}))$, where vec() denotes the operator that stacks the columns of a matrix into a vector. Similar representations are available for the functions β and $\tilde{h}$. When taking derivatives this representation will be implicit.

(1993). A direct computation gives

$$
Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})
$$

$$
= \begin{bmatrix} \mathbf{I}_{n^2} & \mathbf{0}_{n^2,nm} \\ \mathbf{I}_n \otimes \mathbf{B} & \mathbf{I}_{nm} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n^2,nm} & -(\mathbf{Q}' \otimes \mathbf{I}_n)(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma}))^{-1} D\tilde{h}(\boldsymbol{\Sigma}) & \mathbf{I}_n \otimes \tilde{h}(\boldsymbol{\Sigma})^{-1} \\ (\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q})' \otimes \mathbf{I}_m & \mathbf{0}_{nm,n^2} & \mathbf{0}_{nm,n^2} \end{bmatrix}.
$$

To define β, let $\mathbf{L}_n$ denote the elimination matrix and let $\mathbf{D}_n$ denote the duplication matrix.[30] A $n \times n$ matrix $\mathbf{Q}$ is orthogonal if and only if $\mathbf{L}_n \operatorname{vec}(\mathbf{Q}\mathbf{Q}' - \mathbf{I}_n) = \mathbf{0}$. The columns of $\mathbf{D}_n$ are perpendicular and form a basis for the set of all vectorized symmetric matrices. The norm of $\frac{n(n-1)}{2}$ of these columns is $\sqrt{2}$, while the remaining n columns are of norm 1. Thus there is a diagonal matrix $\boldsymbol{\Delta}$ such that the columns of $\mathbf{N}_S = \mathbf{D}_n \boldsymbol{\Delta}$ form an orthonormal basis for the set of vectorized symmetric matrices and the determinant of $\boldsymbol{\Delta}$ is equal to $2^{-\frac{n(n-1)}{4}}$. Let $\mathbf{N}_A$ be any $n^2 \times \frac{n(n-1)}{2}$ matrix such that $[\mathbf{N}_S\ \mathbf{N}_A]$ is orthogonal. A matrix $\boldsymbol{\Sigma}$ is symmetric if and only if $\mathbf{N}_A' \operatorname{vec}(\boldsymbol{\Sigma}) = \mathbf{0}$. We define $\beta : U \to \mathbb{R}^{n^2}$ by $\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = (\mathbf{N}_A' \operatorname{vec}(\boldsymbol{\Sigma}), \mathbf{L}_n \operatorname{vec}(\mathbf{Q}\mathbf{Q}' - \mathbf{I}_n))$.

To compute the derivative of β, we will also need the commutation matrix, which we will denote by $\mathbf{K}_{n,n}$.[31] It follows from the definition of $\mathbf{K}_{n,n}$ that if f is a differentiable function whose range is a subset of the $n \times n$ matrices, then $D(f') = \mathbf{K}_{n,n} Df$. It is also the case that $(\mathbf{X} \otimes \mathbf{Y})\mathbf{K}_{n,n} = \mathbf{K}_{n,n}(\mathbf{Y} \otimes \mathbf{X})$ for $n \times n$ matrices $\mathbf{X}$ and $\mathbf{Y}$.[32] A direct computation gives that the derivative of β is

$$
D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{\frac{n(n-1)}{2},nm} & \mathbf{N}_A' & \mathbf{0}_{\frac{n(n-1)}{2},n^2} \\ \mathbf{0}_{\frac{n(n+1)}{2},nm} & \mathbf{0}_{\frac{n(n+1)}{2},n^2} & \mathbf{L}_n(\mathbf{I}_n + \mathbf{K}_{n,n})(\mathbf{Q} \otimes \mathbf{I}_n) \end{bmatrix}.
$$

Define $\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$ by

$$
\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})} = \begin{bmatrix} \mathbf{I}_{nm} & \mathbf{0}_{nm,\frac{n(n+1)}{2}} & \mathbf{0}_{nm,\frac{n(n-1)}{2}} \\ \mathbf{0}_{n^2,nm} & \mathbf{N}_S & \mathbf{0}_{n^2,\frac{n(n-1)}{2}} \\ \mathbf{0}_{n^2,nm} & \mathbf{0}_{n^2,\frac{n(n+1)}{2}} & (\mathbf{Q}' \otimes \mathbf{I}_n)\mathbf{N}_A \end{bmatrix}.
$$

Because $[\mathbf{N}_S\ \mathbf{N}_A]$ is orthogonal, the columns of $\mathbf{N}_A$ form an orthonormal basis for the set of all vector-

---

[30]$\mathbf{L}_n$ is the unique $\frac{n(n+1)}{2} \times n^2$ such that $\mathbf{L}_n \operatorname{vec}(\mathbf{X}) = \operatorname{vech}(\mathbf{X})$ for every $n \times n$ matrix $\mathbf{X}$ and $\mathbf{D}_n$ is the unique $n^2 \times \frac{n(n+1)}{2}$ matrix such that $\mathbf{D}_n \operatorname{vech}(\mathbf{X}) = \operatorname{vec}(\mathbf{X})$ for every $n \times n$ symmetric matrix $\mathbf{X}$, where $\operatorname{vech}()$ denotes the operator that stacks the elements of a square matrix on or below the diagonal into a vector.

[31]$\mathbf{K}_{n,n}$ is the unique $n^2 \times n^2$ matrix such that $\mathbf{K}_{n,n} \operatorname{vec}(\mathbf{X}) = \operatorname{vec}(\mathbf{X}')$ for every $n \times n$ matrix $\mathbf{X}$.

[32]See Appendix A.12.2 of Lütkephol (1993).

ized $n \times n$ anti-symmetric matrices.[33] This implies that the columns of $\mathbf{N}_A$ also form an orthonormal basis for the null space of $(\mathbf{I}_n + \mathbf{K}_{n,n})$. If $\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \mathbf{0}$, so that $\mathbf{Q}$ is orthogonal, then it is easy to directly verify that $D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})} = \mathbf{0}$ and $\mathbf{N}'_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})} = \mathbf{I}_{n(m+n)}$. Thus the columns of $\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$ are orthonormal and in the null space of $D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$. So, by a dimension argument, the columns of $\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})}$ must form an orthonormal basis for the null space of $D\beta(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$.

A direct computation gives that $|\det(Dg_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})\mathbf{N}_{(\mathbf{B},\boldsymbol{\Sigma},\mathbf{Q})})|$ is equal to

$$|\det((\tilde{h}(\boldsymbol{\Sigma})^{-1}\mathbf{Q})' \otimes \mathbf{I}_m)| \cdot |\det(\mathbf{Q}' \otimes \tilde{h}(\boldsymbol{\Sigma})^{-1})| \cdot \left|\det\left(\left[-(\tilde{h}(\boldsymbol{\Sigma})' \otimes \mathbf{I}_n)^{-1}D\tilde{h}(\boldsymbol{\Sigma})\mathbf{N}_S \quad \mathbf{N}_A\right]\right)\right|. \quad (12)$$

When $\boldsymbol{\Sigma}$ is symmetric, the first two terms are equal to $|\det(\boldsymbol{\Sigma})|^{-\frac{m}{2}}$ and $|\det(\boldsymbol{\Sigma})|^{-\frac{n}{2}}$, respectively.[34] Since $\tilde{h}(\boldsymbol{\Sigma})'\tilde{h}(\boldsymbol{\Sigma}) = \frac{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'}{2}$, it follows that $(\mathbf{I}_{n^2} + \mathbf{K}_{n,n})(\mathbf{I}_n \otimes \tilde{h}(\boldsymbol{\Sigma})')D\tilde{h}(\boldsymbol{\Sigma}) = \frac{1}{2}(\mathbf{I}_{n^2} + \mathbf{K}_{n,n})$. Multiplying by $\mathbf{N}_S$ and rearranging gives $(\mathbf{I}_{n^2} + \mathbf{K}_{n,n})((\mathbf{I}_n \otimes \tilde{h}(\boldsymbol{\Sigma})')D\tilde{h}(\boldsymbol{\Sigma})\mathbf{N}_S - \frac{1}{2}\mathbf{N}_S) = \mathbf{0}$. Because the columns of $\mathbf{N}_A$ form an orthonormal basis for the null space of $(\mathbf{I}_n + \mathbf{K}_{n,n})$, there exists a $\frac{n(n-1)}{2} \times \frac{n(n+1)}{2}$ matrix $\mathbf{X}$ such that $(\mathbf{I}_n \otimes \tilde{h}(\boldsymbol{\Sigma})')D\tilde{h}(\boldsymbol{\Sigma})\mathbf{N}_S = \frac{1}{2}\mathbf{N}_S + \mathbf{N}_A\mathbf{X}$. So, the third term in Expression (12) is equal to $\left|\det\left(\left[-(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}(\frac{1}{2}\mathbf{N}_S + \mathbf{N}_A\mathbf{X}) \quad \mathbf{N}_A\right]\right)\right|$. Similarly, since $(\mathbf{I}_{n^2}+\mathbf{K}_{n,n})(\tilde{h}(\boldsymbol{\Sigma})'\otimes\tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_A = (\tilde{h}(\boldsymbol{\Sigma})'\otimes\tilde{h}(\boldsymbol{\Sigma})')^{-1}(\mathbf{I}_{n^2}+\mathbf{K}_{n,n})\mathbf{N}_A = \mathbf{0}$, it follows that there exists an $\frac{n(n-1)}{2} \times \frac{n(n+1)}{2}$ matrix $\mathbf{Z}$ such that $(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_A\mathbf{X} = \mathbf{N}_A\mathbf{Z}$. This implies that the third term in Expression (12) is also equal to

$$\left|\det\left(\left[-\frac{1}{2}(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_S \quad \mathbf{N}_A\right]\right)\right|$$

$$= \left|\det\left(\begin{bmatrix} \frac{1}{4}\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma}) \otimes \tilde{h}(\boldsymbol{\Sigma}))^{-1}(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_S & -\frac{1}{2}\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma}) \otimes \tilde{h}(\boldsymbol{\Sigma}))^{-1}\mathbf{N}_A \\ -\frac{1}{2}\mathbf{N}'_A(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_S & \mathbf{I}_{\frac{n(n-1)}{2}} \end{bmatrix}\right)\right|^{\frac{1}{2}}, \quad (13)$$

where the equality follows by multiplying by the transpose on the left. The formula for the determinant of a block matrix (see Appendix A.10 of Lütkephol (1993)) implies that Equation (13) is equal to

$$\left|\det\left(\frac{1}{4}\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma}) \otimes \tilde{h}(\boldsymbol{\Sigma}))^{-1}(\mathbf{I}_{n^2} - \mathbf{N}_A\mathbf{N}'_A)(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_S\right)\right|^{\frac{1}{2}}$$

$$= \left|\det\left(\frac{1}{4}\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma}) \otimes \tilde{h}(\boldsymbol{\Sigma}))^{-1}\mathbf{N}_S\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma})' \otimes \tilde{h}(\boldsymbol{\Sigma})')^{-1}\mathbf{N}_S\right)\right|^{\frac{1}{2}}$$

$$= \left|\det\left(\frac{1}{2}\mathbf{N}'_S(\tilde{h}(\boldsymbol{\Sigma})^{-1} \otimes \tilde{h}(\boldsymbol{\Sigma})^{-1})\mathbf{N}_S\right)\right| \quad (14)$$

---

[33]A square matrix $\mathbf{X}$ is anti-symmetric if and only if $\mathbf{X}' = -\mathbf{X}$.

[34]This follows from the fact that $|\det(\mathbf{X} \otimes \mathbf{Y})| = |\det(\mathbf{X})|^n|\det(\mathbf{Y})|^m$ for $m \times m$ matrices $\mathbf{X}$ and $n \times n$ matrices $\mathbf{Y}$, see Appendix A.11 of Lütkephol (1993).

Since $\mathbf{N}_S = \mathbf{D}_n\mathbf{\Delta}$, Equation (16) of Appendix A.12.2 of Lütkephol (1993) implies that Equation (14) is equal to $2^{-\frac{n(n+1)}{2}}|\det(\mathbf{\Sigma})|^{-\frac{n+1}{2}}$. Putting this all together, $v_{g_h|\mathcal{U}}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) = 2^{-\frac{n(n+1)}{2}}|\det(\mathbf{\Sigma})|^{-\frac{2n+m+1}{2}}$ and $v_{f_h}(\mathbf{A}_0, \mathbf{A}_+) = 2^{\frac{n(n+1)}{2}}|\det(\mathbf{A}_0)|^{-(2n+m+1)}$.

A.3 Proofs of Claims in Sections 5.1 and 5.2

In Sections 5.1 and 5.2 the following claims were made, which we now prove.

1. The derivative of $\beta(\mathbf{A}_0, \mathbf{A}_+)$ has full row rank for all structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$.

2. If $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$, with $\mathbf{Q} = (\mathbf{q}_1, \cdots, \mathbf{q}_n)$, are orthogonal reduced-form parameters such that $f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ satisfies the zero restrictions, then $\mathbf{M}_j(\mathbf{B}, \mathbf{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ is of full row rank for $1 \le j \le n$.

3. There exists an open set $V \subset \mathbb{R}^{nm+n^2+n^2}$ such that the functions $\mathbf{K}_j(\mathbf{B}, \mathbf{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ can be defined so that they are differentiable for all $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q}) \in V$, where $\mathbf{Q} = (\mathbf{q}_1, \cdots, \mathbf{q}_n)$. Also, any structural parameters satisfying the zero restrictions is almost surely in the set $U = f_h^{-1}(V)$.

To see that (1) holds, note that $D\beta(\mathbf{A}_0, \mathbf{A}_+) = \text{diag}_{1 \le j \le n}(\mathbf{Z}_j)D\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$.[35] The regularity condition on $\mathbf{F}$ implies that $D\mathbf{F}(\mathbf{A}_0, \mathbf{A}_+)$ is of full row rank and it was assumed that $\mathbf{Z}_j$ was of full row rank for $1 \le j \le n$. This implies that $D\beta(\mathbf{A}_0, \mathbf{A}_+)$ is of full row rank. To see that (2) holds, first note that it follows from (1) that the set of all structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$ that satisfy $\beta(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{0}$ will be an $(n(m+n) - \sum_{j=1}^n z_j)$-dimensional smooth manifold in $\mathbb{R}^{n(m+n)}$. Thus the set of all orthogonal reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ such that $\beta(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = \mathbf{0}$ will be an $(n(m+n) - \sum_{j=1}^n z_j)$-dimensional smooth manifold in $\mathbb{R}^{n(m+2n)}$. If $\mathbf{\Sigma}$ is symmetric and positive definite and $\mathbf{Q} = (\mathbf{q}_1, \cdots, \mathbf{q}_n)$, then $\mathbf{Q}$ will be orthogonal and $\beta(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = \mathbf{0}$ if and only if $\mathbf{M}_j(\mathbf{B}, \mathbf{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})\mathbf{q}_j = \mathbf{0}$ and $\mathbf{q}_j'\mathbf{q}_j = 1$, for $1 \le j \le n$. Thus a dimension argument shows that the set of all orthogonal reduced-form parameters $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ such that $\beta(f_h^{-1}(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})) = \mathbf{0}$ will be an $(n(m+n) - \sum_{j=1}^n z_j)$-dimensional smooth manifold in $\mathbb{R}^{n(m+2n)}$ if and only if $\mathbf{M}_j(\mathbf{B}, \mathbf{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ is of full row rank for any such $(\mathbf{B}, \mathbf{\Sigma}, \mathbf{Q})$ and all $1 \le j \le n$. We show that (3) holds. For non-singular matrices, the Gram-Schmidt process is well-defined, is continuously differentiable, and produces the QR-decomposition with the diagonal of the triangular matrix positive. We shall use this fact to define the set V and the functions $\mathbf{K}_j$.

In Section 5.1, the matrix $\mathbf{M}_j(\mathbf{B}, \mathbf{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ was defined for all $m \times n$ matrices $\mathbf{B}$, all symmetric and positive definite $n \times n$ matrices $\mathbf{\Sigma}$, and all orthonormal n-dimensional vectors $(\mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$.

---

[35]Since $\mathbf{Z}_j$ is $z_j \times r$ for $1 \le j \le n$, $\text{diag}_{1 \le j \le n}(\mathbf{Z}_j)$ is the $(\sum_{j=1}^n z_j) \times (rn)$ block diagonal matrix with the $\mathbf{Z}_j$ appearing along the diagonal.

Using g_h, defined in Appendix A.2, we can extend this to $\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ for all $m \times n$ matrices $\mathbf{B}$, all positive definite $n \times n$ matrices $\boldsymbol{\Sigma}$, and all $n \times n$ matrices $\mathbf{Q}$ by

$$\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{j-1} & (\mathbf{Z}_j \mathbf{F}(g_h(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{I}_n)))' \end{bmatrix}',$$

where $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_n]$. For $1 \leq j \leq n$, let $\mathbf{W}_j$ be some fixed $(n + 1 - j - z_j) \times n$ matrix and let $\tilde{\mathbf{M}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = [\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})' \ \ \mathbf{W}_j']'$. Let V be the set of all $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) \in \mathbb{R}^{n(m+2n)}$ such that $\boldsymbol{\Sigma}$ is positive definite and $\tilde{\mathbf{M}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ is non-singular for $1 \leq j \leq n$, which is an open set. Let

$$\tilde{\mathbf{M}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})' = \begin{bmatrix} \underbrace{\tilde{\mathbf{K}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})}_{n \times (j-1+z_j)} & \underbrace{\mathbf{K}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})}_{n \times (n+1-j-z_j)} \end{bmatrix} \begin{bmatrix} \underbrace{\mathbf{R}_{j,1,1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})}_{(j-1+z_j) \times (j-1+z_j)} & \underbrace{\mathbf{R}_{j,1,2}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})}_{(j-1+z_j) \times (n+1-j-z_j)} \\ \underbrace{\mathbf{0}}_{(n+1-j-z_j),(j-1+z_j)} & \underbrace{\mathbf{R}_{j,2,2}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})}_{(n+1-j-z_j) \times (n+1-j-z_j)} \end{bmatrix}$$

be the QR-decomposition with the diagonals of the upper triangular matrices $\mathbf{R}_{j,1,1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ and $\mathbf{R}_{j,2,2}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ positive. In light of the above discussion, $\mathbf{K}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ is continuously differentiable over V. Since $\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \mathbf{R}_{j,1,1}(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})' \tilde{\mathbf{K}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})'$, the columns of $\mathbf{K}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ form an orthonormal basis for the null space of $\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$.

All that remains to be shown is that any structural parameters satisfying the zero restrictions is almost surely in the set $U = f_h^{-1}(V)$. From (2), we know that $\mathbf{M}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ is of full row rank for any $(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q})$ satisfying the zero restrictions and all $1 \leq j \leq n$. This implies that for almost all choices of the $\mathbf{W}_j$, the matrices $\tilde{\mathbf{M}}_j(\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{q}_1, \cdots, \mathbf{q}_{j-1})$ will be non-singular. From this it follows that almost all structural parameters satisfying the zero restrictions will be in U. Furthermore, this suggests that choosing the $\mathbf{W}_j$ to be some fixed random matrices, say standard normal, will almost surely deliver the desired functions $\mathbf{K}_j$.

B   The IRF Parameterization

In this appendix we describe the mapping between the orthogonal reduced-form parameterization and the IRF parameterization. To do so, first we define the IRF parameterization as a function of the structural parameters. Notice that the IRF of the ith variable to the jth structural shock at horizon $k \leq 0$ is the element in row i and column j of the matrix $\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+)$, where $\mathbf{L}_0(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{A}_0^{-1})'$ and $\mathbf{L}_k(\mathbf{A}_0, \mathbf{A}_+) = \sum_{\ell=1}^{\min\{k,p\}} (\mathbf{A}_\ell \mathbf{A}_0^{-1})' \mathbf{L}_{k-\ell}(\mathbf{A}_0, \mathbf{A}_+)$, for $k > 0$. We combine the IRFs at horizons one through p and the constant term $\mathbf{c}(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{c}$ (which is the last row of $\mathbf{A}_+$) into a single matrix

$\mathbf{L}_+ = \begin{bmatrix} \mathbf{L}_1' & \cdots & \mathbf{L}_p' & \mathbf{c}' \end{bmatrix}'$. We call $(\mathbf{L}_0, \mathbf{L}_+)$ the IRF parameterization.

The expressions above define a mapping, which we will denote by ϕ, from the structural parameterization to the IRF parameterization. This mapping is invertible and its inverse is given by

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

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

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

and $\mathbf{c}(\mathbf{L}_0, \mathbf{L}_+) = \mathbf{c}$ is the last row of $\mathbf{L}_+$.

The mapping between the structural parameterization and the IRF parameterization suggests that Equation (1) can be rewritten in terms of the IRF parameterization as follows

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

Hence, in addition to the orthogonal reduced-form parameterization and the structural parameterization, a VAR can be written in terms of the IRF parameterization. Moreover, we have defined the mapping f_h from the structural parameterization to the orthogonal reduced-form parameterization, and the mapping ϕ from the structural parameterization to the IRF parameterization. This implies that $\phi \circ f_h^{-1}$ defines a mapping from the orthogonal reduced-form parameterization to the IRF parameterization.

C Empirical Application: Estimation and Inference

Following Beaudry, Nam and Wang (2011) we consider four lags and use a normal-inverse-Wishart prior distribution, $NIW(\bar{\nu}, \bar{\boldsymbol{\Phi}}, \bar{\boldsymbol{\Psi}}, \bar{\boldsymbol{\Omega}})$, for the reduced-form parameters. We set $\bar{\nu} = 0$, $\bar{\boldsymbol{\Phi}} = \mathbf{0}$, and $\bar{\boldsymbol{\Psi}} = \mathbf{0}$ and $\bar{\boldsymbol{\Omega}}^{-1} = \mathbf{0}$. 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 prices, consumption, real interest rate, and hours worked. TFP is the factor-utilization-adjusted TFP series from John Fernald's website. Stock prices correspond to the Standard & 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 interest 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 correspond to 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 in log percentages except for the real interest rate, which is expressed in percentages. This choice of scale for the variables seems to produce the most accurate results when approximating the volume elements.

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

Beaudry, P., D. Nam, and J. Wang (2011). Do Mood Swings Drive Business Cycles and is it Rational? *NBER Working Papers* (17651).

Binning, A. (2013). Underidentified SVAR Models: A Framework for Combining Short and Long-Run Restrictions with Sign-Restrictions. *Norges Bank Working Papers* (2013|14).

Caldara, D. and C. Kamps (2012). The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers. *International 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.

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

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016a). Delta-Method Inference for a Class of Set-Identified SVARS. *Working Paper*.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016b). Projection Inference for Set-Identified 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. *Working Paper*.

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

Lütkephol, H. (1993). *Introduction to Multiple Time Series Analysis.* Springer-Verlag.

Moon, H. R. and F. Schorfheide (2012). Bayesian and Frequentist Inference in Partially Identified Models. *Econometrica 80*(2), 755–782.

Moon, H. R., F. Schorfheide, and E. Granziera (2013). Inference for VARs Identified with Sign Restrictions. *Working Paper*.

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

Munkres, J. R. (1991). *Analysis on Manifolds.* Addison-Wesley.

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

Royden, H. L. (1963). *Real Analysis.* MacMillian.

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.

Sims, C. A. and T. Zha (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 949–968.

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.