Inference in Bayesian Proxy-SVARs

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

Abstract: Motivated by the increasing use of external instruments to identify structural vector autoregressions (SVARs), we develop an algorithm for exact finite sample inference in this class of time series models, commonly known as Proxy-SVARs. Our algorithm makes independent draws from any posterior distribution over the structural parameterization of a Proxy-SVAR. Our approach allows researchers to simultaneously use proxies and traditional zero and sign restrictions to identify structural shocks. We illustrate our methods with two applications. In particular, we show how to generalize the counterfactual analysis in Mertens and Montiel-Olea (2018) to identified structural shocks.

1 Introduction

The method of identification of structural vector autoregressions (SVARs) with external instruments, commonly known as Proxy-SVARs, has grown to become influential in empirical macroeconomics. For example, see Stock (2008); Stock and Watson (2012); Mertens and Ravn (2013); Gertler and Karadi (2015); Montiel-Olea, Stock and Watson (2016). This paper describes how to conduct Bayesian inference in this class of structural models.

We contribute to this line of research by developing an efficient algorithm to independently draw from any posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions and the γ-relevance condition. The former requires that the correlation between the proxies and some subset of the structural shocks be zero, while the latter requires that the correlation between the proxies and the remaining shocks be bounded away from zero. The fact that we can draw independently opens the door to using the Bayesian paradigm in larger models. We will write our algorithm as independently drawing from the family of restricted normal-generalized-normal (NGN) posterior distributions over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions and the γ-relevance condition. However, our techniques are not limited to the NGN family and can be applied to any prior over the structural parameterization of a Proxy-SVAR.

We achieve our goal by first independently drawing triangular-block parameters using Waggoner and Zha's (2003) sampler. The triangular-block parameters play the same role as the reduced-form parameters do in the traditional approach. Then, we show that the exogeneity restrictions are linear restrictions on the columns of an orthogonal matrix. This will allow us to draw orthogonal matrices, conditional on each draw of the triangular-block parameters, such that the exogeneity restrictions and the γ-relevance condition hold. Then, we map the orthogonal triangular-block parameters into Proxy-SVAR structural parameters conditional on the exogeneity restrictions and the γ-relevance condition. Finally, we show how to numerically compute the density associated with the implied distribution over the Proxy-SVAR structural parameterization. Hence, we can use those draws as an intermediate step in an importance sampler to draw from any desired posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions and the γ-relevance condition.

Since the exogeneity restrictions may not be enough to identify the Proxy-SVAR equations associated with structural shocks that are correlated with the proxies, additional zero and sign restrictions are needed for identification when more than one proxy is used to identify the same number of Proxy-SVAR equations. Our algorithm can handle these additional restrictions, which could be used to identify not only the Proxy-SVAR equations associated with the structural shocks correlated with the proxies but also the Proxy-SVAR equations

associated with those structural shocks that are uncorrelated with the proxies.

We present two applications to illustrate our algorithm. The first application is aimed at providing applied readers with a succinct and comprehensive description of how to use our techniques. To this end, we begin by revisiting Lunsford's (2016) study on the dynamic effects of consumption and investment total factor productivity (TFP) shocks in a Proxy-SVAR. An important difference between our approach and Lunsford's (2016) is that, while he identifies one structural equation at a time by using a single instrument, which is a common approach in the literature (see Stock and Watson, 2012), we use additional zero and sign restrictions to jointly identify two structural equations using two instruments. In particular, we identify the structural equations by assuming that they are the only equations whose structural shocks are correlated with the two external instruments and by adding some additional sign restrictions to parse out consumption TFP shocks from investment TFP shocks.

The second application is aimed at highlighting that our approach can provide critical insights for a few but highly influential studies using two instruments such as Mertens and Ravn (2013) and Mertens and Montiel-Olea (2018). We will make this clear by revisiting Mertens and Montiel-Olea (2018). That paper relies on two proxies to study the effects of counterfactual changes in marginal and average personal income tax rates. One of its main conclusions is that substitution effects are more important than income effects in the transmission of tax rate changes. We will argue that the counterfactual experiments are narrow because they focus on a particular linear combination of structural shocks rather than on the individual structural shocks. As a result, we propose to separately identify the structural shocks based on a set of sign restrictions. We find that both substitution and income effects play a relevant role in the transmission of tax rate shocks.

Only a handful of papers consider Proxy-SVARs under the Bayesian paradigm. Bahaj (2014), Drautzburg (2016), and Braun and Brüggemann (2017) use Gibbs samplers; therefore the draws are not independent. More importantly, they ignore the effects that the parameter transformations embedded in their approach have on the posterior and as consequence the order of the instruments affects the results; hence, these methods are not appropriate for inference. Giacomini, Kitagawa, and Read (2020) expand the robust Bayesian inference methods in Giacomini and Kitagawa (2018) to Proxy-SVARs.

Next, let's relate our paper to Caldara and Herbst (2016). An advantage of Caldara and Herbst's (2016) approach relative to ours is that they can use more than one proxy to identify a single shock. Nevertheless, the posterior draws are not independent and their Metropolis-Hastings sampler could become computationally inefficient compared with ours in large models. Finally, Jarociński and Karadi (2018) assume that the structural shocks are linear combinations of the proxies; however, a Proxy-SVAR only assumes that the structural shocks

2

are correlated with linear combinations of the proxies.

2 The Framework

This section discusses our general framework. In Section 2.1, we describe the structural parameterization of the Proxy-SVAR. In Sections 2.2 and 2.3, we present the identification problem, the exogeneity restrictions, the γ-relevance condition, and the need for additional zero or sign restrictions. In Section 2.4, we provide an outline of our methodology. In Section 2.5, we explicitly specify the restricted NGN family of prior and posterior distributions over the structural parameterization of a Proxy-SVAR that we will use to illustrate our algorithm. It is important to keep in mind that our methods can be used to independently draw from any posterior distribution. In Section 2.6, we introduce the orthogonal triangular-block parameterization and its mapping into the structural parameterization of the Proxy-SVAR.

2.1 A Proxy-SVAR

Let $\boldsymbol{y}_t$ be an $n \times 1$ vector of endogenous variables, $\boldsymbol{m}_t$ be a $k \times 1$ vector of instruments (also called proxies), $\tilde{\boldsymbol{y}}_t' = [\boldsymbol{y}_t' \ \boldsymbol{m}_t']$, and $\tilde{n} = n + k$. If these are governed by an SVAR, then

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

where $\tilde{\boldsymbol{A}}_i$ is an $\tilde{n} \times \tilde{n}$ matrix for $0 \leq i \leq p$ with $\tilde{\boldsymbol{A}}_0$ invertible, $\tilde{\boldsymbol{c}}$ is a $1 \times \tilde{n}$ row vector, and $\tilde{\boldsymbol{\varepsilon}}_t$ is conditionally standard normal. If $\tilde{\boldsymbol{x}}_t' = [\tilde{\boldsymbol{y}}_{t-1}' \ \cdots \ \tilde{\boldsymbol{y}}_{t-p}' \ 1]$ and $\tilde{\boldsymbol{A}}_+' = [\tilde{\boldsymbol{A}}_1' \cdots \tilde{\boldsymbol{A}}_p' \ \tilde{\boldsymbol{c}}']$, Equation (1) can be written as

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

Let $\tilde{\boldsymbol{\varepsilon}}_t' = [\boldsymbol{\varepsilon}_t' \ \boldsymbol{v}_t']$, where $\boldsymbol{\varepsilon}_t$ is $n \times 1$ and $\boldsymbol{v}_t$ is $k \times 1$. Since $\tilde{\boldsymbol{\varepsilon}}_t$ is conditionally standard normal, $\boldsymbol{v}_t$ is uncorrelated with $\boldsymbol{\varepsilon}_t$. A Proxy-SVAR imposes that $\boldsymbol{y}_t$ evolves according to $\boldsymbol{y}_t' \boldsymbol{A}_0 = \boldsymbol{x}_t' \boldsymbol{A}_+ + \boldsymbol{\varepsilon}_t'$ for $1 \leq t \leq T$, where $\boldsymbol{x}_t' = [\boldsymbol{y}_{t-1}' \ \cdots \ \boldsymbol{y}_{t-p}' \ 1]$ and $\boldsymbol{A}_+' = [\boldsymbol{A}_1' \cdots \boldsymbol{A}_p' \ \boldsymbol{c}']$, with $\boldsymbol{A}_i$ an $n \times n$ matrix for $0 \leq i \leq p$, $\boldsymbol{A}_0$ invertible, and $\boldsymbol{c}$ a $1 \times n$ row vector. The $\boldsymbol{\varepsilon}_t$ are the structural shocks and the $\boldsymbol{v}_t$ are other shocks that affect the proxies, hence

$$\tilde{\boldsymbol{A}}_i = \begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{\Gamma}_{i,1} \\ \boldsymbol{0}_{k \times n} & \boldsymbol{\Gamma}_{i,2} \end{bmatrix},$$

3

where $\boldsymbol{\Gamma}_{i,1}$ is $n \times k$ and $\boldsymbol{\Gamma}_{i,2}$ is $k \times k$ for $0 \leq i \leq p$ and $\boldsymbol{0}_{k \times n}$ is a $k \times n$ matrix of zeros. We could have set $\boldsymbol{\Gamma}_{0,2}$ to be equal to a $k \times k$ identity matrix and allowed the $\boldsymbol{v}_t$ to be correlated among themselves. We call these zero restrictions on $\tilde{\boldsymbol{A}}_0$ and $\tilde{\boldsymbol{A}}_+$ the block restrictions. We call Equation (2), together with the block restrictions, the structural parameterization of the Proxy-SVAR and $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, such that the block restrictions hold, the Proxy-SVAR structural parameters. We call the unrestricted $(\boldsymbol{A}_0, \boldsymbol{A}_+)$ the SVAR structural parameters.

Notice that while the specification of our Proxy-SVAR is similar to the one in Mertens and Ravn (2013) and Stock and Watson (2018), there are two main differences. First, we use a parametric model, whereas the aforementioned papers use a semi-parametric model. Second, we restrict the structural innovations to be conditionally homoscedastic and Gaussian. The latter is a common assumption in set-identified SVAR analysis, but some VAR studies have relaxed it following Gonçalves and Kilian (2004). In any case, considering heteroscedastic structural shocks is still an open question (see Bognanni, 2018).

2.2 The Identification Problem in a Proxy-SVAR

Following Rothenberg (1971) the Proxy-SVAR structural parameters $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent if and only if they imply the same joint distribution of $\tilde{\boldsymbol{y}}_1, \cdots, \tilde{\boldsymbol{y}}_T$. It is easy to show that the Proxy-SVAR structural parameters $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent if and only if $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0 \boldsymbol{Q}$ and $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+ \boldsymbol{Q}$, for some matrix $\boldsymbol{Q} \in \mathcal{Q} \subset \mathcal{O}(\tilde{n})$, where $\mathcal{Q}$ is defined by $\mathcal{Q} = \{\boldsymbol{Q} \in \mathcal{O}(\tilde{n}) | \boldsymbol{Q} = \mathrm{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2), \boldsymbol{Q}_1 \in \mathcal{O}(n), \text{and } \boldsymbol{Q}_2 \in \mathcal{O}(k)\}$, $\mathrm{diag}(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m)$ is the block diagonal matrix with the matrices $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m$ along the diagonal, and $\mathcal{O}(m)$ is the set of all $m \times m$ orthogonal matrices. The identification problem in Proxy-SVARs is commonly a partial identification problem because researchers focus on identifying a subset of the Proxy-SVAR equations. For ease of exposition, we adopt Leeper, Sims, and Zha's (1996) view and use the term identifying structural shocks as equivalent to identifying structural equations. A Proxy-SVAR equation is identified if, for any two sets of observationally equivalent Proxy-SVAR parameters, the parameters in that equation are identical.

The identification problem in Proxy-SVARs is typically addressed by assuming that the k proxies are correlated with k structural shocks in $\boldsymbol{\varepsilon}_t$ and uncorrelated with the remaining structural shocks. Without loss of generality let the structural shocks correlated with the proxies be the last k elements of $\boldsymbol{\varepsilon}_t$ and the structural shocks uncorrelated with the proxies be the first $n - k$ elements of $\boldsymbol{\varepsilon}_t$. We now show that the latter restrictions—which are known in the literature as exogeneity restrictions—are zero restrictions on a non-linear function of the Proxy-SVAR structural parameters. To see this, note that by multiplying Equation (2) by $\tilde{\boldsymbol{A}}_0^{-1}$ and focusing on the last k equations we obtain $\boldsymbol{m}_t' = \tilde{\boldsymbol{y}}_t' \boldsymbol{J}' = \tilde{\boldsymbol{x}}_t' \tilde{\boldsymbol{A}}_+ \tilde{\boldsymbol{A}}_0^{-1} \boldsymbol{J}' + \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{A}}_0^{-1} \boldsymbol{J}'$, for $1 \leq t \leq T$,

where $\boldsymbol{J} = [\boldsymbol{0}_{k \times n} \quad \boldsymbol{I}_k]$. It follows that $\mathbb{E}[\boldsymbol{m}_t \boldsymbol{\varepsilon}_t'] = \mathbb{E}[\boldsymbol{m}_t \tilde{\boldsymbol{\varepsilon}}_t' \boldsymbol{L}'] = \boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}'$, where $\boldsymbol{L} = [\boldsymbol{I}_n \quad \boldsymbol{0}_{n \times k}]$. Thus, the exogeneity restrictions imply that the first $n - k$ columns of matrix $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}'$ must be zero, which makes clear that Proxy-SVARs are identified by zero restrictions on a non-linear function of the Proxy-SVAR structural parameters. In addition to the exogeneity restrictions, we also need the covariance matrix of the last k structural shocks and the k proxies, i.e., the last k columns of $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}'$, to be non-singular. As in the literature, we refer to this as the relevance condition. One may want to control the strength of the relevance condition. In Section 2.5, we show how to do so.

2.3   The Need for Additional Restrictions

The exogeneity restrictions and the relevance condition only allow us to categorize the structural shocks into two groups: the ones that are correlated with the proxies and the ones that are not correlated with the proxies. If we only use the exogeneity restrictions and the relevance condition, we have an identification problem among the structural shocks that are correlated with the proxies unless $k = 1$. We need additional identification restrictions to identify the structural shocks within the set of structural shocks that are correlated with the proxies. The additional restrictions can be either sign or zero restrictions, or both. For example, the zero restrictions can be imposed on the contemporaneous IRFs or on the matrix of contemporaneous coefficients, but can be more general. Jentsch and Lunsford (2019a) describe the same problem and they give two examples of zero restrictions that can be used. Our approach can consider all of their zero restrictions. It is important to note that while Caldara and Herbst's (2016) paper is the one closest to ours, we suspect that it would be challenging to implement additional zero restrictions with their approach.

In particular, it is easy to show that $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$, Proxy-SVAR structural parameters that also satisfy the exogeneity restrictions and the relevance condition, are observationally equivalent if and only if there exists a matrix $\boldsymbol{Q} \in \mathcal{X} \subset \mathcal{Q} \subset \mathcal{O}(\tilde{n})$ such that $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0 \boldsymbol{Q}$ and $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+ \boldsymbol{Q}$, where $\mathcal{X}$ is defined by $\mathcal{X} = \{\boldsymbol{Q} \in \mathcal{Q} | \boldsymbol{Q} = \text{diag}(\boldsymbol{Q}_3, \boldsymbol{Q}_4, \boldsymbol{Q}_5), \boldsymbol{Q}_3 \in \mathcal{O}(n - k), \boldsymbol{Q}_4 \in \mathcal{O}(k), \text{and } \boldsymbol{Q}_5 \in \mathcal{O}(k)\}$. Note that $\boldsymbol{Q}_3$ rotates the columns of the Proxy-SVAR structural parameters associated with the structural shocks that are not correlated with the proxies while, $\boldsymbol{Q}_4$ rotates the columns of the Proxy-SVAR structural parameters associated with the structural shocks that are correlated with the proxies. Often, one is interested only in the partial identification of the k structural shocks that are correlated with the k proxies. If that is the case and $k = 1$, the exogeneity restrictions and the relevance condition exactly identify the structural shock correlated with the proxy, up to a sign.

Although most of the studies relying on Proxy-SVAR analysis use one instrument to identify one structural

shock, a growing literature considers the case in which several instruments are used to identify several structural shocks or to conduct counterfactual experiments based on linear combinations of the latter. Braun and Brüggemann (2017), Piffer and Podstawski (2017), Jarociński and Karadi (2018), Lakdawala (2019), Känzig (2019), Giacomini, Kitagawa, and Read (2020), Jentsch and Lunsford (2019a) and Jentsch and Lunsford (2019b) are examples of papers that explicitly aim to identify multiple structural shocks with multiple instruments. Braun and Brüggemann (2017) identify oil market and monetary policy shocks by combining sign restrictions with information obtained from external instruments. Piffer and Podstawski (2017) identify uncertainty and news shocks by combining exogeneity restrictions arising from external instruments with sign restrictions. Jarociński and Karadi (2018) try to simultaneously identify monetary policy shocks and news shocks using external information. Lakdawala (2019) identifies fed funds rate and forward guidance shocks. Känzig (2019) identifies two types of oil shocks: supply news shocks and unanticipated oil supply shocks by combining exogeneity restrictions arising from external instruments with a zero restriction on a contemporaneous impulse response function. Giacomini, Kitagawa, and Read (2020) expand their robust analysis to Proxy-SVAR models with more than one instrument to identify the same number of structural shocks. Jentsch and Lunsford (2019a) and Jentsch and Lunsford (2019b) expand the techniques in Mertens and Ravn (2013) to larger systems.

2.4   The Road Map

In this paper, we develop an efficient algorithm to independently draw from any posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions. However, we will write our algorithm as independently drawing from the family of restricted NGN posterior distributions over the structural parameterization of a Proxy-SVAR. The NGN distribution is a conjugate family of distributions commonly used in the literature. For instance, the Sims-Zha prior (see Sims and Zha, 1998) is conjugate and in the NGN family. Conjugate priors and posteriors have the appealing property of giving the same weight to observationally equivalent Proxy-SVAR structural parameters.

Although a Proxy-SVAR identified with exogeneity restrictions can be represented by an SVAR identified with zero restrictions, one cannot use Arias, Rubio-Ramírez and Waggoner's (2018) algorithms because of the large number of zero restrictions embedded in a Proxy-SVAR. For this reason, we introduce a new parameterization called the orthogonal triangular-block parameterization, which plays a role parallel to the one played by the orthogonal reduced-form parameterization in Arias, Rubio-Ramírez and Waggoner (2018). This parameterization is useful because one can independently draw orthogonal triangular-block parameters

conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions. These independent draws are mapped into the Proxy-SVAR structural parameterization and weighted, so that they come from the desired posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions.

2.5 Priors and Posteriors

We will use a restricted NGN distribution over the Proxy-SVAR structural parameterization as our prior distribution to describe our algorithm. A restricted NGN distribution over the structural parameterization of the Proxy-SVAR is an NGN distribution over $\mathbb{R}^{\tilde{n}^2+\tilde{m}\tilde{n}}$ conditional on the block restrictions. Let $\tilde{m} = p\tilde{n}+1$; if there are $\tilde{e}$ exogenous variables, then $\tilde{m} = p\tilde{n}+1+\tilde{e}$. Hence, the Proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ have a prior density proportional to $NGN_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, where

$$NGN_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \propto |\det(\tilde{\boldsymbol{A}}_0)|^{\nu-n} e^{-\frac{1}{2}\text{vec}(\tilde{\boldsymbol{A}}_0)'\boldsymbol{\Phi}\,\text{vec}(\tilde{\boldsymbol{A}}_0)} e^{-\frac{1}{2}(\text{vec}(\tilde{\boldsymbol{A}}_+)-\boldsymbol{\Psi}\,\text{vec}(\tilde{\boldsymbol{A}}_0))'\boldsymbol{\Omega}^{-1}(\text{vec}(\tilde{\boldsymbol{A}}_+)-\boldsymbol{\Psi}\,\text{vec}(\tilde{\boldsymbol{A}}_0))}. \quad (3)$$

The density is characterized by four parameters: a scalar $\nu \geq \tilde{n}$, an $\tilde{n}^2 \times \tilde{n}^2$ block diagonal matrix $\boldsymbol{\Phi}$ with $\tilde{n}$ symmetric and positive definite $\tilde{n} \times \tilde{n}$ blocks, an $\tilde{m}\tilde{n} \times \tilde{n}^2$ block diagonal matrix Ψ with $\tilde{n}$ blocks of size $\tilde{m} \times \tilde{n}$, and an $\tilde{m}\tilde{n} \times \tilde{m}\tilde{n}$ block diagonal matrix $\boldsymbol{\Omega}$ with $\tilde{n}$ symmetric and positive definite $\tilde{m} \times \tilde{m}$ blocks.[1] In general, the marginal prior distribution of $vec(\tilde{\boldsymbol{A}}_0)$ is a generalized-normal distribution, but when $\nu = \tilde{n}$ it is a normal distribution with mean zero and variance $\boldsymbol{\Phi}^{-1}$. The parameter ν controls how close the generalized-normal distribution is to a normal distribution. The prior distribution of $vec(\tilde{\boldsymbol{A}}_+)$, conditional on $\tilde{\boldsymbol{A}}_0$, is a normal with mean $\boldsymbol{\Psi}vec(\boldsymbol{A}_0)$ and variance $\boldsymbol{\Omega}$.

If the exogeneity restrictions hold, then we can write $\mathbb{E}[\boldsymbol{m}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}' = [\boldsymbol{0}_{k\times(n-k)} \quad \boldsymbol{V}]$, where the $k \times k$ matrix $\boldsymbol{V}$ is the covariance matrix of the k proxy variables and the last k structural shocks. The relevance condition implies that $\boldsymbol{V}$ is non-singular. Following Caldara and Herbst (2016), one may want to introduce prior beliefs about how much of the variance of the instruments must be related to the underlying structural shocks of interest. We will introduce these prior beliefs by truncating the density described in Equation (3) so that the minimum eigenvalue of the reliability matrix, which is $\left(\boldsymbol{\Gamma}_{0,2}^{-1\prime}\boldsymbol{\Gamma}_{0,2}^{-1} + \boldsymbol{V}\boldsymbol{V}'\right)^{-1}\boldsymbol{V}\boldsymbol{V}'$, is greater than γ, where $0 \leq \gamma < 1$. This implies that at least γ percent of the variance of any linear combination of the instruments must be related to the underlying structural shocks of interest; see Gleser (1992). As we will see in our applications, the truncation may improve the efficiency of our algorithm. In some cases, choosing a very

---

[1] Arias, Rubio-Ramírez, and Waggoner (2018) assumed a Kronecker structure for $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Omega}$, while here only a block diagonal structure is assumed.

low value of γ may imply a slow and inefficient algorithm, because some weights can become very large. We say that the parameters that survive the truncation satisfy the γ-relevance condition.

Our algorithm will independently draw from the restricted NGN posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions, the γ-relevance condition, and any additional zero and sign restrictions implied by the above described prior. More specifically, given a prior density proportional to $NGN_{(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, the posterior density is proportional to $NGN(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\nu} = T + \nu$, $\tilde{\boldsymbol{\Omega}} = (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{X}}' \tilde{\boldsymbol{X}} + \boldsymbol{\Omega}^{-1})^{-1}$, $\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Omega}}(\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{X}}' \tilde{\boldsymbol{Y}} + \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi})$, and $\tilde{\boldsymbol{\Phi}} = \boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{Y}}' \tilde{\boldsymbol{Y}} + \boldsymbol{\Phi} + \boldsymbol{\Psi}' \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi} - \tilde{\boldsymbol{\Psi}}' \tilde{\boldsymbol{\Omega}}^{-1} \tilde{\boldsymbol{\Psi}}$, $\tilde{\boldsymbol{Y}} = [\tilde{\boldsymbol{y}}_1 \; \cdots \; \tilde{\boldsymbol{y}}_T]'$, and $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1 \; \cdots \; \tilde{\boldsymbol{x}}_T]'$.

2.6 The Orthogonal Triangular-Block Parameterization

Since a Proxy-SVAR identified with exogeneity restrictions can be represented by the SVAR in Equation (2), one would like to use Arias, Rubio-Ramírez and Waggoner's (2018) algorithm. However, the techniques of that paper cannot be directly applied in this context because the number of zero restrictions implied by the block restrictions alone is too large. There are $(p+1)k$ block restrictions on each of the first n columns of $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, whereas the maximum number of restrictions that the aforementioned algorithm can handle on the jth column of the structural parameters is $\tilde{n} - j$. So unless $p = 0$, an uninteresting case, the maximum will be exceeded for the nth column, if not before. In this paper we show how to address this shortcoming.

The traditional approach is to map independent draws from the orthogonal reduced-form parameterization conditional on the zero restrictions into the structural parameterization of the SVAR to create a proposal for the desired posterior distribution over the structural parameterization of the SVAR conditional on the zero restrictions. The key to such an approach is to properly characterize the proposal distribution over the structural parameterization. This proposal was then embedded in an importance sampling algorithm. Similarly, we will map what we call the orthogonal triangular-block parameterization conditional on the exogeneity restrictions, the γ-relevance condition, and any additional zero and sign restrictions into the structural parameterization of the Proxy-SVAR to create a proposal for the desired posterior distribution over the structural parameterization of the Proxy-SVAR conditional on the exogeneity restrictions, the γ-relevance condition, and any additional zero and sign restrictions. Again, the key will be to properly weight the draws in order to simulate from the desired distribution over the structural parameterization.

Let $\tilde{\boldsymbol{\Lambda}}_0$ be an $\tilde{n} \times \tilde{n}$ matrix, $\tilde{\boldsymbol{\Lambda}}_+$ be an $\tilde{m} \times \tilde{n}$ matrix, $\boldsymbol{Q}_1 \in \mathcal{O}(n)$, and $\boldsymbol{Q}_2 \in \mathcal{O}(k)$. The matrix $\tilde{\boldsymbol{\Lambda}}_0$ is restricted to be upper-triangular with positive diagonal. The matrix $\tilde{\boldsymbol{\Lambda}}_+' = [\tilde{\boldsymbol{\Lambda}}_1' \cdots \tilde{\boldsymbol{\Lambda}}_p' \; \boldsymbol{d}']$, where $\tilde{\boldsymbol{\Lambda}}_i$ is $\tilde{n} \times \tilde{n}$ for $1 \leq i \leq p$ and $\boldsymbol{d}$ is $1 \times \tilde{n}$, is restricted so that the lower left-hand $k \times n$ block of $\tilde{\boldsymbol{\Lambda}}_i$ is zero for $1 \leq i \leq p$. We

label the zero restrictions on $\tilde{\boldsymbol{\Lambda}}_0$ and $\tilde{\boldsymbol{\Lambda}}_+$ the triangular-block restrictions, and we call $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ such that the triangular-block restrictions hold the triangular-block parameters. We call $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ the orthogonal triangular-block parameters.

We can map the orthogonal triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ into Proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ by

$$(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) \overset{f}{\longrightarrow} (\underbrace{\tilde{\boldsymbol{\Lambda}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)}_{\tilde{\boldsymbol{A}}_0}, \underbrace{\tilde{\boldsymbol{\Lambda}}_+ \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)}_{\tilde{\boldsymbol{A}}_+}).$$

It is easy to verify that $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ will satisfy the block restrictions, so they are Proxy-SVAR structural parameters. Remember that we have assumed that $\tilde{\varepsilon}_t$ is conditionally standard normal; hence, we have normalized the variance matrix of the structural shocks to the identity matrix. One could implement any other normalization—such as the unit effect normalization adopted in Proxy-SVAR studies working under the frequentist paradigm—by appropriately modifying the function f. Our choice keeps the notation as close as possible to the notation in Arias, Rubio-Ramírez, and Waggoner (2018), and hence, it simplifies the implementation and interpretation of additional sign restrictions that we will introduce later. In addition, the normalization we adopt is in line with the normalization adopted by Proxy-SVAR studies working under the Bayesian paradigm; see, e.g., Bahaj (2014), Drautzburg (2016), Caldara and Herbst (2016), and Giacomini, Kitagawa, and Read (2020). For completeness, in Appendix A.1 we show the map associated with the unit effect normalization.

The mapping of f has an inverse. Let $\tilde{\boldsymbol{A}}_0^{-1} = \boldsymbol{P}\boldsymbol{R}$ be the QR-decomposition of $\tilde{\boldsymbol{A}}_0^{-1}$ normalized so that the diagonal of $\boldsymbol{R}$ is positive. Because the lower left-hand $k \times n$ block of $\tilde{\boldsymbol{A}}_0^{-1}$ is zero, $\boldsymbol{P} = \operatorname{diag}(\boldsymbol{P}_1, \boldsymbol{P}_2)$, where $\boldsymbol{P}_1 \in \mathcal{O}(n)$ and $\boldsymbol{P}_2 \in \mathcal{O}(k)$. The inverse of f is

$$(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \overset{f^{-1}}{\longrightarrow} (\underbrace{\tilde{\boldsymbol{A}}_0 \boldsymbol{P}}_{\tilde{\boldsymbol{\Lambda}}_0}, \underbrace{\tilde{\boldsymbol{A}}_+ \boldsymbol{P}}_{\tilde{\boldsymbol{\Lambda}}_+}, \underbrace{\boldsymbol{P}_1'}_{\boldsymbol{Q}_1}, \underbrace{\boldsymbol{P}_2'}_{\boldsymbol{Q}_2}).$$

The matrix $\tilde{\boldsymbol{\Lambda}}_0$ will be upper-triangular with positive diagonal because $\tilde{\boldsymbol{A}}_0 \boldsymbol{P} = \boldsymbol{R}^{-1}$. Furthermore, since $\boldsymbol{P}$ is block diagonal and the lower left-hand $k \times n$ block of $\tilde{\boldsymbol{A}}_i$ is zero, the lower left-hand $k \times n$ block of each $\tilde{\boldsymbol{\Lambda}}_i$ will be zero.

The orthogonal triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ define another parameterization of the Proxy-SVAR. We call this alternative parameterization the orthogonal triangular-block parameterization of a Proxy-SVAR and we write the latter as follows $\tilde{\boldsymbol{y}}_t' \tilde{\boldsymbol{\Lambda}}_0 = \tilde{\boldsymbol{x}}_t' \tilde{\boldsymbol{\Lambda}}_+ + \tilde{\boldsymbol{u}}_t'$ for $1 \le t \le T$, where $\tilde{\boldsymbol{u}}_t' = \tilde{\varepsilon}_t' \boldsymbol{Q}'$ with

$\boldsymbol{Q} = \text{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Like $\tilde{\boldsymbol{\varepsilon}}_t$, the innovations $\tilde{\boldsymbol{u}}_t$ are conditionally standard normal.

Importantly, we can produce independent draws of the triangular-block parameters. Furthermore, as we show in Section 3.2, the exogeneity and any additional zero restrictions are linear restrictions on the columns of the orthogonal matrix $\boldsymbol{Q}_1$, and hence, one can also efficiently and independently draw orthogonal matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$. The resulting draws can be mapped to the Proxy-SVAR structural parameterization using f as defined above. As will become clear in Section 3, these properties play a central role in the algorithm for inference proposed in this paper.

3   The Algorithm

In this section, we present Algorithm 1 to make independent draws from the restricted NGN posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions. Algorithm 1 starts by independently drawing triangular-block parameters, $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, from a restricted NGN posterior using Waggoner and Zha's (2003) Gibbs sampler. A restricted NGN distribution over the triangular-block parameters is an NGN distribution over $\mathbb{R}^{\tilde{n}^2 + \tilde{m}\tilde{n}}$ conditional on the triangular-block restrictions. This will be further discussed in Section 3.1 and Appendix A.2. The exogeneity and any additional zero restrictions are linear restrictions on the columns of the orthogonal matrix $\boldsymbol{Q}_1$, as will be discussed in Section 3.2. This will allow the use of the ideas in Arias, Rubio-Ramírez, and Waggoner (2018) to draw the orthogonal matrices $(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$, conditional on each draw of the triangular-block parameters, such that the exogeneity and any additional zero restrictions hold when $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ is mapped to $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, using the function f defined in Section 2.6. Draws that do not satisfy the γ-relevance condition or any additional sign restrictions are discarded. This is feasible because the set of Proxy-SVAR structural parameters that satisfy the γ-relevance condition and any additional sign restrictions is a subset of positive measure in the set of all Proxy-SVAR structural parameters that satisfy the exogeneity and any additional zero restrictions. These draws of $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ are not from the restricted NGN posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions, but in Section 3.4 we show how to numerically compute the density of each of these draws. In this section we also define the volume element that we will use to weight the draws. Thus we can importance weight these draws and re-sample to obtain independent draws from the desired distribution.[2] Section 3.5 highlights some practicalities when

---

[2]Re-sampling is not always necessary or desirable. Even without re-sampling, our draws are independent. This makes certain computations, such as computing moments, very efficient using all the weighted draws.

implementing Algorithm 1 and emphasizes some easy extensions of the algorithm. Finally, Section 3.6 discusses the importance of the volume element.

3.1 Independent Draws of the Triangular-Block Parameters

We use the Gibbs sampler of Waggoner and Zha (2003) to independently draw from a restricted NGN posterior distribution over the triangular-block parameters characterized by $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$. This Gibbs sampler can be used to draw from an NGN distribution subject to linear restrictions, as long as the restrictions do not involve cross-equation restrictions and the matrices $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ are block diagonal. The Gibbs sampler of Waggoner and Zha (2003) was developed to draw from the posterior distribution of a structural VAR with linear non-cross-equation restrictions using a certain class of normal priors. The class of posterior distributions that can be obtained with this class of priors is the set of NGN distributions, conditional on the linear non-cross-equation restrictions, described in Section 2.5. Since the triangular and block restrictions on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ do not involve cross-equation restrictions, and $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ can be chosen to be block diagonal, the conditions for using the Gibbs sampler are satisfied. Furthermore, because $\tilde{\boldsymbol{\Lambda}}_0$ is restricted to be upper-triangular, it follows from Theorem 2 of Waggoner and Zha (2003) that the Gibbs sampler draws will be independent. In Appendix A.2, we describe how to adapt their paper to our purposes.

Often, it suffices to choose $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ to be equal to $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, the parameters associated with the desired restricted NGN posterior distribution over the structural parameterization of the Proxy-SVAR conditional on the exogeneity restrictions, the relevance condition, and any additional zero and sign restrictions. However, sometimes this can lead to small effective sample sizes in our importance sampler. In Appendix A.3, we describe a more tailored choice of $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ that can avoid this loss of efficiency.

3.2 Restrictions on the Orthogonal Triangular-Block Parameters

As noted in Section 2.3, the exogeneity restrictions and γ-relevance condition do not fully identify the Proxy-SVAR parameters so that one may need to impose additional zero and sign restrictions. In this section we define the allowable additional zero and sign restrictions.

Because of the arguments made in Section 2.2, if $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ are Proxy-SVAR structural parameters, the exogeneity restrictions are of the form $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'\boldsymbol{e}_{n,j} = \boldsymbol{0}_{k \times 1}$, for $1 \leq j \leq n - k$, where $\boldsymbol{e}_{n,j}$ is the jth column of an identity matrix of dimension n. The index j stops at $n - k$ because there are no exogeneity restrictions

for $n - k < j \leq n$. In terms of the orthogonal triangular-block parameterization, this is equivalent to

$$\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'\boldsymbol{e}_{n,j} = \boldsymbol{J}((\tilde{\boldsymbol{\Lambda}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2))^{-1})'\boldsymbol{L}'\boldsymbol{e}_{n,j} = \boldsymbol{J}(\tilde{\boldsymbol{\Lambda}}_0^{-1})'\boldsymbol{L}'\boldsymbol{Q}_1\boldsymbol{e}_{n,j} = \boldsymbol{0}_{k \times 1} \text{ for } 1 \leq j \leq n - k. \quad (4)$$

Thus, conditional on a draw of triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, the exogeneity restrictions are linear restrictions on the columns of $\boldsymbol{Q}_1$. As in Arias, Rubio-Ramírez, and Waggoner (2018), this will be used to draw the orthogonal matrix $\boldsymbol{Q}_1$ conditional on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$.

The exogeneity restrictions are linear restrictions on the function of the Proxy-SVAR parameters given by $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'$ and the key condition that this function needed to satisfy was $\boldsymbol{J}((\tilde{\boldsymbol{A}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2))^{-1})'\boldsymbol{L} = \boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'\boldsymbol{Q}_1$. Let $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ be a function from the set of Proxy-SVAR structural parameters to the set of $r \times n$ matrices that satisfies

$$\boldsymbol{F}_z(\tilde{\boldsymbol{A}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2), \tilde{\boldsymbol{A}}_+ \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)) = \boldsymbol{F}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)\boldsymbol{Q}_1 \text{ for every } \boldsymbol{Q}_1 \in \mathcal{O}(n) \text{ and } \boldsymbol{Q}_2 \in \mathcal{O}(k). \quad (5)$$

We call functions that satisfy Equation (5) orthogonally commutative. Let $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) = [\boldsymbol{L}\tilde{\boldsymbol{A}}_0^{-1}\boldsymbol{J}' \quad \boldsymbol{F}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)']'$. Note that $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ is also orthogonally commutative. In addition to $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ being orthogonally commutative, a regularity condition is needed to ensure that there is sufficient variation in $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$. The exact condition is discussed in Appendix A.4. Allowable additional zero restrictions are linear restrictions on $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$.

Let $\tilde{\boldsymbol{Z}}_j$ be a $\tilde{z}_j \times (k+r)$ matrix of full row rank, where $k \leq \tilde{z}_j \leq n-j$, for $1 \leq j \leq n-k$, and $0 \leq \tilde{z}_j \leq n-j$, for $n-k < j \leq n$. Then, the exogeneity restrictions and allowable additional zero restrictions are of the form

$$\tilde{\boldsymbol{Z}}_j \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)\boldsymbol{e}_{n,j} = \boldsymbol{0}_{\tilde{z}_j \times 1}, \text{ for } 1 \leq j \leq n. \quad (6)$$

Because the zero restrictions represented in Equation (6) encode both the exogeneity and any additional zero restrictions, the first k rows of $\tilde{\boldsymbol{Z}}_j$ are equal to $[\boldsymbol{I}_k \quad \boldsymbol{0}_{k \times r}]$, for $1 \leq j \leq n-k$, and $\tilde{z}_j$ is the total number of restrictions, including both the exogeneity and any additional zero restrictions. Note that we are identifying only $\boldsymbol{\varepsilon}_t$. The number of restrictions could be zero for $n-k < j \leq n$. In this case, $\tilde{\boldsymbol{Z}}_j$ would be the empty $0 \times (k+r)$ matrix and $\tilde{\boldsymbol{Z}}_j \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)\boldsymbol{e}_{n,j}$ would be the empty 0×1 matrix. In principle, one could also use additional zero restrictions to identify $\boldsymbol{v}_t$, but that will rarely be of interest. Many restrictions used in the literature are of this form. For instance, we can impose linear restrictions on the last k columns of $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'$, which means that we can impose linear restrictions on the covariance matrix of the proxies and the shocks

correlated with the proxies, as long as the bounds on the number of restrictions are respected. Furthermore, we can impose linear restrictions on the impulse response of endogenous variables to structural shocks or on the SVAR structural parameters themselves.

From Equation (5) and the definition of f, the zero restrictions in the orthogonal triangular-block parameterization are

$$\tilde{\boldsymbol{Z}}_j \tilde{\boldsymbol{F}}_z(f(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)) \boldsymbol{e}_{n,j} = \underbrace{\tilde{\boldsymbol{Z}}_j \tilde{\boldsymbol{F}}_z(f(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{I}_n, \boldsymbol{I}_k))}_{\tilde{\boldsymbol{G}}_j(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)} \boldsymbol{Q}_1 \boldsymbol{e}_{n,j} = \boldsymbol{0}_{\tilde{z}_j \times 1}, \quad \text{for } 1 \leq j \leq n. \tag{7}$$

The function $\tilde{\boldsymbol{G}}_j(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ is used to impose both the exogeneity and any additional zero restrictions, which we see from Equation (7) are equivalent to linear restrictions on the columns of $\boldsymbol{Q}_1$ conditional on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$. To have a unified and compact notation, let d_i denote the size of the orthogonal matrix $\boldsymbol{Q}_i$, which is n when $i = 1$ and k when $i = 2$; let $\tilde{z}_{i,j}$ denote the number of restrictions on the jth column of $\boldsymbol{Q}_i$, which is $\tilde{z}_j$ for $i = 1$ and zero for $i = 2$; let $\tilde{\boldsymbol{G}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ be $\tilde{\boldsymbol{G}}_j(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ if $i = 1$ and the empty $0 \times k$ matrix if $i = 2$, and, finally, let $n_{i,j} = \tilde{z}_{i,j} + j - 1$.

We also allow for additional sign restrictions. The allowable sign restrictions are of the form $\tilde{\boldsymbol{F}}_s(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) > \boldsymbol{0}_{\tilde{s} \times 1}$, where $\tilde{\boldsymbol{F}}_s$ is any continuous function from $\mathbb{R}^{\tilde{n}^2 + \tilde{n}\tilde{m}}$ to $\mathbb{R}^{\tilde{s}}$. As with the exogeneity restrictions, we will express the γ-relevance condition as a sign restriction. We assume that the first row of $\tilde{\boldsymbol{F}}_s(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ is the minimum eigenvalue of the reliability matrix less γ, which is a continuous function. Because $\tilde{\boldsymbol{F}}_s$ is continuous, the set of all Proxy-SVAR structural parameters satisfying the zero and sign restrictions is an open subset of the Proxy-SVAR structural parameters satisfying just the zero restrictions. Thus, if the restrictions are non-degenerate, so that there is at least one value of the Proxy-SVAR structural parameters satisfying the zero and sign restrictions, then the set of all Proxy-SVAR structural parameters satisfying the zero and sign restrictions will be an open set of positive measure in the set of all Proxy-SVAR structural parameters satisfying just the zero restrictions. Because of this, it is feasible to make draws of Proxy-SVAR structural parameters satisfying just the zero restrictions, and then retain only the ones that also satisfy the sign restrictions.

As we have seen, the exogeneity restrictions are allowable zero restrictions and the γ-relevance condition is an allowable sign restriction. Henceforth, when we refer to the zero restrictions, this will include both the exogeneity and any allowable additional zero restrictions, and when we refer to the sign restrictions, this will include both the γ-relevance condition and any allowable additional sign restrictions.

3.3 The Algorithm

We now have the notation and concepts to state our simulation algorithm.

Algorithm 1. *The following algorithm makes independent draws from the restricted NGN posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the zero and sign restrictions.*

1. *Draw triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ independently from the restricted $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ distribution using Waggoner and Zha's (2003) Gibbs sampler.*

2. *For $i = 1, 2$ and $1 \leq j \leq d_i$, draw $\boldsymbol{\alpha}_{i,j} \in \mathbb{R}^{d_i - n_{i,j}}$ independently from a standard normal distribution and set $\boldsymbol{w}_{i,j} = \boldsymbol{\alpha}_{i,j} / \parallel \boldsymbol{\alpha}_{i,j} \parallel$.*

3. *For $i = 1, 2$ recursively define $\boldsymbol{Q}_i = [\boldsymbol{q}_{i,1} \ \cdots \ \boldsymbol{q}_{i,d_i}]$ by $\boldsymbol{q}_{i,j} = \boldsymbol{K}_{i,j} \boldsymbol{w}_{i,j}$ for any $d_i \times (d_i - n_{i,j})$ matrix $\boldsymbol{K}_{i,j}$ whose columns form an orthonormal basis for the null space of the $n_{i,j} \times d_i$ matrix*

$$\boldsymbol{M}_{i,j} = \left[\tilde{\boldsymbol{G}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)' \quad \boldsymbol{q}_{i,1} \quad \cdots \quad \boldsymbol{q}_{i,j-1} \right]'.$$

4. *Define $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) = f(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$.*

5. *If the sign restrictions are satisfied, retain the draw; otherwise, discard the draw and return to Step 1.*

6. *For each retained draw, set its importance weight to*

$$w_i = \frac{NGN_{(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)}{p(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)},$$

where $p(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ denotes the density of the draws obtained in Steps 1 thorough 4.

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

8. *Optionally, re-sample with replacement using the importance weights.*

The density p used in Step 6 will be explicitly computed in Section 3.4. In order for this algorithm to work, it must be the case that $\boldsymbol{M}_{i,j}$ is of full row rank; otherwise, the dimension of the null space of $\boldsymbol{M}_{i,j}$ would be strictly greater than $d_i - n_{i,j}$ and the matrix $\boldsymbol{K}_{i,j}$ would not exist. In Appendix A.4, we will show that $\boldsymbol{M}_{i,j}$ is almost surely of full row rank.

When there are no additional zero restrictions, $\boldsymbol{M}_{i,j}$, being of full row rank, has a nice interpretation in terms of the relevance condition. If the exogeneity restrictions hold, then $\mathbb{E}[\boldsymbol{m}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}' = [\boldsymbol{0}_{k \times (n-k)} \quad \boldsymbol{V}]$, where the $k \times k$ matrix $\boldsymbol{V}$ is the covariance matrix of the k proxy variables and the last k structural shocks. So, the relevance condition, which requires $\boldsymbol{V}$ to be non-singular, holds if and only if $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}'$ is of full row rank. When there are no additional zero restrictions, the matrix $\boldsymbol{M}_{i,j}$ will clearly be of full row rank when $i = 1$ and

$n - k < j \leq n$ or when $i = 2$ and $1 \leq j \leq k$. When $i = 1$ and $1 \leq j \leq n - k$, the matrix $\boldsymbol{M}_{i,j}$ will be of full row rank if and only if $\boldsymbol{J}(\tilde{\boldsymbol{\Lambda}}_0^{-1})'\boldsymbol{L}'$ is of full row rank. This is because, by construction, the $\boldsymbol{q}_{1,1}, \cdots, \boldsymbol{q}_{1,j-1}$ are perpendicular to each other and the rows of $\boldsymbol{J}(\tilde{\boldsymbol{\Lambda}}_0^{-1})'\boldsymbol{L}'$. Because $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}' = \boldsymbol{J}(\tilde{\boldsymbol{\Lambda}}_0^{-1})'\boldsymbol{L}'\boldsymbol{Q}_1$, the matrix $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'$ is of full row rank if and only if the matrix $\boldsymbol{J}(\tilde{\boldsymbol{\Lambda}}_0^{-1})'\boldsymbol{L}'$ is of full row rank. So, when there are no additional zero restrictions, the relevance condition is equivalent to $\boldsymbol{M}_{i,j}$ being of full row rank for all i and j. Of course, in practice, we not only want the covariance matrix to be non-singular, but we also would like it to be well conditioned so that it is far from being singular. We accomplish this by using the stronger γ-relevance condition, as explained in Section 2.5.

Even if $\boldsymbol{M}_{i,j}$ is of full row rank, the matrix $\boldsymbol{K}_{i,j}$ is not unique. If the columns of $\boldsymbol{K}_{i,j}$ form an orthonormal basis for the null space of $\boldsymbol{M}_{i,j}$, then any matrix whose columns form an orthonormal basis for the null space of $\boldsymbol{M}_{i,j}$ will be of the form $\boldsymbol{K}_{i,j}\boldsymbol{Q}$ for some $\boldsymbol{Q} \in \mathcal{O}(d_i - n_{i,j})$. Since $\boldsymbol{\alpha}_{i,j}$ is drawn from the standard normal distribution, $\boldsymbol{w}_{i,j}$ is drawn from the uniform distribution over the unit sphere in $\mathbb{R}^{d_i-n_{i,j}}$; so the distribution of $\boldsymbol{K}_{i,j}\boldsymbol{w}_{i,j}$ is identical to the distribution of $\boldsymbol{K}_{i,j}\boldsymbol{Q}\boldsymbol{w}_{i,j}$. So, when making draws, the choice of $\boldsymbol{K}_{i,j}$ does not matter. In terms of efficiency, we recommend taking $\boldsymbol{K}_{i,j}$ to be the last $d_i - n_{i,j}$ columns of the orthogonal component of the full QR-decomposition of $\boldsymbol{M}_{i,j}'$.[3]

Finally, we must show that Algorithm 1 does, in fact, independently draw from the posterior distribution over the Proxy-SVAR structural parameterization conditional on zero and sign restrictions. Steps 1 and 2 produce independent draws, so the algorithm also produces independent draws. Step 5 ensures that the sign restrictions are satisfied for the retained draws. Because the columns of $\boldsymbol{K}_{i,j}$ form a basis for the null space of $\boldsymbol{M}_{i,j}$, we have that $\boldsymbol{M}_{i,j}\boldsymbol{q}_{i,j} = \boldsymbol{M}_{i,j}\boldsymbol{K}_{i,j}\boldsymbol{w}_{i,j} = \boldsymbol{0}_{n_{i,j}}$. This implies $\tilde{\boldsymbol{G}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)\boldsymbol{q}_{i,j} = \boldsymbol{0}_{\tilde{z}_{i,j}}$, so that the zero restrictions are satisfied. It also implies that $\boldsymbol{q}_{i,\ell}'\boldsymbol{q}_{i,j} = 0$, for $1 \leq \ell < j$, so $\boldsymbol{q}_{i,\ell}$ and $\boldsymbol{q}_{i,j}$ are perpendicular. Because the columns of $\boldsymbol{K}_{i,j}$ are orthonormal, $\|\boldsymbol{q}_{i,j}\| = \|\boldsymbol{w}_{i,j}\| = 1$. So, Step 3 of Algorithm 1 ensures both that the zero restrictions are satisfied and that the matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are orthogonal. Finally, Step 6 ensures that the weighted draws are from the desired posterior, provided that almost all Proxy-SVAR structural parameters satisfying the zero restrictions are in the image of the mapping defined by Steps 3 and 4. This will be shown in Section 3.4.

---

[3]If the full QR-decomposition of $\boldsymbol{M}_{i,j}'$ is $\boldsymbol{M}_{i,j}' = [\hat{\boldsymbol{Q}} \ \tilde{\boldsymbol{Q}}][\boldsymbol{R}' \ \boldsymbol{0}_{n_{i,j} \times (d_i - n_{i,j})}]' = \hat{\boldsymbol{Q}}\boldsymbol{R}$, where $\hat{\boldsymbol{Q}}$ is $d_i \times n_{i,j}$, $\tilde{\boldsymbol{Q}}$ is $d_i \times (d_i - n_{i,j})$, and $\boldsymbol{R}$ is $n_{i,j} \times n_{i,j}$, then $\boldsymbol{M}_{i,j}\tilde{\boldsymbol{Q}} = \boldsymbol{R}'\hat{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}} = \boldsymbol{0}$. So, if $\boldsymbol{M}_{i,j}$ is of full row rank, then the columns of $\tilde{\boldsymbol{Q}}$ form an orthonormal basis for the null space of $\boldsymbol{M}_{i,j}$.

3.4 The Density Implied by Steps 1–4 of Algorithm 1

Step 1 of Algorithm 1 independently draws the triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ from the restricted $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ distribution. Step 2 independently draws $\boldsymbol{w}_{i,j}$ from the uniform distribution on the unit sphere in $\mathbb{R}^{d_i - n_{i,j}}$. Hence, the density over $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w})$, where $\boldsymbol{w} = (\boldsymbol{w}_{1,1}, \cdots, \boldsymbol{w}_{1,n}, \boldsymbol{w}_{2,1}, \cdots, \boldsymbol{w}_{2,k})$, is proportional to $NGN_{(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$. Step 3 defines a mapping from $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w})$ to the orthogonal triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. This mapping depends on the choice of $\boldsymbol{K}_{i,j}$ and we denote any choice of this mapping by g. Step 4 maps $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ to the Proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ using the function f. The density of the draws produced by Steps 1–4 will be the density coming from Steps 1 and 2, which is proportional to $NGN_{(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, times the volume element associated with the inverse of the mapping $f \circ g$ defined in Steps 3 and 4. The function $f \circ g$ is invertible because both f and g are one-to-one. A volume element can be thought of as a generalization of the Jacobian that appears in the usual change of variable theorem. We will use the change of variable theorem outlined in Arias, Rubio-Ramírez, and Waggoner (2018). Because we need to transform densities defined over smooth manifolds, we will use Theorem 3 of that paper, which is reproduced here as Theorem 1.

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

$$\int_A \lambda(\boldsymbol{v}) d_{\mathcal{V}} \boldsymbol{v} = \int_{\zeta^{-1}(A) \cap \mathcal{U}} \lambda(\zeta(\boldsymbol{u})) \underbrace{\left| \det(\boldsymbol{N}_{\boldsymbol{u}}' \cdot D\zeta(\boldsymbol{u})' \cdot D\zeta(\boldsymbol{u}) \cdot \boldsymbol{N}_{\boldsymbol{u}}) \right|^{\frac{1}{2}}}_{volume\ element} d_{\mathcal{U}} \boldsymbol{u},$$

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

To apply Theorem 1, several choices must be made. These choices will not affect the value of the volume element, but will affect the implementation. The vector $\boldsymbol{u} \in \mathbb{R}^b$ will be a vectorized version of the Proxy-SVAR structural parameter $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, which contains blocks of zeros. So, one could take $b = \tilde{n}^2 + \tilde{n}\tilde{m}$ and have β encode both the block and zero restrictions or take $b = \tilde{n}^2 + \tilde{n}\tilde{m} - (p+1)nk$ and have β encode only the zero restrictions. We choose the latter, which implies that $b - d = \sum_{j=1}^n \tilde{z}_{1,j}$ and β is given by $(\tilde{\boldsymbol{Z}}_1 \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \boldsymbol{e}_{n,1}, \cdots, \tilde{\boldsymbol{Z}}_n \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \boldsymbol{e}_{n,n})$. In Appendix A.4 it is shown that the derivative of β is of rank $b - d$ over all of $\mathbb{R}^b$.

The vector $\boldsymbol{v} \in \mathbb{R}^a$ will be a vectorized version of $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{1,1}, \cdots, \boldsymbol{\alpha}_{1,n}, \boldsymbol{\alpha}_{2,1}, \cdots, \boldsymbol{\alpha}_{2,k})$. As with $\boldsymbol{u}$, we choose to squeeze the zeros out of the block-triangular parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, which implies

that $a = d + n + k$. This expression comes from summing the dimensions of the $\boldsymbol{\alpha}_{i,j}$ and imposing the block-triangular restrictions. The d-dimensional smooth manifold $\mathcal{V}$ is the set of all $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{\alpha})$ such that the norm of each $\boldsymbol{\alpha}_{i,j}$ is one. The function λ is given by $\lambda(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{\alpha}) = NGN_{(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, which is proportional to the density implied by Steps 1 and 2 of Algorithm 1 when λ is restricted to $\mathcal{V}$.

All that remains is to define the open set $U \subset \mathbb{R}^b$ and the function ζ. We want $\zeta = (f \circ g)^{-1}$ and we want ζ to be continuously differentiable. The function g can be defined only if the matrices $\boldsymbol{M}_{i,j}$ are of full row rank for all i and j. Since $\boldsymbol{M}_{2,j}$ is always of full row rank, this suggests that we take $U \subset \mathbb{R}^b$ to be the set of all $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ such that $\boldsymbol{M}_{1,j}(f^{-1}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+))$ is of full row rank for all $1 \leq j \leq n$.[4] The following proposition implies that for this choice of U, the function g can be defined so that it is continuously differentiable, at least locally.

Proposition 1. *The set U is open and the complement of $U \cap \beta^{-1}(\{\boldsymbol{0}\})$ is of measure zero in $\beta^{-1}(\{\boldsymbol{0}\})$. For every $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+) \in U$, the function $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$, for $i = 1, 2$ and $1 \leq j \leq d_i$, can be defined in a neighborhood of $f^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ so that it is continuously differentiable and depends only on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ and the first $j - 1$ columns of $\boldsymbol{Q}_i$.*

Proof. See Appendix A.4. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proposition 1 ensures that the functions $\boldsymbol{K}_{i,j}$ can be defined so that they are continuously differentiable, at least locally. Thus, the function g, and hence ζ, can be defined locally so that they are continuously differentiable, which is enough to apply Theorem 1. A natural question to ask is can $\boldsymbol{K}_{i,j}$ be defined so that it is continuously differentiable, or even just continuous, over all of $f^{-1}(U)$? There are deep theorems from algebraic topology that imply that the answer is no, in general, but this does not matter for our purposes. As was noted in the discussion after Algorithm 1, we need the complement of $U \cap \beta^{-1}(\{\boldsymbol{0}\})$ to be of measure zero in $\beta^{-1}(\{\boldsymbol{0}\})$ in order for the weighted draws obtained from Algorithm 1 to be from the desired posterior. While the proof that the complement of $U \cap \beta^{-1}(\{\boldsymbol{0}\})$ is of measure zero in $\beta^{-1}(\{\boldsymbol{0}\})$ is involved and left to Appendix A.4, the local construction of the function $\boldsymbol{K}_{i,j}$ is straightforward.

If $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+) \in U$, then $\boldsymbol{M}_{i,j}(f^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+))$ is of full row rank. So, there exists a $(d_i - n_{i,j}) \times d_i$ matrix $\boldsymbol{R}_{i,j}$ such that $[\boldsymbol{M}_{i,j}(f^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+))' \;\; \boldsymbol{R}'_{i,j}]$ is non-singular. Let $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ be the last $d_i - n_{i,j}$ columns of the orthogonal component of the QR-decomposition of $\hat{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) = [\boldsymbol{M}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)' \;\; \boldsymbol{R}'_{i,j}]$, normalized so that the diagonal of the triangular component is positive. By an argument similar to the one in Section 3.3, the columns of $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ will form an orthonormal basis for the null space

---

[4]The matrices $\boldsymbol{M}_{i,j}$ and $\boldsymbol{K}_{i,j}$ depend only on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ and the first $j - 1$ columns of Q_i, but we can consider them to be functions of $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) = f^{-1}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$.

of $\boldsymbol{M}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Because $\hat{\boldsymbol{M}}_{i,j}(f^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+))$ is non-singular, $\hat{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ will be non-singular in some open set $\hat{U}$ about $f^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$. The matrix $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ can be obtained using the Gram-Schmidt orthogonalization process, and thus is continuously differentiable over $\hat{U}$.[5] Finally, the function $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ will depend only on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ and the first $j-1$ columns of $\boldsymbol{Q}_i$, because $\hat{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ depends only on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ and the first $j-1$ columns of $\boldsymbol{Q}_i$.

The matrices $\boldsymbol{R}_{i,j}$ are called local reference matrices and are a coordination device that allows one to define the $\boldsymbol{K}_{i,j}$ so that they are continuously differentiable. One might ask if it is really necessary to go to the expense of forming the local reference matrices and just use the much simpler technique described in Section 3.3. No technique can produce $\boldsymbol{K}_{i,j}$ that are continuously differentiable globally, but it is the case that most techniques will produce $\boldsymbol{K}_{i,j}$ that are continuously differentiable almost everywhere. Whether or not this holds for the technique described in Section 3.3 depends on the algorithm used to produce the full QR-decomposition. However, most linear algebra programs will produce the QR-decomposition using either Householder reflections or Given rotations, both of which will produce a continuously differentiable QR-decomposition almost everywhere. While it is true that our numeric computations of the derivatives appearing in the volume element can go awry if they are being evaluated sufficiently close to a point where one of the $\boldsymbol{K}_{i,j}$ is not continuously differentiable, experience leads us to recommend the simpler algorithm to obtain the $\boldsymbol{K}_{i,j}$.

Proposition 1 implies that Theorem 1 can be applied and the density of the draws given by Steps 1–4 of Algorithm 1 is

$$p(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \propto NGN_{(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+) \underbrace{\left| \det(\boldsymbol{N}'_{(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)} \cdot D\zeta(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)' \cdot D\zeta(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) \cdot \boldsymbol{N}_{(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)}) \right|^{\frac{1}{2}}}_{\text{volume element}}, \quad (8)$$

where $f^{-1}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) = (\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. The derivatives of ζ and β can be computed numerically as in Arias, Rubio-Ramírez, and Waggoner (2018). Equation (8) does not depend on which orthonormal basis for the null space of the derivative of β is chosen. So, up to a multiplicative constant, the density p can be explicitly computed, at least numerically. Note that the zero restrictions affect the density because the volume element depends on the derivative of β, which is not constant, but the sign restrictions only affect the density up to a multiplicative constant because it restricts the parameters to an open subset of positive measure of the parameter space.

---

[5]The Gram-Schmidt orthogonalization process applied to a non-singular matrix can be explicitly expressed using the operators $+$, $-$, $\times$, and $\div$ and division by zero will not occur, so it is continuously differentiable. Because the QR-decomposition of a non-singular matrix, normalized so that the diagonal of the triangular component is positive, is unique, any algorithm producing the QR-decomposition can be used, so long as the diagonal of the triangular component is normalized to be positive.

3.5 Practical Considerations, Extensions, and Limitations of Algorithm 1

In this section, we highlight some practicalities when implementing Algorithm 1. We also emphasize some easy extensions of the algorithm.

3.5.1 Effective Sample Size

Importance samplers generate weighted draws. If all of the weights were equal, then we would actually have unweighted draws and the effective sample size would be the actual sample size. However, the weights from Algorithm 1 are not equal, so it is critical that one computes the effective sample size. The effective sample size is

$$\frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2},$$

where w_i is the importance weight associated with the ith draw and N is the total number of retained draws. It is also useful to keep track of the percentage of the total number of draws that were retained in Step 5 to get a sense of how restrictive the sign restrictions are, but the effective sample size is the key statistic for this sampler. If one chooses to re-sample in order to have unweighted draws, the number of re-sampled draws should never exceed the effective sample size, and some would argue that the number re-sampled should be much smaller than the effective size. In all our applications our effective sample size is 10,000 draws.

The reliable use of the importance sampler requires the importance weights to possess finite variance. We use the tests proposed by Koopman, Shephard, and Creal (2009) as described in Appendix A.6. In Appendices A.7–A.10 we show that these tests imply that the finite variance requirement holds for the applications analyzed in this paper.

3.5.2 Efficiency and Approximating the Derivatives

It is also important to note that computing the volume element in Step 6 is the most expensive part in implementing Algorithm 1 because of computing the derivative $D\zeta(\boldsymbol{u})$. The rest of Algorithm 1 is quite fast. Both the domain and the range of ζ are of fairly high dimension in practice, and so $D\zeta(\boldsymbol{u})$ is a fairly large matrix and numerically computing it requires many evaluations of the function ζ. The jth column of $D\zeta(\boldsymbol{u})$ can be approximated as either $(\zeta(\boldsymbol{u} + \epsilon\boldsymbol{e}_{b,j}) - \zeta(\boldsymbol{u}))/\epsilon$ or $(\zeta(\boldsymbol{u} + \epsilon\boldsymbol{e}_{b,j}) - \zeta(\boldsymbol{u} - \epsilon\boldsymbol{e}_{b,j}))/(2\epsilon)$, which are the one-sided and two-sided approximations. The two-sided approximation is usually more accurate, but requires almost twice as many function evaluations. For this reason, we generally prefer the one-sided approximation, but we also recommend that a test run be employed to see if there is a difference between the two approximations.

If so, then we would recommend the two-sided approximation. Also, we recommend choosing ϵ to be 10^{-6}. Because ζ is a complicated function, we do not recommend choosing ϵ to be smaller than 10^{-7}, though values as large as 10^{-4} can give good approximations in some applications. Similar advice is given for approximating $D\beta(\boldsymbol{u})$.

3.5.3 Normalization

The reader should note that Algorithm 1 should be used with a normalization to determine the sign of each equation. This is because if we change the sign of the jth column of $\tilde{\boldsymbol{A}}_0$ and $\tilde{\boldsymbol{A}}_+$, the zero restrictions and γ-relevance condition will still hold. Furthermore, the resulting two sets of parameters are observationally equivalent and will have the same posterior value. A normalization will eliminate one of these two sets of parameters. Typically, SVARs are normalized by restricting the sign of the contemporaneous response of a given variable to a shock of interest. Proxy-SVARs can be normalized analogously. Since this is well understood and one simply has to discard the draw when the normalization is not satisfied, we do not explicitly state this in the algorithm. If there is only one sign restriction of this type for the jth shock, then instead of discarding the draw, one could change the sign of the jth column of $\tilde{\boldsymbol{A}}_0$ and $\tilde{\boldsymbol{A}}_+$. As $\tilde{n}$ becomes large, this can result in significant efficiency gains.

3.5.4 Drawing from Other Parameterizations or Posterior Distributions

Algorithm 1 is stated in terms of the Proxy-SVAR structural parameterization, but it will work for any parameterization as long as one can explicitly compute the transformation between that parameterization and the orthogonal triangular-block parameterization. Similarly, Algorithm 1 independently draws from the restricted $NGN(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$ posterior distribution over the Proxy-SVAR structural parameterization conditional on the zero and sign restrictions. As mentioned several times already, the algorithm can be modified to independently draw from any desired restricted posterior distribution. When that is the case, one will need to modify Step 6 in Algorithm 1 to include the density associated with the desired restricted posterior distribution instead of $NGN_{(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$. The rest of the steps will not change. In either of these cases, there is no natural choice for the hyper-parameters $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ needed in Step 1 of Algorithm 1 and one will have to use the techniques outlined in Appendix A.3 to choose values that will not lead to unreasonably small effective sample sizes.

In many applications, the Proxy-SVAR model will only be set-identified. As Giacomini, Kitagawa, and Read (2020) and Giacomini and Kitagawa (2018) highlight, Bayesian analysis of set-identified models may

not be robust to small changes in the prior. Because our techniques can be applied to any prior, some of the techniques advocated in Giacomini and Kitagawa (2018) could be applied if needed.

3.5.5 Identifying More Shocks Than Proxies

Our algorithm can handle cases in which a researcher wants to consider k instruments that are correlated with $\tilde{k}$ shocks, with $\tilde{k} \geq k$. In such cases, Equation (4) will only hold for $1 < j < n - \tilde{k}$. This could be of interest, for example, when a researcher assumes that a proxy is not correlated with a particular structural shock while leaving the correlation with the remaining structural shocks unrestricted. As with the case that $k = \tilde{k}$, the proxies only divide the shocks into two groups: those that are correlated with the proxies and those that are uncorrelated with the proxies. To identify the shocks within each of these groups, additional zero and sign restrictions would be required.

The intuition of why our approach works for k instruments that are correlated with $\tilde{k}$ structural shocks, with $\tilde{k} > k$, is as follows. As more structural shocks than instruments are considered, fewer zero restrictions need to be imposed, since the instruments are correlated with more structural shocks. It is the case that in this scenario, the set of Proxy-SVAR structural parameters that satisfy the exogeneity restrictions and the γ-relevance condition is larger.

To conclude this section, let's clarify that our methodology cannot be used to impose the restriction that every proxy is only correlated with one structural shock, at least when $k > 1$. This requires imposing a diagonal structure in the $k \times k$ matrix $\boldsymbol{V}$ using additional zero restrictions, which is outside the scope of our algorithm given the constraints embedded in the size of $\tilde{z}_{1,j}$ for $n - k + 1 \leq j \leq n$.

3.6 The Importance of the Volume Element

Given that the main expense of our algorithm is computing the importance weights, one might be tempted to dispense with Step 6 of Algorithm 1 and simply use the unweighted draws. Of course, the unweighted draws are not from the desired posterior distribution, but if the weights do not vary too much, then the draws would be approximately from the desired posterior distribution. The reader should be aware of at least one dangerous feature of the distribution over the structural parameterization implied by this strategy. The distribution is not invariant to a reordering of the instruments, and hence it is invalid for inference.

To illustrate that Algorithm 1 without weighting is not invariant to a reordering of the instruments, let us consider a Proxy-SVAR with three variables, no lags, no constant, and two external instruments; so $n = 3$, $k = 2$, and $\tilde{\boldsymbol{A}}_0$ is the only Proxy-SVAR structural parameter. We impose only the exogeneity restrictions. The

following matrices will be of use in expressing a different ordering of the instruments,

$$\boldsymbol{P}_m = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n \times k} \\ \boldsymbol{0}_{k \times n} & \boldsymbol{P}_m \end{bmatrix}.$$

There are two ways to order the external instruments, $\tilde{\boldsymbol{y}}'_{1,t} = [\boldsymbol{y}'_t, \boldsymbol{m}'_t]$ and $\tilde{\boldsymbol{y}}'_{2,t} = \tilde{\boldsymbol{y}}'_{1,t} \tilde{\boldsymbol{P}}$. If $\tilde{\boldsymbol{A}}_0$ satisfies the exogeneity restrictions under the first ordering, then $\tilde{\boldsymbol{P}}' \tilde{\boldsymbol{A}}_0$ satisfies the exogeneity restrictions under the second ordering. Because $\tilde{\boldsymbol{A}}_+$ does not appear, there are only two hyper-parameters controlling the prior and posterior distributions. If the hyper-parameters for prior and posterior densities under the first ordering were $(\bar{\nu}, \bar{\boldsymbol{\Phi}})$ and $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}})$, then the equivalent prior and posterior hyper-parameters under the second ordering would be $(\bar{\nu}, (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}') \bar{\boldsymbol{\Phi}} (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}))$ and $(\tilde{\nu}, (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}') \tilde{\boldsymbol{\Phi}} (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}))$. With these priors and posteriors, it is easy to see that the prior and posterior density of $\tilde{\boldsymbol{A}}_0$ under the first ordering and $\tilde{\boldsymbol{P}}' \tilde{\boldsymbol{A}}_0$ under the second ordering are equal.

One might speculate that if the hyper-parameters controlling the draws in Step 1 of Algorithm 1 were $(\hat{\nu}, \hat{\boldsymbol{\Phi}})$ under the first ordering and were $(\hat{\nu}, (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}') \hat{\boldsymbol{\Phi}} (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{P}}))$ under the second ordering, then the unweighted density of $\tilde{\boldsymbol{A}}_0$ under the first ordering and $\tilde{\boldsymbol{P}}' \tilde{\boldsymbol{A}}_0$ under the second ordering might be equal. However, this is not true. To see this, we make 10 draws of $\tilde{\boldsymbol{A}}_0$ from the first ordering and compute the density kernel. Then we compute the density kernel under the second ordering of $\tilde{\boldsymbol{P}}' \tilde{\boldsymbol{A}}_0$. If our speculation is correct, the ratio of these two density kernels would be constant. The results are reported in Table 1. If the order of the instruments did not matter, then all the entries in the table would be equal, which is clearly not the case. For simplicity, we set the hyper-parameters for the first ordering to be $(\tilde{\nu}, \boldsymbol{I}_{n^2})$. This is a disturbing result because it means that

Table 1: Ratio of densities

Draw	1	2	3	4	5	6	7	8	9	10
	0.64	0.10	0.07	0.71	1.93	0.04	0.34	0.27	0.01	4.57

Ratio of densities for ten draws of the structural parameters using a different ordering for the instruments.

inference based on Algorithm 1 without Step 1 hinges on an arbitrary decision regarding the order in which one sets the instruments in $\boldsymbol{m}_t$. A similar argument can be made regarding the order of the variables within $\boldsymbol{y}_t$. These results point out that without the importance sampling step, one cannot control the distribution implied by Algorithm 1.

4 Application I: The Dynamic Effects of TFP Shocks

In this section we illustrate our methodology by studying the dynamic effects of two types of TFP shocks, a consumption TFP shock and an investment TFP shock, in a quarterly frequency Proxy-SVAR featuring five endogenous variables and two proxies for the structural shocks of interest. More specifically, we adopt the specification of the SVAR and the proxies from Lunsford (2016). Accordingly, the endogenous variables are real GDP growth, employment growth, inflation, real consumption growth, and real investment in equipment growth. The remaining details on the data are provided in Appendix A.5. The proxies are a consumption TFP proxy and an investment TFP proxy based on Fernald's (2014) consumption and investment TFP series, respectively. In particular, we use Lunsford's (2016) proxies, which are obtained by regressing each of the TFP series just mentioned on four lags of the endogenous variables and by labeling the residuals associated with each of these regressions as consumption and investment TFP proxies, respectively.[6]

The Proxy-SVAR features four lags and a constant, and the sample runs from 1947Q2 until 2015Q4. Consequently, in this application $T = 275$, $n = 5$, $k = 2$, $p = 4$, and $\tilde{m} = p\tilde{n} + 1$. We set $\nu = \tilde{n} = 7$, $\boldsymbol{\Phi}^{-1} = \mathbf{0}_{\tilde{n},\tilde{n}}$, and $\boldsymbol{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \tilde{\boldsymbol{\Omega}}^{-1}$ to characterize our proposal over the orthogonal triangular-block parameterization. We also set $\gamma = 0.2$. Clearly, we could consider γ as a hyper-parameter and define a prior over it. Our approach, which is equivalent to having a dogmatic prior over γ, was chosen for simplicity, but can be easily extended to a more general prior.

Let $\boldsymbol{\varepsilon}_t^{TFP}$ be a vector containing the consumption and investment TFP shocks, i.e., $\boldsymbol{\varepsilon}_t^{TFP} = [\varepsilon_{C,t} \quad \varepsilon_{I,t}]'$, and let $\boldsymbol{\varepsilon}_t^O$ be a vector containing all other structural shocks. The exogeneity restrictions and the relevance condition are

$$\mathbb{E}\left[\boldsymbol{m}_t \boldsymbol{\varepsilon}_t^{O\prime}\right] = \mathbf{0}_{2\times 3} \text{ and } \mathbb{E}\left[\boldsymbol{m}_t \boldsymbol{\varepsilon}_t^{TFP\prime}\right] = \boldsymbol{V} \neq \mathbf{0}_{2\times 2},$$

where $\boldsymbol{m}_t' = [\boldsymbol{m}_{C,t} \quad \boldsymbol{m}_{I,t}]$ are the proxies for the consumption and investment TFP shocks. As mentioned in Section 2.3, without additional restrictions, these conditions are not enough to distinguish a consumption TFP shock from an investment TFP shock. As a consequence we also impose the additional sign restrictions $\mathbb{E}[\boldsymbol{m}_{C,t}\varepsilon_{C,t}] > 0$, $\mathbb{E}[\boldsymbol{m}_{I,t}\varepsilon_{I,t}] > 0$, $\mathbb{E}[\boldsymbol{m}_{C,t}\varepsilon_{C,t}] > \mathbb{E}[\boldsymbol{m}_{C,t}\varepsilon_{I,t}]$ and $\mathbb{E}[\boldsymbol{m}_{I,t}\varepsilon_{I,t}] > \mathbb{E}[\boldsymbol{m}_{I,t}\varepsilon_{C,t}]$ on the entries of $\boldsymbol{V}$. These sign restrictions make sense because the structural shocks are standardized to have unit variance. If

---

[6]We downloaded the proxies from Kurt Lunsford's website at https://sites.google.com/site/kurtglunsford/research.
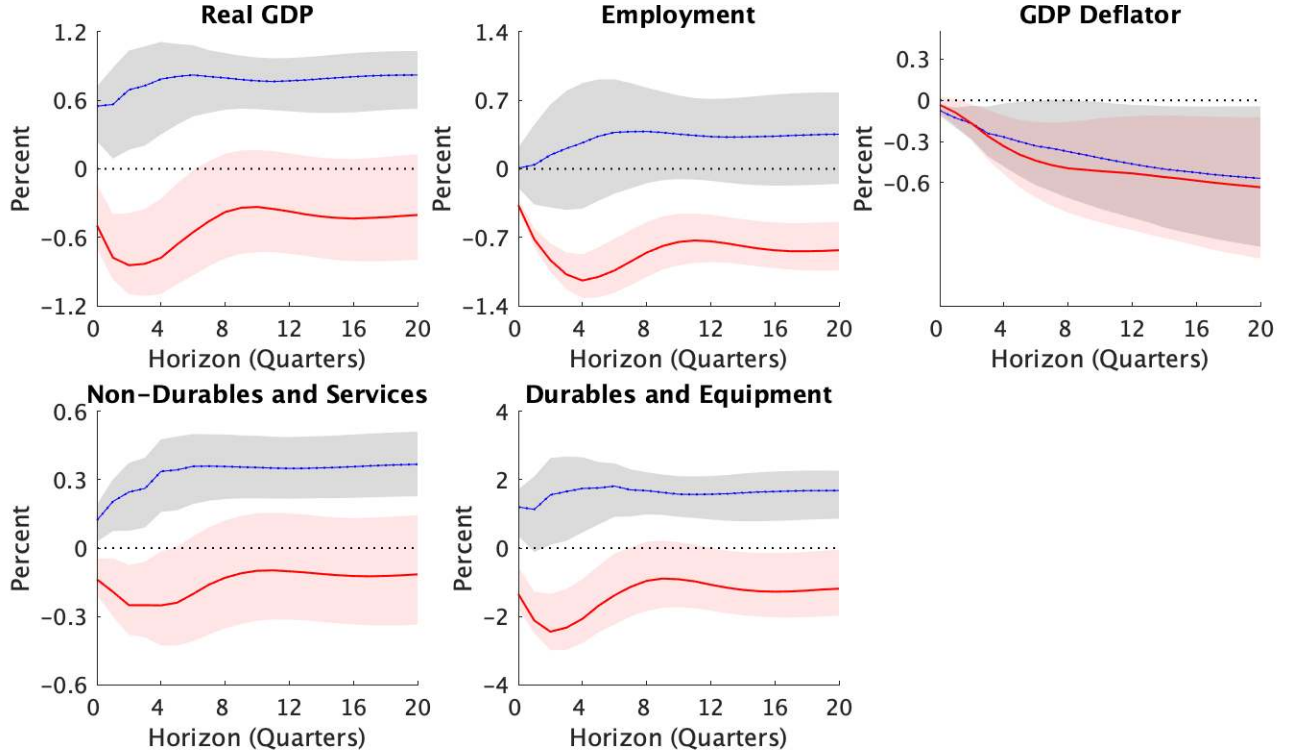
Figure 1: IRFs to positive one standard deviation consumption and investment TFP shocks. The blue solid-dotted curves represent the point-wise posterior medians and the gray shaded areas represent the 68 percent equal-tailed point-wise probability bands to a consumption TFP shock. The red solid curves represent the point-wise posterior medians and the red shaded areas represent the 68 percent equal-tailed point-wise probability bands to an investment TFP shock.

we order the two structural shocks of interest last, this implies setting $s = 4$,

$$\boldsymbol{S}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \; \boldsymbol{S}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

and

$$\tilde{\boldsymbol{F}}_s(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) = \begin{bmatrix} \boldsymbol{e}'_{2,1} \boldsymbol{S}_4 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},4} \\ \boldsymbol{e}'_{2,1} \boldsymbol{S}_4 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},4} - \boldsymbol{e}'_{2,2} \boldsymbol{S}_5 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},5} \\ \boldsymbol{e}'_{2,1} \boldsymbol{S}_5 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},5} \\ \boldsymbol{e}'_{2,1} \boldsymbol{S}_5 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},5} - \boldsymbol{e}'_{2,2} \boldsymbol{S}_4 (\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{e}_{\tilde{n},4} \end{bmatrix},$$

where, for ease of exposition, we have abstracted from explicitly stating the γ-relevance condition.

Figure 1 shows the IRFs to positive one standard deviation consumption and investment TFP shocks.[7] The blue solid-dotted curves represent the point-wise posterior medians and the gray shaded areas represent the 68 percent equal-tailed point-wise probability bands to a consumption TFP shock. The red solid curves

---

[7] While Lunsford (2016) reports the IRFs of the endogenous variables in the SVAR, we report the cumulative IRFs.

represent the point-wise posterior medians and the red shaded areas represent the 68 percent equal-tailed point-wise probability bands to an investment TFP shock. These IRFs are qualitatively consistent with the results reported by Lunsford (2016).

In particular, a consumption TFP shock causes an increase in real GDP, consumption in non-durables and services, consumption in durables and equipment, and employment while the price level gradually decreases. Although the probability bands associated with the latter variable contain zero, the findings are in line with those reported by Lunsford (2016). Accordingly, a consumption TFP shock implies opposite movements in quantities and prices, supporting the conventional wisdom about the effects of standard TFP shocks. In contrast, a positive investment TFP shock leads with high probability to a decrease in real GDP, employment, consumption, and the price level. As highlighted by Lunsford (2016), these results are inconsistent with the conventional wisdom of standard TFP shocks but in line with the findings in Liu, Fernald, and Basu (2012).

5   Application II: The Dynamic Effects of Personal Income Tax Shocks

In this section we use our methodology to revisit a recent study by Mertens and Montiel-Olea (2018) presenting new time series evidence—based on Proxy-SVARs—on the effects of personal income tax rate cuts on reported pre-tax income and other indicators of real activity such as GDP and the unemployment rate. Their three main reported findings can be summarized as follows. First, negative average marginal tax rate (AMTR) shocks lead not only to increases in real GDP and declines in the unemployment rate but also to increases in reported income. Second, they find that substitution effects rather than income effects are important for the transmission of personal income tax policy changes in the U.S. economy post-World War II. Third, the dynamic effects of tax reforms depend on how different income groups are affected by the reforms. One important point of Mertens and Montiel-Olea's (2018) analysis is that they use counterfactual experiments. A counterfactual experiment is a linear combination of structural shocks that imposes a particular dynamic relation between some endogenous variables.

Our approach will basically replicate Mertens and Montiel-Olea's (2018) first and third findings. When analyzing their second finding, we will show that their conclusions depend on the particular counterfactual tax experiments that they conduct to assess the effectiveness of changes in marginal relative to average tax rates. While counterfactual experiments could be potentially useful and have been used by other SVAR-based studies of fiscal policy (e.g., Mountford and Uhlig, 2009; Ramey, 2013; Mertens and Ravn, 2013), they frequently hinge on imposing certain relations between endogenous variables that some researchers could find questionable. To address this issue, we separately identify structural shocks using additional sign restrictions.

5.1 Macroeconomic Responses to Marginal Tax Rates

In their benchmark specification, Mertens and Montiel-Olea (2018) use yearly data from 1946 through 2012 to estimate a Proxy-SVAR including nine endogenous variables, two exogenous variables, and one proxy for AMTR shocks.[8] The endogenous variables are the negative of log net-of-tax rate, log reported income, log real GDP per tax unit, the unemployment rate, the log real stock market index, inflation, the federal funds rate, log real government spending per tax unit, and the change in log real federal government debt per tax unit. Net-of-tax rate is defined as 1 minus the AMTR. The exogenous variables are dummy variables for the years 1949 and 2008. The proxy (which we call the AMTR proxy) is a collection of instances of variation in marginal tax rates that the authors reasonably consider to be contemporaneously exogenous changes in the AMTR. The net-of-tax rate is based on Barro and Redlick (2011). Accordingly, the identification of the AMTR shock is achieved by assuming that the proxy is only correlated with the AMTR. The SVAR features two lags and a constant term. Altogether, in this application $T = 65$, $n = 9$, $k = 1$, $p = 2$, $\tilde{e} = 2$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$.
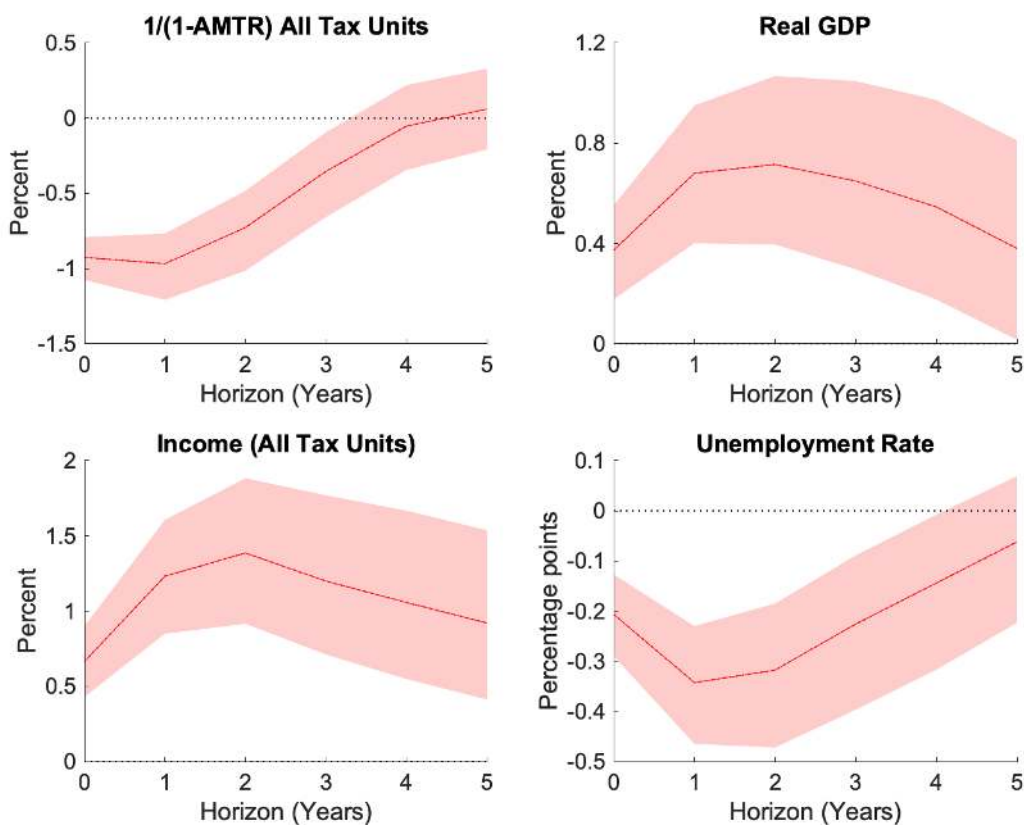
Figure 2: IRFs to positive AMTR shock (rate cut). The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent equal-tailed point-wise probability bands.

---

[8]Link to the dataset: https://karelmertenscom.files.wordpress.com/2018/01/data_mmo.xlsx.

We set $\nu = \tilde{n}$, $\boldsymbol{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\boldsymbol{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\boldsymbol{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \tilde{\boldsymbol{\Omega}}^{-1}$ to characterize our proposal over the orthogonal triangular-block parameterization. We also set $\gamma = 0.2$.

Figure 2 shows the point-wise median and the 68 percent equal-tailed point-wise probability bands for the IRFs of the key variables of interest to a positive one standard deviation AMTR shock. Clearly, the positive and sizable IRFs of real GDP and the negative and sizable IRFs of the unemployment rate coincide with a positive and sizable response of income. Therefore, our results clearly align with those reported in Figure 5 of Mertens and Montiel-Olea (2018).

5.2 Average versus Marginal Tax Rates

To assess whether tax policy mainly operates through direct effects on individual incentives, Mertens and Montiel-Olea (2018) expand the SVAR used in Section 5.1 by adding the log ATR as an endogenous variable and they jointly identify two personal income tax rate shocks using two proxies. ATR is defined as total revenue and contributions as a ratio of the Piketty and Saez (2003) measure of aggregate market income. Analogously to the case of the AMTR proxy, the new proxy (which we call the ATR proxy) is a collection of instances of variation in ATRs that the authors reasonably consider to be contemporaneously exogenous changes in the ATR. The identification of the AMTR and ATR shocks is achieved assuming that the proxies are only correlated with the tax rate shocks and using two counterfactual tax experiments to establish causal effects. In the AMTR counterfactual, they consider an unanticipated change in the marginal tax rate that does not have a direct effect on the average tax rate. In the ATR counterfactual, they consider an unanticipated change in the average tax rate that does not have a direct effect on the marginal tax rate.

In this case $T = 65$, $n = 10$, $k = 2$, $p = 2$, $\tilde{e} = 2$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$. We set $\nu = \tilde{n}$, $\boldsymbol{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\boldsymbol{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\boldsymbol{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Phi}}_0 \neq \tilde{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \tilde{\boldsymbol{\Omega}}^{-1}$ to characterize our proposal over the orthogonal triangular-block parameterization. We choose $\hat{\boldsymbol{\Phi}}_0$ to maximize the efficiency of the importance sampler. If we set $\hat{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}$ the algorithm becomes very inefficient. The basic description of the approach used for the selection of $\hat{\boldsymbol{\Phi}}_0$ is described in Appendix A.3. We also set $\gamma = 0.2$.

Figure 3a shows the point-wise median and the 68 percent equal-tailed point-wise probability bands for the IRFs of the key variables of interest to the AMTR and ATR counterfactuals; i.e., we are just replicating their analysis using a Bayesian approach. Essentially, Figure 3a shows that Mertens and Montiel-Olea's (2018) findings regarding the effects of their counterfactuals can be supported by our approach. As
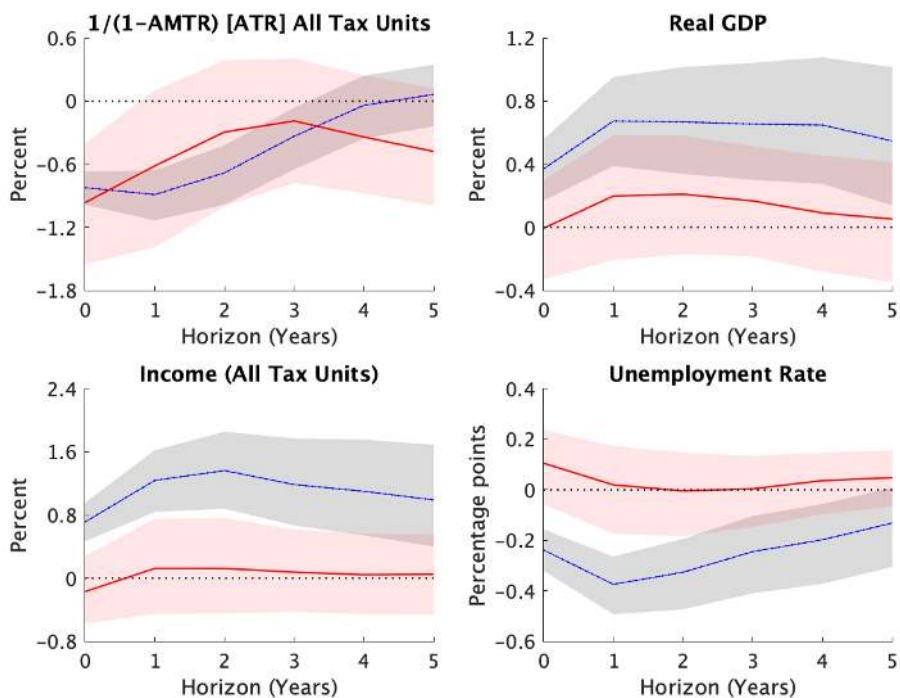
(a) Mertens and Montiel-Olea (2018)
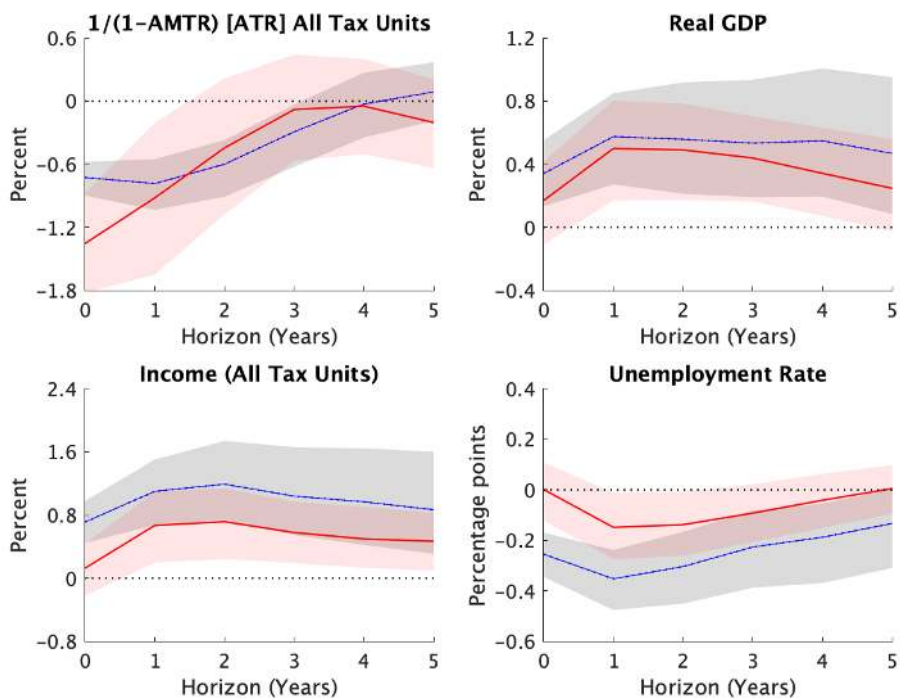
(b) Sign Restrictions

Figure 3: Panel (a): IRFs to counterfactuals. The solid curves (blue for the AMTR policy counterfactual and red for the ATR policy counterfactual) represent the point-wise posterior medians, and the shaded areas (gray for the AMTR policy counterfactual and red for the ATR policy counterfactual) represent the 68 percent equal-tailed point-wise probability bands. The IRFs are with respect to a one standard deviation counterfactual. Panel (b): IRFs to structural shocks. The solid curves (blue for the AMTR shock and red for the ATR shock) represent the point-wise posterior medians, and the shaded areas (gray for the AMTR shock and red for the ATR shock) represent the 68 percent equal-tailed point-wise probability bands. The IRFs are with respect to a one standard deviation shock.

the reader can see, the panel closely resembles the IRFs reported in Panels (B) and (C) of *Figure 10* of Mertens and Montiel-Olea (2018). Overall, this figure justifies the following claims: "*On the other hand, there is no evidence for any effect on incomes when average tax rates decline but marginal rates do not*" (Mertens and Montiel-Olea, 2018, page 1805) and "*The main finding is that, in sharp contrast to the results for marginal tax rate changes after controlling for average tax rates, there is no evidence that income responds strongly to average tax rate changes once marginal rate changes are controlled for. The point estimates are in fact slightly negative, although they are not statistically significant at any horizon*" (Mertens and Montiel-Olea, 2018, page 1860).

Researchers familiar with SVAR analysis may instead want to report the causal effects of each of the two structural shocks, which could naturally be labeled marginal tax rate shock and average tax rate shock. For this reason, we now complement Mertens and Montiel-Olea's (2018) results by identifying a marginal tax rate shock and an average tax rate shock using a set of sign restrictions. The results obtained using such an approach suggest caution while reading Mertens and Montiel-Olea's (2018) findings. But, before discussing them in more detail, let us describe our sign restrictions.

Sign Restrictions for Identifying AMTR vs ATR Shocks. *(i) The proxy for the AMTR shock is positively correlated with the AMTR shocks; (ii) the proxy for the ATR shock is positively correlated with the ATR shocks; (iii) the covariance between the AMTR shock and the AMTR proxy is bigger than the covariance between the ATR shock and the AMTR proxy; and (iv) the covariance between the ATR shock and the ATR proxy is bigger than the covariance between the AMTR shock and the ATR proxy.*

The implementation of our sign restrictions needs a function $\tilde{\boldsymbol{F}}_s$ and matrices $\boldsymbol{S}_j$ very similar to the ones described in Section 4. In Figure 3b the 68 percent equal-tailed point-wise probability bands for the IRF of income are significantly above zero for both AMTR and ATR shocks. The 68 percent equal-tailed point-wise probability bands for the IRF of real GDP to both structural shocks are also positive and similar. Turning to the unemployment rate, the IRFs to both structural shocks are broadly similar and mostly negative. The differences between the results reported in Figures 3a and 3b are confirmed when analyzing Table 2.

This table shows the short-run elasticities of income and real GDP to AMTR and ATR. In the case of Mertens and Montiel-Olea (2018), the short-run elasticities are measured by the ratio between the IRF of income (real GDP) one period after the tax cut counterfactual and the impact IRF of the AMTR (ATR) to an AMTR (ATR) tax cut counterfactual. In the case of our approach, the short-run elasticities are measured by the ratio between the IRF of income (real GDP) one period after the corresponding shock—that is the AMTR (ATR) shock when computing the elasticity with respect to the AMTR (ATR)—and the impact IRF of the

Table 2: Short-run elasticities of income (Inc) and real GDP to tax rates

Mertens and Montiel-Olea (2018)				
Ratio of IRFs	$\text{Inc}_{t+1}/\text{AMTR}_t$	$\text{Inc}_{t+1}/\text{ATR}_t$	$\text{GDP}_{t+1}/\text{AMTR}_t$	$\text{GDP}_{t+1}/\text{ATR}_t$
Median	1.51	0.12	0.82	0.20
68% Prob. Interval	[1.00; 2.15]	[-0.47; 1.05]	[0.47; 1.25]	[-0.19; 0.81]
Sign Restrictions				
Ratio of IRFs	$\text{Inc}_{t+1}/\text{AMTR}_t$	$\text{Inc}_{t+1}/\text{ATR}_t$	$\text{GDP}_{t+1}/\text{AMTR}_t$	$\text{GDP}_{t+1}/\text{ATR}_t$
Median	1.54	0.51	0.79	0.38
68% Prob. Interval	[0.94; 2.35]	[0.15; 1.09]	[0.39; 1.32]	[0.13; 0.80]

NOTE: Panel (a) The entries in the table denote the posterior moments of the ratio between the IRF of income (Inc) and real GDP one period after the start of the AMTR (ATR) counterfactual and the IRF of the AMTR (ATR) on impact following the AMTR (ATR) counterfactual. Panel (b): The entries in the table denote the posterior moments of the ratio between the IRF of income (Inc) and real GDP one period after the shock and the IRF of the AMTR and ATR on impact following an AMTR and ATR shock, respectively. See the main text for details. The table is based on the draws used in Figure 3.

AMTR (ATR) to an AMTR (ATR) shock. The reader can see that, when we use the counterfactuals, the 68 percent posterior probability intervals for the short-run elasticities of income and real GDP to ATR include negative numbers and that the posterior median is quite low when compared to the AMTR case. That is not the case when we use the set of sign restrictions instead.

Comparing Figures 3a and 3b and reading the results in Table 2, it becomes clear that definitive claims such as "*There is, on the other hand, no evidence for any effect on incomes when ATRs decline but marginal rates do not*" (Mertens and Montiel-Olea, 2018, page 1805) or "*there is no evidence that income responds strongly to ATR changes once marginal rate changes are controlled for*" (Mertens and Montiel-Olea, 2018, page 1860) are not robust to individually identifying the structural shocks underlying the counterfactuals. It is true that there may be other restrictions consistent with the results in Mertens and Montiel-Olea (2018). Nevertheless, our results show that there is no categorical evidence to rule out the income effects of exogenous changes in average tax cut rates.

5.3 Marginal Rate Cuts for the Top and Bottom of the Income Distribution

To assess whether the effects of tax reforms depend on how different income groups are affected by the reforms, Mertens and Montiel-Olea (2018) modify the SVAR used in Section 5.1 to include disaggregated measures of AMTRs and reported income, and they jointly identify two marginal personal income tax rate shocks using two proxies. More specifically, they replace the negative of the aggregate log net-of-tax rate with the negative

of the log net-of-tax rate for the top 1 percent and bottom 99 percent of the income distribution, and the aggregate log income level with the log income levels for the top 1 percent and bottom 99 percent of the income distribution. In addition, they modify the reduced-form specification by including a linear and a quadratic trend to capture longer trends in income inequality following Saez (2004) and Saez, Slemrod, and Giertz (2012). The proxies are two newly built disaggregated measures of exogenous variation in the AMTR for taxpayers at the top 1 percent of the income distribution and in the AMTR for taxpayers at the bottom 99 percent of the income distribution.

As was the case in Section 5.2, Mertens and Montiel-Olea (2018) do not aim to separately identify the two underlying tax rate shocks and instead they rely on two counterfactual tax experiments to establish causal effects. In the the top 1 percent counterfactual, they consider an unanticipated change in the marginal tax rate for taxpayers at the top 1 percent that does not have a direct effect on the marginal tax rate for taxpayers at the bottom 99 percent. In the bottom 99 percent counterfactual, they consider an unanticipated change in the marginal tax rate for taxpayers at the bottom 99 percent that does not have a direct effect on the marginal tax rate for taxpayers at the top 1 percent.

In this case $T = 65$, $n = 11$, $k = 2$, $p = 2$, $\tilde{e} = 4$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$. We set $\nu = \tilde{n}$, $\boldsymbol{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\boldsymbol{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\boldsymbol{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Phi}}_0 \neq \tilde{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \tilde{\boldsymbol{\Omega}}^{-1}$ to characterize our proposal over the orthogonal triangular-block parameterization. We choose $\hat{\boldsymbol{\Phi}}_0$ to maximize the efficiency of the importance sampler. If we set $\hat{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}$ the algorithm becomes very inefficient. The basic description of the approach used for the selection of $\hat{\boldsymbol{\Phi}}_0$ is described in Appendix A.3. We also set $\gamma = 0.2$.

Figure 4 shows the point-wise median and the 68 percent equal-tailed point-wise probability bands for the IRFs of the key variables of interest to the top 1 percent and bottom 99 percent counterfactuals. Essentially, this figure replicates Mertens and Montiel-Olea (2018) and it is very easy to conclude that there are strong positive short-run effects on income, GDP, and the unemployment rate. In contrast, the IRFs associated with the bottom 99 percent counterfactual contain much more uncertainty than the IRFs reported in *Figure XII* of Mertens and Montiel-Olea (2018) and it becomes evident that our approach gives less support for claims such as "*Bottom 99% incomes show approximately no response in the short run but increase only from the second year after the cut onwards*" and "*The timing of GDP and unemployment responses is similar to the reaction of bottom 99% incomes and shows a substantial delay relative to the more immediate effects estimated for the top 1% cut in the Figure XII*" (Mertens and Montiel-Olea, 2018, page 1865). Nevertheless, next we will show that the wide uncertainty surrounding the effects of exogenous tax cuts to the bottom 99 percent vanishes
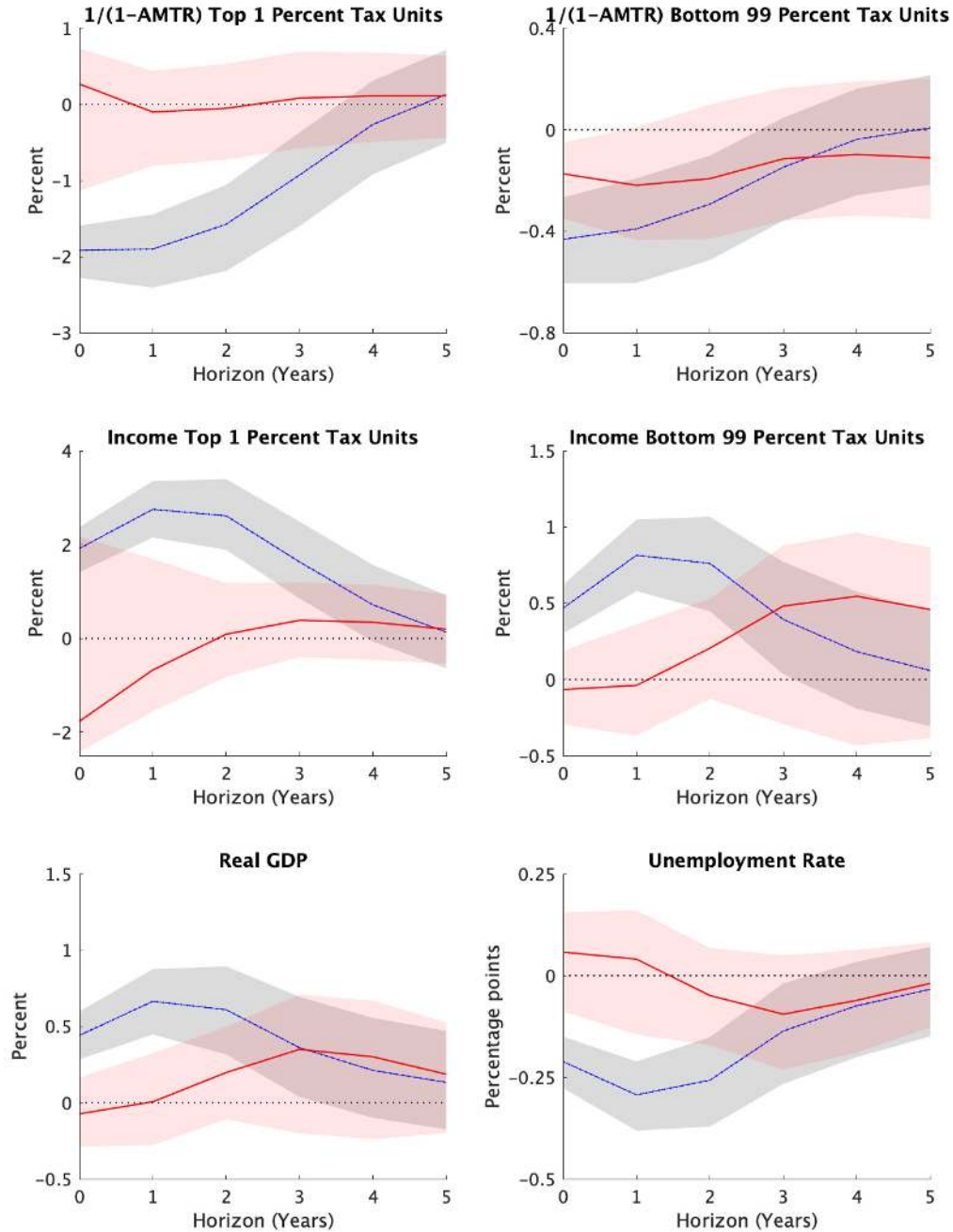
Figure 4: IRFs to counterfactuals. The solid curves (blue for the tax cut for the top 1 percent policy counterfactual and red for the tax cut for the bottom 99 percent policy counterfactual) represent the point-wise posterior medians, and the shaded areas (gray for the tax cut for the top 1 percent policy counterfactual and red for the tax cut for the bottom 99 percent policy counterfactual) represent the 68 percent equal-tailed point-wise probability bands. The IRFs are with respect to a one standard deviation counterfactual.

once we focus on the identification of fundamental structural shocks as done in Section 5.1.

In particular, we use a set of sign restrictions to identify the top 1 percent and bottom 99 percent AMTR

shocks analogous to the one used to study AMTR and ATR shocks.
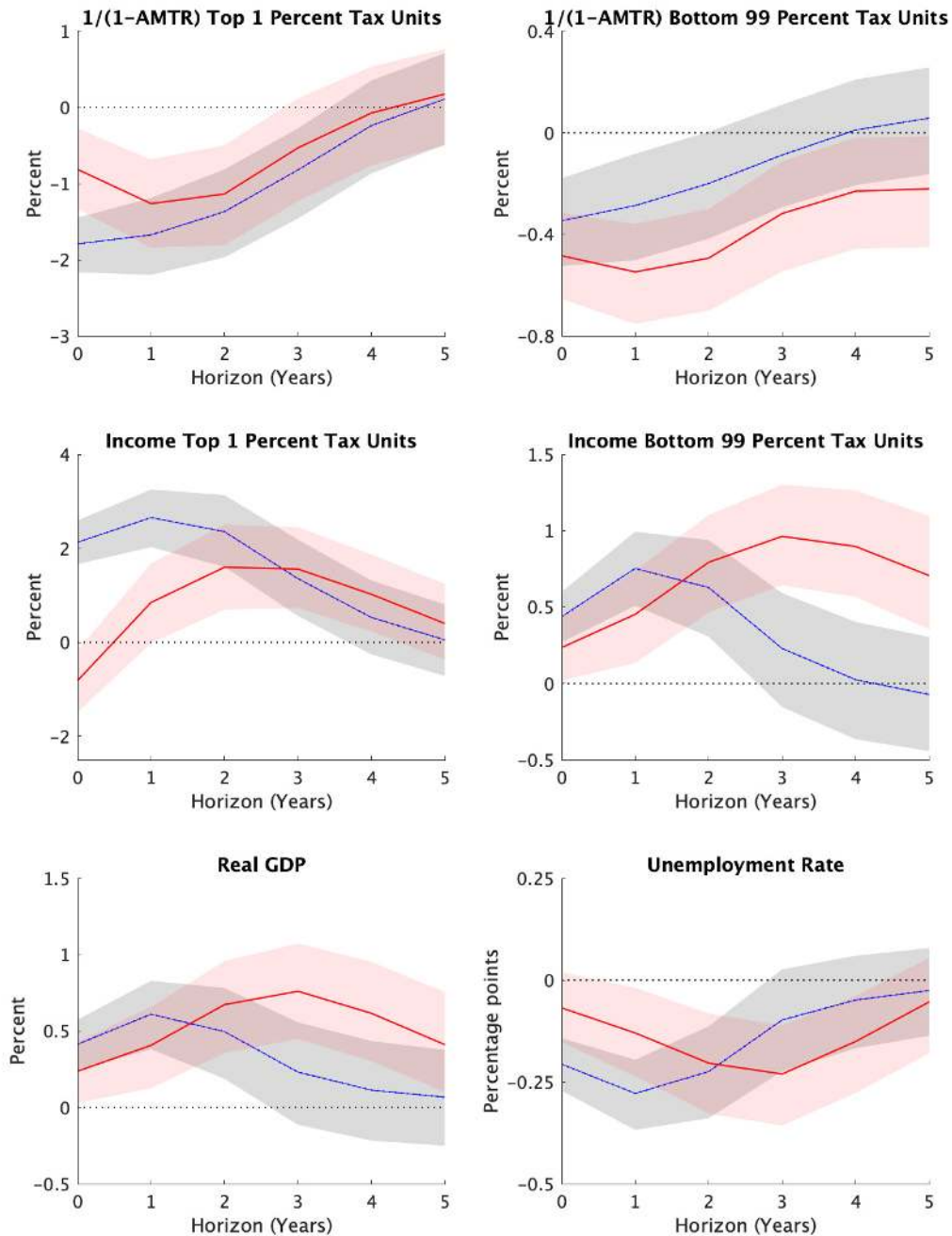
Figure 5: IRFs to structural shocks. The solid curves (blue for the 1 percent AMTR shock and red for the 99 percent AMTR shock) represent the point-wise posterior medians, and the shaded areas (gray for the 1 percent AMTR shock and red for the 99 percent AMTR shock) represent the 68 percent equal-tailed point-wise probability bands. The IRFs are with respect to a one standard deviation shock.

Sign Restrictions for Identifying AMTR Shocks to the Top 1 and Bottom 99 Percent. *(i) The proxy for the AMTR shock to the top 1 percent is positively correlated with the AMTR shock to the top 1*

percent; (ii) the proxy for the AMTR shock to the bottom 99 percent is positively correlated with the AMTR shock to the bottom 99 percent; (iii) the covariance between the AMTR shock to the top 1 percent and the proxy for the AMTR shock to the top 1 percent is bigger than the covariance between the AMTR shock to the bottom 99 percent and the proxy for the AMTR shock to the top 1 percent; and (iv) the covariance between the AMTR shock to the bottom 99 percent and the proxy for the AMTR shock to the bottom 99 percent is bigger than the covariance between the AMTR shock to the top 1 percent and the proxy for the AMTR shock to the bottom 99 percent.

Figure 5 shows the IRFs to a top 1 percent and a bottom 99 percent AMTR shock. Comparing Figures 4 and 5 it is clear that the uncertainty associated with the bottom 99 percent AMTR cut is not present when the set of sign restrictions is used. Hence, the figure shows that Mertens and Montiel-Olea's (2018) conclusions regarding the effects of tax rate cuts at the top and bottom of the income distribution will be quite robust to any linear combination of shocks.

6   Conclusion

This paper develops an efficient algorithm to independently draw from any posterior distributions over the structural parameterization of a Bayesian Proxy-SVAR. In addition, our approach expands the type of identification schemes currently considered (e.g., Montiel-Olea, Stock and Watson, 2016). More specifically, influential papers rely on counterfactuals when more than one instrument is used to identify more than one structural shock. In contrast, our approach allows researchers to individually identify structural shocks.

References

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications. *Econometrica 86*(2), 685–720.

Bahaj, S. A. (2014). Systemic Sovereign Risk: Macroeconomic Implications in the Euro Area. *Centre For Macroeconomics Working Paper*.

Barro, R. J. and C. J. Redlick (2011). Macroeconomic Effects from Government Purchases and Taxes. *The Quarterly Journal of Economics 126*(1), 51–102.

Bognanni, M. (2018). A Class of Time-Varying Parameter Structural VARs for Inference under Exact or Set Identification. *Federal Reserve Bank of Cleveland Working Paper 1*(18-11), 1–61.

Braun, R. and R. Brüggemann (2017, August). Identification of SVAR Models by Combining Sign Restrictions With External Instruments. Working Paper Series of the Department of Economics, University of Konstanz 2017-07.

Caldara, D. and E. Herbst (2016). Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs. *IFDP (2016-049), Federal Reserve Board*.

Drautzburg, T. (2016). A Narrative Approach to a Fiscal DSGE model. *Working Paper, FRB Philadelphia*.

Fernald, J. (2014). A Quarterly, Utilization-Adjusted Series on Total Factor Productivity. *Working Paper 2012-19, Federal Reserve Bank of San Francisco*.

Gertler, M. and P. Karadi (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Giacomini, R. and T. Kitagawa (2018, November). Robust Bayesian Inference for Set-identified Models. CeMMAP working papers CWP61/18, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Giacomini, R., T. Kitagawa, and M. Read (2020). Robust Bayesian Inference in Proxy SVARs. *Journal of Econometrics (Forthcoming)*.

Gleser, L. J. (1992). The Importance of Assessing Measurement Reliability in Multivariate Regression. *Journal of the American Statistical Association 87*(419), 696–707.

Gonçalves, S. and L. Kilian (2004, 02). Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics 123*, 89–120.

Jarociński, M. and P. Karadi (2018, February). Deconstructing Monetary Policy Surprises: The Role of Information Shocks. *ECB Working Paper Series No. 2133*.

Jentsch, C. and K. G. Lunsford (2019a, May). Asymptotically Valid Bootstrap Inference for Proxy SVARs. Working Paper 19-08, Federal Reserve Bank of Cleveland.

Jentsch, C. and K. G. Lunsford (2019b, July). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment. *American Economic Review 109*(7), 2655–78.

Känzig, D. R. (2019). The Macroeconomic Effects of Oil Supply Shocks: New Evidence from OPEC Announcements. *Available at SSRN 3185839*.

Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the Assumptions Behind Importance Sampling. *Journal of Econometrics 149*(1), 2–11.

Lakdawala, A. (2019). Decomposing the Effects of Monetary Policy Using an External Instruments SVAR. *Journal of Applied Econometrics 34*(6), 934–950.

Leeper, E. M., C. A. Sims, and T. Zha (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity 27*(2), 1–78.

Liu, Z., J. Fernald, and S. Basu (2012). Technology Shocks in a Two-Sector DSGE model. *Meeting Paper 1017, Society for Economic Dynamics*.

Lunsford, K. G. (2016). Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy. *Federal Reserve Bank of Cleveland Working Paper 15-28*.

Mertens, K. and J. L. Montiel-Olea (2018). Marginal Tax Rates and Income: New Time Series Evidence. *Quarterly Journal of Economics 133*(4), 1803–1884.

Mertens, K. and M. O. Ravn (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review 103*(4), 1212–47.

Montiel-Olea, J. L., J. H. Stock, and M. W. Watson (2016). Inference in Structural VARs with External Instruments. *Working Paper*.

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

Piffer, M. and M. Podstawski (2017, 12). Identifying Uncertainty Shocks Using the Price of Gold. *The Economic Journal 128*(616), 3266–3284.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913-1998. *Quarterly Journal of Economics 118*(1), 1–39.

Ramey, V. A. (2013). Government Spending and Private Activity. *Fiscal Policy after the Financial Crisis*, 19.

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

Saez, E. (2004). Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications. *Tax Policy and the Economy 18*, 117–173.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature 50*(1), 3–50.

Sims, C. A. and T. Zha (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review 39*(4), 949–968.

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

Stock, J. H. (2008). What's New in Econometrics: Time Series, Lecture 7. *Short course lectures, NBER Summer Institute at* $http: //www. nber. org/ minicourse_ 2008. html$.

Stock, J. H. and M. W. Watson (2012). Disentangling the Channels of the 2007-09 Recession. *Brookings Papers on Economic Activity: Spring 2012*, 81.

Stock, J. H. and M. W. Watson (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal 128*(610), 917–948.

Waggoner, D. F. and T. Zha (2003). A Gibbs Sampler for Structural Vector Autoregressions. *Journal of Economic Dynamics and Control 28*(2), 349–366.

A Appendix Not for Publication

A.1 The Mapping for the Unit Effect Normalization

Consider a Proxy-SVAR written as

$$\tilde{\boldsymbol{y}}_t'\hat{\boldsymbol{A}}_0 = \tilde{\boldsymbol{x}}_t'\hat{\boldsymbol{A}}_+ + \tilde{\boldsymbol{\varepsilon}}_t'\hat{\boldsymbol{D}}^{-1/2} \text{ for } 1 \leq t \leq T,$$

where $\tilde{\boldsymbol{\varepsilon}}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_{\tilde{n}})$, $\hat{\boldsymbol{A}}_0$ is non-singular with $\boldsymbol{e}_{\tilde{n},j}'\left(\hat{\boldsymbol{A}}_0^{-1}\right)' \boldsymbol{e}_{\tilde{n},j} = 1$ for $1 \leq j \leq \tilde{n}$, $\hat{\boldsymbol{A}}_+ = \left[\hat{\boldsymbol{A}}_1' \cdots \hat{\boldsymbol{A}}_p' \ \hat{\boldsymbol{c}}'\right]'$, where $\hat{\boldsymbol{c}}$ is a $1 \times \tilde{n}$ row vector, $\hat{\boldsymbol{D}}$ is a positive diagonal matrix, and the lower left-hand $k \times n$ block of $\hat{\boldsymbol{A}}_\ell$ is zero for $0 \leq \ell \leq p$. We call $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+, \hat{\boldsymbol{D}})$ the Proxy-SVAR structural parameters with unit effect normalization. The mapping from the orthogonal triangular-block parameters, $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$, to the Proxy-SVAR structural parameters with unit effect normalization, $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+, \hat{\boldsymbol{D}})$, is given by

$$\hat{f}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) = (\underbrace{\tilde{\boldsymbol{\Lambda}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)\hat{\boldsymbol{D}}^{-1/2}}_{\hat{\boldsymbol{A}}_0}, \underbrace{\tilde{\boldsymbol{\Lambda}}_+ \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)\hat{\boldsymbol{D}}^{-1/2}}_{\hat{\boldsymbol{A}}_+}, \underbrace{\operatorname{diag}(\hat{\boldsymbol{d}})^2}_{\hat{\boldsymbol{D}}}),$$

where $\hat{\boldsymbol{d}} = (\boldsymbol{e}_{\tilde{n},1}'(\tilde{\boldsymbol{\Lambda}}_0^{-1})' \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)\tilde{\boldsymbol{e}}_{\tilde{n},1}, \ldots, \hat{\boldsymbol{e}}_{\tilde{n},\tilde{n}}'(\tilde{\boldsymbol{\Lambda}}_0^{-1})' \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)\boldsymbol{e}_{\tilde{n},\tilde{n}})$. Direct calculations show that $\boldsymbol{e}_{\tilde{n},j}'\left(\hat{\boldsymbol{A}}_0^{-1}\right)' \boldsymbol{e}_{\tilde{n},j} = 1$ for $1 \leq j \leq \tilde{n}$ and the lower left-hand $k \times n$ block of $\hat{\boldsymbol{A}}_\ell$ is zero for $0 \leq \ell \leq p$. The inverse of $\hat{f}$ is

$$\hat{f}^{-1}(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+, \hat{\boldsymbol{D}}) = (\underbrace{\hat{\boldsymbol{A}}_0\hat{\boldsymbol{D}}^{1/2}\hat{\boldsymbol{P}}}_{\tilde{\boldsymbol{\Lambda}}_0}, \underbrace{\hat{\boldsymbol{A}}_+\hat{\boldsymbol{D}}^{1/2}\hat{\boldsymbol{P}}}_{\tilde{\boldsymbol{\Lambda}}_+}, \underbrace{\hat{\boldsymbol{P}}_1'}_{\hat{\boldsymbol{Q}}_1}, \underbrace{\hat{\boldsymbol{P}}_2'}_{\hat{\boldsymbol{Q}}_2}).$$

where $\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}_0^{-1} = \hat{\boldsymbol{P}}\hat{\boldsymbol{R}}$ is the QR-decomposition of $\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}_0^{-1}$, normalized so that the diagonal of $\hat{\boldsymbol{R}}$ is positive. Because the lower left-hand $k \times n$ block of $\hat{\boldsymbol{D}}^{1/2}\hat{\boldsymbol{A}}_0^{-1}$ is zero, $\hat{\boldsymbol{P}} = \operatorname{diag}(\hat{\boldsymbol{P}}_1, \hat{\boldsymbol{P}}_2)$, where $\hat{\boldsymbol{P}}_1 \in \mathcal{O}(n)$ and $\hat{\boldsymbol{P}}_2 \in \mathcal{O}(k)$. The matrix $\tilde{\boldsymbol{\Lambda}}_0$ will be upper-triangular with positive diagonal because $\tilde{\boldsymbol{\Lambda}}_0 = \hat{\boldsymbol{R}}^{-1}$. Since $\hat{\boldsymbol{P}}$ is block diagonal and the lower left-hand $k \times n$ block of each $\hat{\boldsymbol{A}}_\ell$ is zero, the lower left-hand $k \times n$ block of each $\tilde{\boldsymbol{\Lambda}}_\ell$ is zero. Algorithm 1 can now be used, but with the function f in Step 4 replaced by the function $\hat{f}$. Clearly, when using the unit effect normalization one also has to set a prior over $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+, \hat{\boldsymbol{D}})$.

A.2 Gibbs Sampler

In Waggoner and Zha (2003), a Gibbs sampler is described for sampling from a posterior distribution of a structural VAR over a certain class of normal priors and subject to a certain class of linear non-cross-equation restrictions. In that paper, the restrictions are described in terms of free parameters. In particular, if $\tilde{\boldsymbol{\lambda}}_{0,j}$ and $\tilde{\boldsymbol{\lambda}}_{+,j}$ denote the jth columns of $\tilde{\boldsymbol{\Lambda}}_0$ and $\tilde{\boldsymbol{\Lambda}}_+$, respectively, then it is assumed that the $\tilde{\boldsymbol{\lambda}}_{0,j}$ and $\tilde{\boldsymbol{\lambda}}_{+,j}$ that satisfy

the restrictions are of the form

$$\tilde{\boldsymbol{\lambda}}_{0,j} = \boldsymbol{U}_j \boldsymbol{\gamma}_{0,j} \text{ and } \tilde{\boldsymbol{\lambda}}_{+,j} = \boldsymbol{V}_j \boldsymbol{\gamma}_{+,j},$$

where both $\boldsymbol{U}_j$ and $\boldsymbol{V}_j$ have orthonormal columns for $1 \leq j \leq \tilde{n}$. Because $\tilde{\boldsymbol{\Lambda}}_0$ must be upper-triangular, $\boldsymbol{U}_j$ can be taken to be the first j columns of $\boldsymbol{I}_{\tilde{n}}$. Because $\tilde{\boldsymbol{\Lambda}}_+$ satisfies the block restrictions, $\boldsymbol{V}_j$ can be taken to be $\boldsymbol{I}_{\tilde{m}}$ for $n + 1 \leq j \leq \tilde{n}$. When $1 \leq j \leq n$, $\boldsymbol{V}_j$ will be block diagonal with the first p blocks equal to the first n columns of $\boldsymbol{I}_{\tilde{n}}$ and the last block the scalar one.

The Gibbs sampler is described in terms of a non-negative scalar T and matrices $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$, for $1 \leq j \leq \tilde{n}$. In Waggoner and Zha (2003), the goal was to sample from a posterior and so T, $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$ were given in terms of restrictions, prior, and data. In this paper our goal is to sample from an NGN distribution conditional on the above restrictions, so we will describe T, $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$ in terms of $\hat{\nu}$, $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ and the above restrictions. The $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ must be block diagonal, so we assume that $\hat{\boldsymbol{\Phi}} = \text{diag}(\hat{\boldsymbol{\Phi}}_1, \ldots, \hat{\boldsymbol{\Phi}}_j, \ldots, \hat{\boldsymbol{\Phi}}_{\tilde{n}})$, $\hat{\boldsymbol{\Psi}} = \text{diag}(\hat{\boldsymbol{\Psi}}_1, \ldots, \hat{\boldsymbol{\Psi}}_j, \ldots, \hat{\boldsymbol{\Psi}}_{\tilde{n}})$, and $\hat{\boldsymbol{\Omega}} = \text{diag}(\hat{\boldsymbol{\Omega}}_1, \ldots, \hat{\boldsymbol{\Omega}}_j, \ldots, \hat{\boldsymbol{\Omega}}_{\tilde{n}})$. More specifically, we draw $\boldsymbol{\gamma}_{0,j}$ from a generalized-normal distribution with parameters ν and $\boldsymbol{S}_j^{-1}$ and we draw $\boldsymbol{\gamma}_{+,j}$ given $\boldsymbol{\gamma}_{0,j}$ from a normal distribution with mean $\boldsymbol{P}_j \boldsymbol{\gamma}_{0,j}$ and variance $\boldsymbol{H}_j$, where

$$\nu = T + \tilde{n}$$

$$\boldsymbol{H}_j = (\boldsymbol{V}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \boldsymbol{V}_j)^{-1}$$

$$\boldsymbol{P}_j = \boldsymbol{H}_j \boldsymbol{V}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \hat{\boldsymbol{\Psi}}_j \boldsymbol{U}_j$$

$$\boldsymbol{S}_j = (\boldsymbol{U}_j' \hat{\boldsymbol{\Phi}}_j \boldsymbol{U}_j + \boldsymbol{U}_j' \hat{\boldsymbol{\Psi}}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \hat{\boldsymbol{\Psi}}_j \boldsymbol{U}_j - \boldsymbol{P}_j' \boldsymbol{H}_j^{-1} \boldsymbol{P}_j)^{-1}.$$

A.3 Proposal Normal-Generalized-Normal Parameters

As mentioned in Section 3, while it often suffices to choose $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ to be equal to $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, there are instances in which this can lead to small effective sample sizes in our importance sampler. In such cases we find it useful to tailor the choice of $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ by choosing the value of $\hat{\boldsymbol{\Phi}}$ that minimizes the square of the difference between the target and the proposal density evaluated at a given number of draws of the posterior distribution over the structural parameterization obtained when $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ is set equal to $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$.

A.4 The Functions β and ζ

To show that the derivative of β has full row rank and to prove Proposition 1, both of which are needed to apply Theorem 1, there will need to be regularity conditions to ensure there is enough variation in the

functions used to define the zero restrictions. Let $\tilde{\boldsymbol{A}}(\boldsymbol{u}) = (\tilde{\boldsymbol{A}}_0(\boldsymbol{u}), \tilde{\boldsymbol{A}}_+(\boldsymbol{u}))$ denote any one-to-one linear mapping from $\mathbb{R}^b$ onto the set of all Proxy-SBVAR structural parameters. The regularity conditions require that $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u}) = D\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))D\tilde{\boldsymbol{A}}(\boldsymbol{u})$ be of full row rank for all $\boldsymbol{u} \in \mathbb{R}^b$.[9] Because of the block restrictions, $D\tilde{\boldsymbol{A}}(\boldsymbol{u})$ is not of full row rank. So, it is not sufficient for $D\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ to be of full row rank, though it is necessary. At the end of this appendix, we will return to the regularity conditions and explore some of the types of restrictions that can easily be imposed in this framework.

Proposition 2. *The derivative of the function $\beta(\boldsymbol{u})$, which defines the zero restrictions, has full row rank for every $\boldsymbol{u} \in \mathbb{R}^b$.*

Proof. The function $\beta : \mathbb{R}^b \to \mathbb{R}^{b-d}$ and its derivative are given by

$$
\beta(\boldsymbol{u}) = \begin{bmatrix} \tilde{\boldsymbol{Z}}_1 \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))\boldsymbol{e}_{n,1} \\ \vdots \\ \tilde{\boldsymbol{Z}}_n \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))\boldsymbol{e}_{n,n} \end{bmatrix} \quad \text{and} \quad D\beta(\boldsymbol{u}) = \begin{bmatrix} \boldsymbol{e}'_{n,1} \otimes \tilde{\boldsymbol{Z}}_1 \\ \vdots \\ \boldsymbol{e}'_{n,n} \otimes \tilde{\boldsymbol{Z}}_n \end{bmatrix} D\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))D\tilde{\boldsymbol{A}}(\boldsymbol{u}).
$$

The first term in the expression for $D\beta(\boldsymbol{u})$ is of full row rank because $\tilde{\boldsymbol{Z}}_j$ is of full row rank for $1 \leq j \leq n$. By the regularity conditions, $D\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))D\tilde{\boldsymbol{A}}(\boldsymbol{u})$ has full row rank for every $\boldsymbol{u} \in \mathbb{R}^b$. So, $D\beta(\boldsymbol{u})$ has full row rank for every $\boldsymbol{u} \in \mathbb{R}^b$. $\qquad\square$

If B and C are sets, let $B\backslash C$ denote the complement of C in B. Note that we do not require $C \subset B$.

Proof of Proposition 1. Section 3.4 showed how to construct the functions $\boldsymbol{K}_{i,j}$. All that remains to be shown is that U is open and that $\beta^{-1}(\{\boldsymbol{0}\})\backslash U$ is of measure zero in $\beta^{-1}(\{\boldsymbol{0}\})$. The vector $\boldsymbol{u} \in U$ if and only if the matrix $\boldsymbol{M}_{1,j}(f^{-1}(\tilde{\boldsymbol{A}}(\boldsymbol{u})))$ is of full row rank for $1 \leq j \leq n$, and $\boldsymbol{M}_{1,j}(f^{-1}(\tilde{\boldsymbol{A}}(\boldsymbol{u})))$ is of full row rank for $1 \leq j \leq n$ if and only if $\det(\boldsymbol{M}_{1,j}(f^{-1}(\tilde{\boldsymbol{A}}(\boldsymbol{u})))\boldsymbol{M}_{1,j}(f^{-1}(\tilde{\boldsymbol{A}}(\boldsymbol{u})))') \neq 0$ for $1 \leq j \leq n$. Since the determinant is continuous, this implies that U is open.

Proposition 2 states that $D\beta(\boldsymbol{u})$ is of full row rank for every $\boldsymbol{u} \in \mathbb{R}^b$. This implies that $\beta^{-1}(\{\boldsymbol{0}\})$ is a $(b - \sum_{j=1}^n \tilde{z}_{1,j})$-dimensional smooth manifold in $\mathbb{R}^b$.[10] Thus, there is a natural measure on $\beta^{-1}(\{\boldsymbol{0}\})$ called the volume measure.[11] We show that $\beta^{-1}(\{\boldsymbol{0}\})\backslash U$ is of measure zero with respect to the volume measure over $\beta^{-1}(\{\boldsymbol{0}\})$.

---

[9]If $\boldsymbol{T}(\boldsymbol{x})$ is a matrix value function of the vector $\boldsymbol{x}$, then $D\boldsymbol{T}(\boldsymbol{x})$ denotes the total derivative of $(\text{vec} \circ \boldsymbol{T})(\boldsymbol{x})$, where vec is the operator that stacks the columns of a matrix into a vector. If $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, then $D_{\boldsymbol{x}_1}\boldsymbol{T}(x_1, x_2)$ denotes the partial derivative with respect to $\boldsymbol{x}_1$ of $(\text{vec} \circ \boldsymbol{T})(\boldsymbol{x})$. Most of the properties of matrix derivatives follow from the properties of the vec operator. For instance, the product rule, which is repeatedly used in this appendix, follows from the fact that $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\,\text{vec}(\boldsymbol{B})$, for all conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$.

[10]See Theorem 5-1 of Spivak (1965).

[11]See Arias, Rubio-Ramírez, and Waggoner (2018) for a discussion of the volume measure over smooth manifolds.

The implicit function theorem implies that for every $\tilde{\boldsymbol{u}} \in \beta^{-1}(\{\boldsymbol{0}\}) \subset \mathbb{R}^b$, there are open sets $A_{\tilde{\boldsymbol{u}}} \subset \mathbb{R}^b$ about $\tilde{\boldsymbol{u}}$ and $B_{\tilde{\boldsymbol{u}}} = B_{\tilde{\boldsymbol{u}}}^1 \times B_{\tilde{\boldsymbol{u}}}^2 \subset \mathbb{R}^{b-(k+r)n} \times \mathbb{R}^{(k+r)n}$ and a diffeomorphism $h_{\tilde{\boldsymbol{u}}} : B_{\tilde{\boldsymbol{u}}} \to A_{\tilde{\boldsymbol{u}}}$ such that $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(h_{\tilde{\boldsymbol{u}}}(\boldsymbol{u}^1, \boldsymbol{u}^2))) = \boldsymbol{u}^2$, for every $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in B_{\tilde{\boldsymbol{u}}}^1 \times B_{\tilde{\boldsymbol{u}}}^2$, where $\boldsymbol{u}^2$ is interpreted as a $(k+r) \times n$ matrix.[12] Because smooth manifolds are second countable, there exist $\tilde{\boldsymbol{u}}_i \in \beta^{-1}(\{\boldsymbol{0}\})$, for $i \in \{1, 2, \cdots\}$, such that $\beta^{-1}(\{\boldsymbol{0}\}) \subset \bigcup_{i=1}^{\infty} A_{\tilde{\boldsymbol{u}}_i}$.[13] So, $\beta^{-1}(\{\boldsymbol{0}\}) \backslash U = \bigcup_{i=1}^{\infty}((A_{\tilde{\boldsymbol{u}}_i} \cap \beta^{-1}(\{\boldsymbol{0}\})) \backslash U)$, and it thus suffices to show that for every $\tilde{\boldsymbol{u}} \in \beta^{-1}(\{\boldsymbol{0}\})$ the set $(A_{\tilde{\boldsymbol{u}}} \cap \beta^{-1}(\{\boldsymbol{0}\})) \backslash U$ is of measure zero with respect to the volume measure over $\beta^{-1}(\{\boldsymbol{0}\})$.

Let $\mathcal{U}_{\tilde{\boldsymbol{u}}} = h_{\tilde{\boldsymbol{u}}}^{-1}(A_{\tilde{\boldsymbol{u}}} \cap \beta^{-1}(\{\boldsymbol{0}\}))$. Since $\beta(h_{\tilde{\boldsymbol{u}}}(\boldsymbol{u}^1, \boldsymbol{u}^2)) = (\tilde{\boldsymbol{Z}}_1 \boldsymbol{u}^2 \boldsymbol{e}_{n,1}, \cdots, \tilde{\boldsymbol{Z}}_n \boldsymbol{u}^2 \boldsymbol{e}_{n,n})$ is a linear function, $\mathcal{U}_{\tilde{\boldsymbol{u}}}$ is the intersection of a linear subspace of $\mathbb{R}^b$ and $B_{\tilde{\boldsymbol{u}}}^1 \times B_{\tilde{\boldsymbol{u}}}^2$. Thus, the volume measure is defined over $\mathcal{U}_{\tilde{\boldsymbol{u}}}$.

For $0 \le d \le c$, let $\boldsymbol{L}_{c,d} = [\boldsymbol{I}_d \ \boldsymbol{0}_{d \times c-d}]$ and let $\boldsymbol{J}_{c,d} = [\boldsymbol{0}_{d \times c-d} \ \boldsymbol{I}_d]$. For $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in B_{\tilde{\boldsymbol{u}}}^1 \times B_{\tilde{\boldsymbol{u}}}^2$, define

$$\tilde{\boldsymbol{M}}_{1,j}(\boldsymbol{u}_1, \boldsymbol{u}_2) = \begin{bmatrix} \tilde{\boldsymbol{Z}}_j \tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(h_{\tilde{\boldsymbol{u}}}(\boldsymbol{u}^1, \boldsymbol{u}^2))) \\ \boldsymbol{L}_{n,j-1} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{Z}}_j \boldsymbol{u}_2 \\ \boldsymbol{L}_{n,j-1} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{Z}}_j \boldsymbol{u}_2 \boldsymbol{L}'_{n,j-1} & \tilde{\boldsymbol{Z}}_j \boldsymbol{u}_2 \boldsymbol{J}'_{n,n-j+1} \\ \boldsymbol{I}_{j-1} & \boldsymbol{0}_{(j-1) \times (n-j+1)} \end{bmatrix}. \tag{9}$$

Let $f^{-1}(\tilde{\boldsymbol{A}}(h_{\tilde{\boldsymbol{u}}}(\boldsymbol{u}^1, \boldsymbol{u}^2))) = (\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Because $\tilde{\boldsymbol{F}}_z$ is orthogonally commutative, $\tilde{\boldsymbol{M}}_{1,j}(\boldsymbol{u}^1, \boldsymbol{u}^2) = \boldsymbol{M}_{1,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) \boldsymbol{Q}_1$. Thus, $\boldsymbol{M}_{1,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ is of full row rank if and only if $\tilde{\boldsymbol{M}}_{1,j}(\boldsymbol{u}^1, \boldsymbol{u}^2)$ is of full row rank. From the last expression in Equation (9), $\tilde{\boldsymbol{M}}_{1,j}(\boldsymbol{u}^1, \boldsymbol{u}^2)$ is of full row rank if and only $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,n-j+1}$ is of column rank at least $\tilde{z}_{1,j}$. Let $U_{\tilde{\boldsymbol{u}}}$ be the set of all $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in B_{\boldsymbol{u}}^1 \times B_{\boldsymbol{u}}^2$ such that $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$ is of column rank at least $\tilde{z}_{1,j}$ for $1 \le j \le n$.

Since $h_{\tilde{\boldsymbol{u}}}^{-1}(A_{\tilde{\boldsymbol{u}}} \cap \beta^{-1}(\{\boldsymbol{0}\})) \backslash U) = \mathcal{U}_{\tilde{\boldsymbol{u}}} \backslash U_{\tilde{\boldsymbol{u}}}$, it suffices to show that $\mathcal{U}_{\tilde{\boldsymbol{u}}} \backslash U_{\tilde{\boldsymbol{u}}}$ is of measure zero with respect to the volume measure over $\mathcal{U}_{\tilde{\boldsymbol{u}}}$, which will follow from showing that for almost all $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{U}_{\boldsymbol{u}}$, the columns of $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$ are linearly independent. For $1 \le \ell \le \tilde{z}_{1,j}$, the ℓth column of $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$ is $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{e}_{n,\hat{j}}$, where $\hat{j} = n - (\tilde{z}_{i,j} - \ell)$. Because the dimension of the span of the set of all $\boldsymbol{u}^2 \boldsymbol{e}_{n,\hat{j}} \in \mathbb{R}^{k+r}$ with $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{U}_{\tilde{\boldsymbol{u}}}$ is $r + k - \tilde{z}_{1,\hat{j}}$ and the dimension of the row space of $\tilde{\boldsymbol{Z}}_j$ in $\mathbb{R}^{k+r}$ is $\tilde{z}_{1,j}$, the intersection of these two linear spaces is of dimension at least $\tilde{z}_{1,j} - \tilde{z}_{1,\hat{j}} \ge \ell$. Thus, for $0 \le \ell \le \tilde{z}_{1,j}$, the dimension of the span of the set of all $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{e}_{n,\hat{j}} \in \mathbb{R}^{\tilde{z}_{1,j}}$ with $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{U}_{\tilde{\boldsymbol{u}}}$ is at least ℓ. By a simple dimension argument, this implies that for almost all $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{U}_{\tilde{\boldsymbol{u}}}$ the ℓth column of $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$ is not in the span of the first $\ell - 1$ columns of $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$. Thus, for almost all $(\boldsymbol{u}^1, \boldsymbol{u}^2) \in \mathcal{U}_{\tilde{\boldsymbol{u}}}$ the columns of $\tilde{\boldsymbol{Z}}_j \boldsymbol{u}^2 \boldsymbol{J}'_{n,\tilde{z}_{1,j}}$ are linearly independent. $\qquad\square$

We now return to the regularity conditions and discuss the kinds of restrictions that can be easily imposed within this framework. Write $\tilde{\boldsymbol{F}}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ as $[\boldsymbol{F}_e(\tilde{\boldsymbol{A}}_0(\boldsymbol{u}))' \quad \boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))']'$, where $\boldsymbol{F}_e(\tilde{\boldsymbol{A}}_0(\boldsymbol{u})) = \boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})' \boldsymbol{L}' =$

---

[12]See Theorem 2-13 of Spivak (1965).
[13]A topological space is second countable if and only if the space has a countable basis.

$(\boldsymbol{A}_0^{-1}\boldsymbol{\Gamma}_{0,1}\boldsymbol{\Gamma}_{0,2}^{-1})'$. We can assume without loss of generality that the first nk elements of $\boldsymbol{u}$ correspond to the elements of $\boldsymbol{\Gamma}_{0,1}$ and the next $(p+1)n^2+n$ elements of $\boldsymbol{u}$ correspond to the SBVAR structural parameters $(\boldsymbol{A}_0, \boldsymbol{A}_+)$. So we can write $\boldsymbol{u}$ as $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) \in \mathbb{R}^{nk} \times \mathbb{R}^{(p+1)n^2+n} \times \mathbb{R}^{a-nk-(p+1)n^2-n}$. Note that

$$D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u}) = \begin{bmatrix} D_{\boldsymbol{u}_1}(\boldsymbol{F}_e \circ \tilde{\boldsymbol{A}}_0)(\boldsymbol{u}) & D_{\boldsymbol{u}_2}(\boldsymbol{F}_e \circ \tilde{\boldsymbol{A}}_0)(\boldsymbol{u}) & D_{\boldsymbol{u}_3}(\boldsymbol{F}_e \circ \tilde{\boldsymbol{A}}_0)(\boldsymbol{u}) \\ D_{\boldsymbol{u}_1}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u}) & D_{\boldsymbol{u}_2}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u}) & D_{\boldsymbol{u}_3}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u}) \end{bmatrix}.$$

Since $D_{\boldsymbol{u}_1}(\boldsymbol{F}_e \circ \tilde{\boldsymbol{A}}_0)(\boldsymbol{u}) = (\boldsymbol{A}_0^{-1} \otimes (\boldsymbol{\Gamma}_{0,2}^{-1})')\boldsymbol{P}$, for some permutation matrix $\boldsymbol{P}$, $D_{\boldsymbol{u}_1}(\boldsymbol{F}_e \circ \tilde{\boldsymbol{A}}_0)(\boldsymbol{u})$ is of full row rank. Thus, if $D_{\boldsymbol{u}_1}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ is zero and $D_{\boldsymbol{u}_2}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ were of full row rank, then $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ would be of full row rank. If each row of $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ were equal to a row of $\boldsymbol{A}_0$, or a row of $\boldsymbol{A}_+$, or the impulse responses of one endogenous variable to all the structural shocks at one horizon, then $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ would be orthogonally commutative and $D_{\boldsymbol{u}_1}(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ would be zero since $F_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ would depend only on $(\boldsymbol{A}_0, \boldsymbol{A}_+)$. As long as not too many such rows are included, $D(\boldsymbol{F}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ will also be of full row rank. For instance, if $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u})) = (\boldsymbol{A}_0^{-1})'$, which is the contemporaneous impulse response of all the endogenous variables to all the structural shocks, then $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ would be of full row rank. Similarly, if $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u})) = \boldsymbol{A}_0$, then $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ would be of full row rank. However, if $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u})) = [\boldsymbol{A}_0' \ \boldsymbol{A}_0^{-1}]'$, then $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ would not be of full row rank. Since the number of zero restrictions is small, $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$ can usually be defined so that the desired restrictions can be imposed and $D(\tilde{\boldsymbol{F}}_z \circ \tilde{\boldsymbol{A}})(\boldsymbol{u})$ is of full row rank.

So, in this framework, we could have zero restrictions on the elements of $\boldsymbol{A}_0$ or $\boldsymbol{A}_+$ or on impulse responses of endogenous variables to structural shocks. These correspond to restrictions on $\boldsymbol{F}_z(\tilde{\boldsymbol{A}}(\boldsymbol{u}))$. We also could have additional zero restrictions on $\boldsymbol{F}_e(\tilde{\boldsymbol{A}}_0(\boldsymbol{u}))$, which is the covariance matrix of the proxies and the structural shocks. The exogeneity restrictions already require the first $n-k$ columns of $\boldsymbol{F}_e(\tilde{\boldsymbol{A}}_0(\boldsymbol{u}))$ to be zero, but one could impose additional zero restrictions on the last k columns. The last k columns of $\boldsymbol{F}_e(\tilde{\boldsymbol{A}}_0(\boldsymbol{u}))$ are the covariance matrix of the proxies and the structural shocks correlated with the proxies.

A.5 Data Appendix for Section 4

Here we describe the data used in Section 4 in more detail. The time series used to construct the endogenous variables used in the Proxy-SVAR are:

1. Real Gross Domestic Product, BEA, NIPA table 1.1.6, line 1, billions of chained (2009) dollars, seasonally adjusted at annual rates. Downloaded from https://www.bea.gov.

2. Total Private Employment, BLS, Current Employment Statistics survey (National), series Id CES0500000001, thousands, seasonally adjusted. Downloaded from https://www.bls.gov.

3. Price Index for Gross Domestic Product, BEA, NIPA table 1.1.4, line 1, index 2009=100, seasonally adjusted. Downloaded from https://www.bea.gov.

4. Personal Consumption Expenditures on Non-durable Goods, BEA NIPA table 1.1.5, line 5, billions of dollars, seasonally adjusted at annual rate. Downloaded from https://www.bea.gov.

5. Personal Consumption Expenditures on Services, BEA NIPA table 1.1.5, line 6, billions of dollars, seasonally adjusted at annual rate. Downloaded from https://www.bea.gov.

6. Personal Consumption Expenditures on Durable Goods, BEA NIPA table 1.1.5, line 4, billions of dollars, seasonally adjusted at annual rate. Downloaded from https://www.bea.gov.

7. Fixed Investment in Equipment, BEA NIPA table 1.1.5, line 8, billions of dollars, seasonally adjusted at annual rate. Downloaded from https://www.bea.gov.

8. Real Consumption = (4)+(5) / (3)

9. Real Investment in Equipment = (6)+(7) / (3)

The endogenous variables in the SVAR are series (1), (2), (3), (8), and (9) transformed to percent log differences.

A.6 Finite Variance Tests of Importance Sampling Weights

We numerically test for the variance of the importance sampler weights to be finite in each of the applications of the paper. In particular we use the Wald, score, and likelihood ratio (LR) tests as described in Koopman, Shephard, and Creal (2009). These tests assume that the importance sampler weights are independent draws from a Pareto distribution characterized by the shape parameter ξ. The null of each one of these tests is

$$H_0 : \xi = \frac{1}{2} \text{ and } H_1 : \xi > \frac{1}{2}$$

because for $\xi > \frac{1}{2}$ the Pareto distribution variance does not exist.

We conduct the tests for several thresholds of the importance sampler weights ranging from the largest 50 percent to the largest 1 percent of the importance sampler weights. These thresholds determine the number of importance sampler weights used to implement the tests. The 95 percent critical values for the Wald, score, and LR tests are 1.64, 1.64, and 2.69 respectively.

A.7 Tests for the Analysis in Section 4

Table A.1 shows the value of the tests described above for several thresholds—as shown in the first row of the table—applied to the analysis performed in Section 4. The second row of the table shows the Wald test

statistics, the third row shows the score test statistics, and the fourth row shows the LR test statistics. None of the values displayed in Table A.1 exceed the critical values reported above. Hence, these tests indicate that the importance sampler weights have finite variance.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-7.50	-6.94	-6.18	-3.77	-1.28
Score	-11.52	-9.84	-8.13	-4.41	-1.14
LM	0	0	0	0	0

Table A.1: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 4.

A.8 Tests for the Analysis in Section 5.1

Table A.2 shows that Koopman, Shephard and Creal's (2009) tests indicate that the importance weights used in Section 5.1 have finite variance.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-4525.94	-3949.29	-3417.48	-1515.45	-279.51
Score	-5.44	-5.22	-5.19	-3.31	-1.34
LM	0	0	0	0	0

Table A.2: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.1.

A.9 Tests for the Analysis in Section 5.2

Tables A.3 and A.4 present Koopman, Shephard and Creal's (2009) tests for the case in which the average and marginal tax rates are identified using the counterfactuals in Mertens and Montiel-Olea (2018) described in Section 5.2. Table A.5 shows Koopman, Shephard and Creal's (2009) tests for the Proxy-SVAR identified with the proxy and sign restrictions described in Section 5.2.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-21.15	-17.97	-14.76	-7.19	-1.21
Score	-11.12	-10.09	-9.05	-6.32	-1.77
LM	0	0	0	0	0

Table A.3: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.2 for the AMTR Counterfactual.

The tables indicate that the importance weights have finite variance.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-16.91	-14.56	-12.20	-5.75	-1.29
Score	-10.59	-9.86	-9.25	-6.21	-2.54
LM	0	0	0	0	0

Table A.4: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.2 for the ATR Counterfactual.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-13.36	-11.40	-9.65	-4.88	-0.73
Score	-8.75	-7.99	-7.58	-5.79	-1.44
LR	0	0	0	0	0

Table A.5: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.2 for the Proxy-SVAR Identified with Proxy and Sign Restrictions.

A.10   Tests for the Analysis in Section 5.3

Tables A.6 and A.7 present Koopman, Shephard and Creal's (2009) tests for the case in which the bottom and top tax rate cut shocks are identified using the counterfactuals in Mertens and Montiel-Olea (2018) described in Section 5.3. Table A.8 presents Koopman, Shephard and Creal's (2009) tests for the case in which the bottom and top tax rate cut shocks are identified using the less restrictive scheme described in Section 5.3.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-60.73	-51.60	-41.26	-17.82	-3.83
Score	-21.99	-18.55	-14.76	-7.36	-2.99
LM	0	0	0	0	0

Table A.6: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.3 for the Top 1 percent Counterfactual.

The tables indicate that the importance weights have finite variance.

A.11   Robustness Analysis of the Reliability Matrix

To get a sense of the effects that the truncation on γ has on the results, we suggest using the non-zero entries of such a matrix to compute the minimum eigenvalue of the reliability matrix of the external instruments, a metric that has been used by previous studies such as Mertens and Ravn (2013) and Caldara and Herbst (2016) to measure the strength of the instruments. If there is substantial probability mass concentrated around the threshold, then the γ-relevance condition plays an important role for the results. In contrast, if there is low probability mass concentrated around the threshold, then the γ-relevance condition plays a minor role for

| Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-39.66	-33.38	-26.16	-9.81	-1.31
Score	-14.48	-12.42	-9.75	-4.21	-1.17
LM	0	0	0	0	0

Table A.7: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.3 for the Bottom 99 Percent Counterfactual.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-27.69	-23.75	-18.92	-7.84	-1.33
Score	-10.10	-8.88	-6.87	-3.28	-0.85
LM	0	0	0	0	0

Table A.8: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.3 for the Proxy-SVAR Identified Using the Less Restrictive Identification Scheme.

the results.

Figure A.1 plots a histogram computed using the posterior draws of the minimum eigenvalue of the reliability matrix for the case in which the threshold (γ) is set equal to 0.2 and for the case in which the threshold (γ) is set equal to 0. The former is depicted by bars with dot-dashed blue edges, the latter is depicted by bars that feature a gray face. As can be seen, the histograms are essentially identical. The same occurs when plotting a cumulative histogram (see Figure A.2). A similar conclusion can be obtained in the
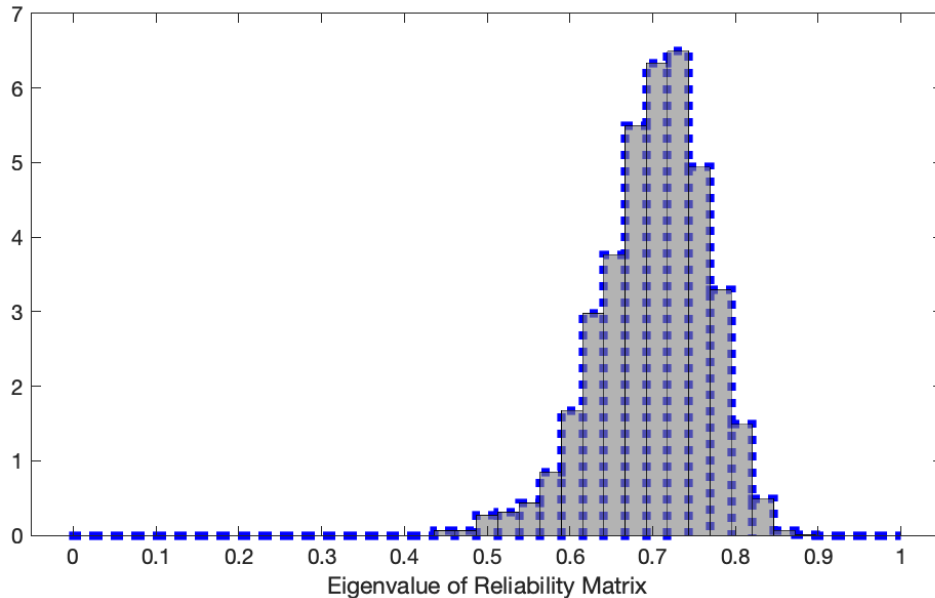
Figure A.1: Reliability Matrix Eigenvalue: Histogram

application described in Section 4. Finally, when using this approach for the applications in Sections 5.2 and 5.3, most of the probability mass is in values of γ well beyond the 0.2 threshold, but the algorithm becomes
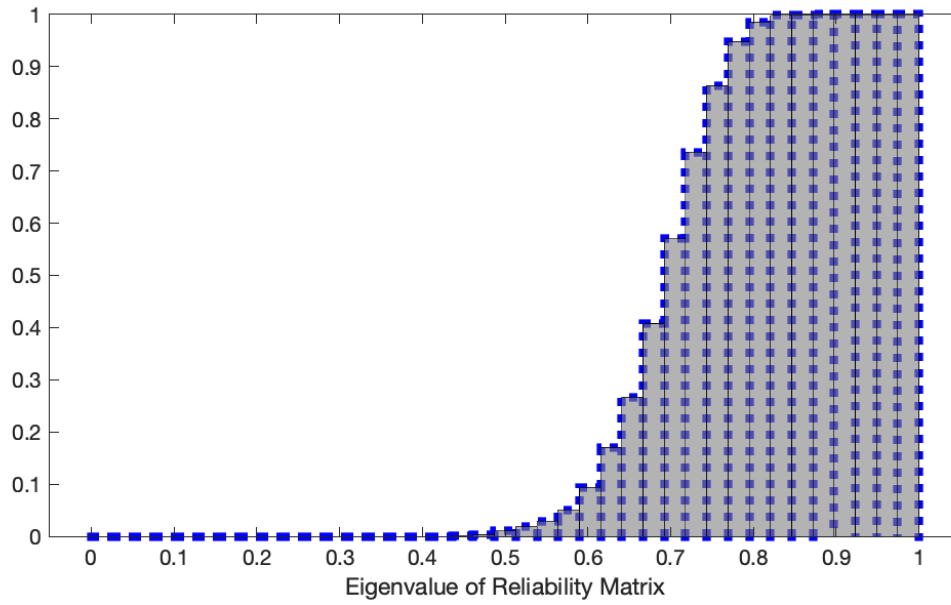
Figure A.2: Reliability Matrix Eigenvalue: Cumulative Histogram

very inefficient and unreliable for $\gamma = 0$.