

Inference in Bayesian Proxy-SVARs

Jonas E. Arias*

Juan F. Rubio-Ramírez†

Daniel F. Waggoner ‡

February 5, 2020

Abstract

Motivated by the increasing use of external instruments to identify structural vector autoregressions (SVARs), we develop algorithms for exact finite sample inference in this class of time series models, commonly known as Proxy-SVARs. Our algorithms make independent draws from the normal-generalized-normal family of conjugate posterior distributions over the structural parameterization of a Proxy-SVAR. Our approach allows researchers to simultaneously use proxies and traditional sign and zero 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.

JEL classification: C15; C32

Keywords: SVARs; External Instruments; Importance Sampler

The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. No statements here should be treated as legal advice.

*FEDERAL RESERVE BANK OF PHILADELPHIA. Email: jonas.arias@phil.frb.org

†EMORY UNIVERSITY, FEDERAL RESERVE BANK OF ATLANTA, AND BBVA RESEARCH. Email: juan.rubio-ramirez@emory.edu

‡FEDERAL RESERVE BANK OF ATLANTA. Email: daniel.f.waggoner@atl.frb.org

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. Currently, most of the papers using Proxy-SVARs work under the frequentist paradigm.¹ But, while substantial progress has been made on this front, less is known about conducting Bayesian inference in this class of structural time series models.

In this paper we contribute to this line of research by developing efficient algorithms to independently draw from any desired posterior distribution over the structural parameterization of a Proxy-SVAR conditional on exogeneity restrictions. These restrictions require that the correlation between the proxies and some subset of the structural shocks be zero. The fact that we can draw independently opens the door to use the Bayesian paradigm in larger models. We will write all our algorithms as independently drawing from the family of restricted normal-generalized-normal (NGN) posterior distribution over the structural parameterization of a Proxy-SVAR conditional on exogeneity restrictions. However, our techniques are not limited to the NGN family and can be applied to any desired prior over the structural parameterization of a Proxy-SVAR. The rationale for writing our algorithms in terms of the NGN distribution over the structural parameterization is that it is a conjugate family of distributions commonly used in the literature that has the appealing property of giving the same prior and posterior weight to observationally equivalent Proxy-SVAR structural parameters.

Our main algorithm combines the sampler developed by [Waggoner and Zha \(2003\)](#) with a variant of the importance sampler developed by [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#). The fundamental insight in the latter was to produce independent draws from a normal-inverse-Wishart distribution over the reduced-form parameters and generalize the QR decomposition to produce independent draws from a distribution over the orthogonal-reduced-form parameterization of the SVAR conditional on sign and zero restrictions. These draws were then mapped into the structural parameterization of the SVAR. By taking appropriate care of the volume elements, the draws were weighted so that they came from any desired posterior distribution over the structural parameterization of the SVAR conditional on the identification restrictions. Since a Proxy-SVAR identified with exogeneity restrictions can be represented by a SVAR identified with zero restrictions one may be tempted to use [Arias, Rubio-Ramírez and Waggoner's \(2018\)](#) algorithm. However, the techniques of that paper cannot be directly applied in this environment because the zero restrictions embedded in a Proxy-SVAR restrict its reduced-form representation, which invalidates the use of the orthogonal-reduced-form parameterization for

¹For example, see [Stock \(2008\)](#), [Stock and Watson \(2012\)](#), [Mertens and Ravn \(2013\)](#), [Gertler and Karadi \(2015\)](#), and [Montiel-Olea, Stock, and Watson \(2016\)](#).

our purpose. To handle this issue, we introduce a new parameterization called the orthogonal-triangular-block parameterization—composed of triangular-block parameters and orthogonal matrices.

Using [Waggoner and Zha \(2003\)](#), we will produce independent draws from a normal-generalized-normal distribution over the triangular-block parameters and further generalize the QR decomposition to produce independent draws from a distribution over the orthogonal-triangular-block parameterization conditional on the exogeneity restrictions. We will then map these draws into the Proxy-SVAR structural parameterization. These draws are weighted by the appropriate volume elements, so that they come from any desired posterior distribution over the structural parameterization of the Proxy-SVAR conditional on the exogeneity restrictions of interest.

We also show that those restrictions may not be enough to identify the Proxy-SVAR equations associated with structural shocks that are correlated with the proxies. In particular, additional sign and zero restrictions are needed for identification when more than one proxy is used to identify the same number of Proxy-SVAR equations. For this reason, we adapt our main algorithm to consider these additional restrictions, which could be used not only to identify the Proxy-SVAR equations associated with the structural shocks correlated with the proxies but also to identify the Proxy-SVAR equations associated with those structural shocks that are uncorrelated with the proxies.

We present two applications to illustrate our algorithms. 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 studying the dynamic effects of consumption and investment total factor productivity (TFP) shocks in a Proxy-SVAR where the equations associated with the shocks of interest are identified using [Fernald's \(2014\)](#) TFP series as external instruments as in [Lunsford \(2016\)](#). An important difference between our approach and [Lunsford's \(2016\)](#) approach is that, while he identifies one structural equation at a time by using a single instrument, we jointly identify two structural equations using two instruments.² Hence, the application allows us to emphasize how to use additional sign and zero restrictions to simultaneously identify more than one equation. In particular, we identify the structural equations by assuming that they are the only equations whose 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. Like [Lunsford \(2016\)](#), we find that a positive consumption TFP shock causes an increase in real GDP and consumption in non-durables and services as well as in durables and equipment while the price level gradually decreases. Accordingly, such a shock resembles a standard TFP shock. In contrast, a positive investment TFP shock leads to a decrease of real

²[Lunsford's \(2016\)](#) approach is a common approach in the literature (see [Stock and Watson, 2012](#)).

GDP, employment, consumption, and the price level. These results are inconsistent with the conventional wisdom of standard TFP shocks but in line with the findings in [Liu, Fernald, and Basu \(2012\)](#).

The second application is aimed at highlighting that the distinctive feature of our approach illustrated in the previous application—i.e., using more than one instrument to simultaneously set identify more than one structural equation—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 \(2018\)](#). As we will discuss later, the fundamental issue with their approach is that it is limited to counterfactual experiments. We will make this clear by revisiting [Mertens and Montiel-Olea \(2018\)](#). Such 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 shocks rather in the individual structural shocks. As a result, we propose to identify the structural shocks based on a set of sign restrictions. Once this is done, we find that both substitution and income effects play a relevant role for the transmission of tax rate shocks.

1.1 Relationship with the Bayesian Literature on Proxy-SVARs

As mentioned above, to the best of our knowledge, only three papers consider Proxy-SVARs under the Bayesian paradigm: [Bahaj \(2014\)](#), [Drautzburg \(2016\)](#), and [Caldara and Herbst \(2016\)](#).³ [Bahaj \(2014\)](#) and [Drautzburg \(2016\)](#) draw from the posterior distribution of the orthogonal reduced-form parameterization and then transform the draws to the structural Proxy-SVAR parameterization. Since the reduced-form parameters are restricted they both used Gibbs samplers to draw from the posterior of the reduced-form parameters; thereby the draws are not independent. More importantly, they ignore the volume element when transforming the draws into the structural Proxy-SVAR parameterization. As it will be shown in Section 5, when the volume element is ignored the order of the instruments affects the results, hence these methods are not appropriate for inference.

Finally, let's relate our paper to [Caldara and Herbst \(2016\)](#). As in our approach, this paper draws directly from the structural parameterization of the Proxy-SVAR. Nevertheless, the posterior draws are not independent and their Metropolis-Hastings sampler could become computationally inefficient compared with ours in large models. 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. Regarding the latter, researchers using our approach can also explore the

³[Jarociński and Karadi \(2018\)](#) use sign restrictions to identify shocks in a related framework. Nevertheless, while they assume that the shocks are linear combinations of the proxies, a Proxy-SVAR only assumes that the shocks are correlated with linear combinations of the proxies.

sensitivity of the results to the quality of the instruments by imposing thresholds on their reliability using additional sign restrictions.

The remainder of the paper is organized as follows. Section 2 introduces the methodology. Section 3 describes the algorithm. Section 4 shows that when identifying more than one shock using more than one external instrument additional sign or zero restrictions are necessary, and presents an algorithm to consider them. Sections 6 and 7 present two applications. Section 8 concludes. Technical details are deferred to the Appendix.

2 The Framework

This section discusses our general framework. In Section 2.1, we describe the structural parameterization of a Proxy-SVAR. In Section 2.2, we present the identification problem and the exogeneity restrictions. In Section 2.3, we explicitly specify the family of prior and posterior distributions over the structural parameterization of a Proxy-SVAR. We use this distribution to describe our algorithms because it has appealing properties. The chief among them is that it is a conjugate distribution and, hence, assigns equal prior and posterior weight to observationally equivalent Proxy-SVAR structural parameters. Even so, as mentioned in the introduction, it is important to keep in mind throughout the remainder of the paper that our methods can be used to independently draw from any posterior distribution over the structural parameterization of a Proxy-SVAR. In Section 2.4, we introduce the orthogonal-triangular-block parameterization. As will become clear later, this is a useful parameterization of a Proxy-SVAR and our algorithms will rely on it. In this section we will also introduce the mapping that we will be using to move between parameterizations.

2.1 A Proxy-SVAR

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

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

where $\tilde{\mathbf{A}}_i$ is an $\tilde{n} \times \tilde{n}$ matrix for $0 \leq i \leq p$ with $\tilde{\mathbf{A}}_0$ invertible, $\tilde{\mathbf{c}}$ is a $1 \times \tilde{n}$ row vector, and $\tilde{\boldsymbol{\varepsilon}}_t$ is conditionally standard normal with mean zero and identity variance-covariance matrix.⁴ If $\tilde{\mathbf{x}}'_t = [\tilde{\mathbf{y}}'_{t-1} \cdots \tilde{\mathbf{y}}'_{t-p} \ 1]$ and $\tilde{\mathbf{A}}'_+ = [\tilde{\mathbf{A}}'_1 \cdots \tilde{\mathbf{A}}'_p \ \tilde{\mathbf{c}}']$, Equation (1) can be more compactly written as

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

Let $\tilde{\boldsymbol{\varepsilon}}'_t = [\boldsymbol{\varepsilon}'_t \ \boldsymbol{\nu}'_t]$, where $\boldsymbol{\varepsilon}_t$ is $n \times 1$ and $\boldsymbol{\nu}_t$ is $k \times 1$. Because $\tilde{\boldsymbol{\varepsilon}}_t$ is conditionally standard normal with mean zero and identity variance-covariance matrix, we are assuming that $\boldsymbol{\nu}_t$ is uncorrelated with $\boldsymbol{\varepsilon}_t$. A Proxy-SVAR imposes that \mathbf{y}_t evolves according to

$$\mathbf{y}'_t \mathbf{A}_0 = \mathbf{x}'_t \mathbf{A}_+ + \boldsymbol{\varepsilon}'_t \text{ for } 1 \leq t \leq T, \quad (3)$$

where $\mathbf{x}'_t = [\mathbf{y}'_{t-1} \cdots \mathbf{y}'_{t-p} \ 1]$ and $\mathbf{A}'_+ = [\mathbf{A}'_1 \cdots \mathbf{A}'_p \ \mathbf{c}']$, with \mathbf{A}_i an $n \times n$ matrix for $0 \leq i \leq p$, \mathbf{A}_0 invertible, and \mathbf{c} a $1 \times n$ row vector.

Hence, $\boldsymbol{\varepsilon}_t$ are the structural shocks and $\boldsymbol{\nu}_t$ are other shocks that affect the proxies. Equation (3) implies that

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

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 $\mathbf{0}_{k \times n}$ is a $k \times n$ matrix of zeros. We call the zero restrictions on $\tilde{\mathbf{A}}_0$ and $\tilde{\mathbf{A}}_+$ the block restrictions. We call Equation (2) plus the block restrictions the structural parameterization of the Proxy-SVAR and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ such that the block restrictions hold the Proxy-SVAR structural parameters. Hence, for the set of Proxy-SVAR structural parameters the block restrictions always hold. While our Proxy-SVAR representation is very similar to the representation of [Mertens and Ravn \(2013\)](#) and [Stock and Watson \(2018\)](#), it is somewhat more general because it allows for the lags of \mathbf{y}_t to Granger cause \mathbf{m}_t .

2.2 The Identification Problem in a Proxy-SVAR

Following [Rothenberg \(1971\)](#) the Proxy-SVAR structural parameters $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent if and only if they imply the same joint distribution of $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_T$. Proposition 1 extends an insight

⁴We can always include a vector of exogenous variables, $\tilde{\mathbf{z}}_t$ of dimension $\tilde{e} \times 1$. In that case the model will be written as

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

where $\tilde{\mathbf{d}}$ is a $\tilde{e} \times \tilde{n}$ row vector.

of Rubio-Ramírez, Waggoner, and Zha (2010) to the case of Proxy-SVARs. Let $\text{diag}(X_1, \dots, X_j)$ denote a block diagonal matrix with the matrices X_1, \dots, X_j along the diagonal.

Proposition 1. *The Proxy-SVAR structural parameters $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent if and only if $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$ and $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$, for some matrix $\mathbf{Q} \in \mathcal{Q} \subset \mathcal{O}(\tilde{n})$, where $\mathcal{O}(\tilde{n})$ is the set of $\tilde{n} \times \tilde{n}$ orthogonal matrices and \mathcal{Q} is defined by*

$$\mathcal{Q} = \{\mathbf{Q} \in \mathcal{O}(\tilde{n}) \mid \mathbf{Q} = \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2), \mathbf{Q}_1 \in \mathcal{O}(n), \text{ and } \mathbf{Q}_2 \in \mathcal{O}(k)\}.$$

Proof. See Appendix A.1. □

Rubio-Ramírez, Waggoner, and Zha (2010) prove that SVARs are not identified. Corollary 1 shows that Proxy-SVARs are not identified.

Corollary 1. *A Proxy-SVAR is not identified.*

Proof. Because elements of \mathcal{Q} are block diagonal, if $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are Proxy-SVAR parameters, then so are $(\tilde{\mathbf{A}}_0 \mathbf{Q}, \tilde{\mathbf{A}}_+ \mathbf{Q})$ for all $\mathbf{Q} \in \mathcal{Q}$. By Proposition 1, $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ and $(\tilde{\mathbf{A}}_0 \mathbf{Q}, \tilde{\mathbf{A}}_+ \mathbf{Q})$ are observationally equivalent. So Proxy-SVARs are not identified. □

The identification problem in Proxy-SVARs is commonly a partial identification problem because researchers focus on identifying a subset of the Proxy-SVAR structural equations.⁵ For easy of exposition, henceforward we adopt Leeper, Sims, and Zha’s (1996) view and often talk about identifying structural shocks as equivalent to identifying structural equations because, under the Proxy-SVAR framework, each equation contains only one shock so that it is possible to directly relate equations and shocks.⁶

More specifically, the identification problem in Proxy-SVARs is typically solved by assuming that the k proxies are correlated with k structural shocks in ε_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 ε_t and the structural shocks uncorrelated with the proxies be the first $n - k$ elements of ε_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, first note that for the Proxy-SVAR

⁵A Proxy-SVAR equation is identified if for any two sets of observationally equivalent Proxy-SVAR parameters, the parameters in that equation are identical.

⁶As in Arias, Rubio-Ramírez and Waggoner (2018), the theory and the algorithms of this paper can be replicated for a Proxy-SVAR in which the structural parameters are written in terms of impulse response functions (IRFs)—i.e., the IRF parameterization (see Appendix B in Arias, Rubio-Ramírez and Waggoner (2018)). In such a case, identifying structural shocks is equivalent to identifying structural IRFs.

structural parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ it is the case that

$$\tilde{\mathbf{A}}_0^{-1} = \begin{bmatrix} \mathbf{A}_0^{-1} & -\mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1} \\ \mathbf{0}_{k \times n} & \mathbf{\Gamma}_{0,2}^{-1} \end{bmatrix}.$$

Then, note that by multiplying Equation (2) by $\tilde{\mathbf{A}}_0^{-1}$ and focusing on the last k equations we obtain

$$\mathbf{m}'_t = \tilde{\mathbf{x}}'_t \tilde{\mathbf{A}}_+ \begin{bmatrix} -\mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1} \\ \mathbf{\Gamma}_{0,2}^{-1} \end{bmatrix} - \boldsymbol{\varepsilon}'_t \mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1} + \boldsymbol{\nu}'_t \mathbf{\Gamma}_{0,2}^{-1} \text{ for } 1 \leq t \leq T.$$

It follows that

$$\mathbb{E}[\boldsymbol{\varepsilon}'_t \mathbf{m}'_t] = -\mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1}.$$

Thus, the identifying restrictions imply that the first $n - k$ rows of matrix $\mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1}$ must be zero. Taking transposes, this implies that the lower left-hand $k \times (n - k)$ block of $(\tilde{\mathbf{A}}_0^{-1})'$ must be zero. This makes clear that Proxy-SVARs are typically identified by zero restrictions on a function of the Proxy-SVAR structural parameters. That is, the exogeneity restrictions are zero restrictions on $(\tilde{\mathbf{A}}_0^{-1})'$. In addition to exogeneity restrictions, we also need the covariance matrix of the last k shocks and the k proxies, which is given by the last k rows of $-\mathbf{A}_0^{-1}\mathbf{\Gamma}_{0,1}\mathbf{\Gamma}_{0,2}^{-1}$, to be non-singular. Following the literature, we refer to this as the relevance condition. Following [Caldara and Herbst \(2016\)](#) one may want to control how strong the relevance condition is. In Section 2.3, to follow, we show how to use the priors to do so.

2.3 Priors, Posteriors, and a Useful Parameterization and Mapping

As mentioned above, we will use a restricted normal-generalized-normal distribution over the structural parameterization of the Proxy-SVAR as our prior distribution to describe our algorithms.⁷ Hence, the Proxy-SVAR structural parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ have a prior density proportional to

$NGN_{(\nu, \Phi, \Psi, \Omega)}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$, where

$$NGN_{(\nu, \Phi, \Psi, \Omega)}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \propto |\det(\tilde{\mathbf{A}}_0)|^{\nu-n} e^{-\frac{1}{2} \text{vec}(\tilde{\mathbf{A}}_0)' \Phi \text{vec}(\tilde{\mathbf{A}}_0)} e^{-\frac{1}{2} (\text{vec}(\tilde{\mathbf{A}}_+) - \Psi \text{vec}(\tilde{\mathbf{A}}_0))' \Omega^{-1} (\text{vec}(\tilde{\mathbf{A}}_+) - \Psi \text{vec}(\tilde{\mathbf{A}}_0))}. \quad (4)$$

⁷By a restricted normal-generalized-normal distribution over the structural parameterization of the Proxy-SVAR we mean a normal-generalized-normal distribution over $\mathbb{R}^{\tilde{n}^2 + \tilde{m}\tilde{n}}$ conditional on the block restrictions, where $\tilde{m} = p\tilde{n} + 1$. If there are exogenous variables, $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$.

The density is characterized by four parameters: a scalar $\nu \geq \tilde{n}$, an $\tilde{n}^2 \times \tilde{n}^2$ symmetric and positive definite matrix Φ , an $\tilde{m}\tilde{n} \times \tilde{n}^2$ matrix Ψ , and an $\tilde{m}\tilde{n} \times \tilde{m}\tilde{n}$ symmetric and positive definite matrix Ω .⁸ The normal-generalized-normal distribution over the structural parameterization is a conjugate family of distributions commonly used in the literature. For instance, when implemented with dummy variables, the Sims-Zha prior (see [Sims and Zha, 1998](#)) is also conjugate. Among others, a conjugate family has the appealing property that it assigns equal prior and posterior weight to observationally equivalent Proxy-SVAR structural parameters.

If the exogeneity restrictions hold, then we can write

$$\mathbb{E}[\mathbf{m}_t \boldsymbol{\varepsilon}'_t] = - \left(\mathbf{A}_0^{-1} \boldsymbol{\Gamma}_{0,1} \boldsymbol{\Gamma}_{0,2}^{-1} \right)' = [\mathbf{0}_{k \times (n-k)} \quad \mathbf{V}],$$

where the $k \times k$ matrix \mathbf{V} is the covariance matrix of the k proxy variables and the last k structural shocks. The relevance condition implies that \mathbf{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 (4) so that the minimum eigenvalue of the reliability matrix $\left(\boldsymbol{\Gamma}_{0,2}^{-1} \boldsymbol{\Gamma}_{0,2}^{-1} + \mathbf{V} \mathbf{V}' \right)^{-1} \mathbf{V} \mathbf{V}'$ implied by the prior distribution of the structural parameters is greater than γ . 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 \(1992\)](#).

Our objective is to independently draw from the restricted normal-generalized-normal posterior distribution over the structural parametrization of a Proxy-SVAR conditional on the exogeneity restrictions implied by Equation (4). More specifically, such posterior distribution is proportional to $NGN(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})$, where $\tilde{\nu} = T + \nu$, $\tilde{\Omega} = (\mathbf{I}_{\tilde{n}} \otimes \tilde{\mathbf{X}}' \tilde{\mathbf{X}} + \Omega^{-1})^{-1}$, $\tilde{\Psi} = \tilde{\Omega} (\mathbf{I}_{\tilde{n}} \otimes \tilde{\mathbf{X}}' \tilde{\mathbf{Y}} + \Omega^{-1} \Psi)$, and $\tilde{\Phi} = \mathbf{I}_{\tilde{n}} \otimes \tilde{\mathbf{Y}}' \tilde{\mathbf{Y}} + \Phi + \Psi' \tilde{\Omega}^{-1} \Psi - \tilde{\Psi}' \tilde{\Omega}^{-1} \tilde{\Psi}$, $\tilde{\mathbf{Y}} = [\tilde{\mathbf{y}}_1 \quad \cdots \quad \tilde{\mathbf{y}}_T]'$, and $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1 \quad \cdots \quad \tilde{\mathbf{x}}_T]'$.

[Arias, Rubio-Ramírez, and Waggoner \(2018\)](#) showed how to independently draw from any desired posterior distribution over the structural parameterization of a SVAR conditional on sign and zero restrictions.⁹ Since a Proxy-SVAR identified with exogeneity restrictions can be represented by the SVAR in Equation (2) identified with the zero restrictions on $\tilde{\mathbf{A}}_0$, $\tilde{\mathbf{A}}_+$, and $(\tilde{\mathbf{A}}_0^{-1})'$ associated with the block and the exogeneity restrictions, 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 $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$,

⁸[Arias, Rubio-Ramírez, and Waggoner \(2018\)](#) assumed a Kronecker structure for Φ , Ψ , and Ω .

⁹As it is the case here, [Arias, Rubio-Ramírez and Waggoner \(2018\)](#) use the normal-generalized-normal distribution to illustrate their methods, but their algorithm can be used to draw from any desired posterior distribution over the structural parameterization.

while the maximum number of restrictions that the aforementioned algorithm can handle on the j^{th} column of the structural parameters is $\tilde{n} - j$. So unless $p = 0$, an uninteresting case, the maximum will be exceeded for the n^{th} column, if not before. In this paper we show that the techniques in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#) can be adapted to accomplish our objective.

The idea in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#) was 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 conditional on the zero restrictions. The key to their approach is to properly account for the volume element associated with that mapping in order to characterize the proposal. This proposal was then embedded in an importance sampling algorithm.

The fact that the number of zeros is too large in a Proxy-SVAR identified with exogeneity restrictions implies that the reduced-form is restricted. This prevents us from obtaining independent draws from the orthogonal-reduced-form parameterization. Hence, instead of using the orthogonal-reduced-form parameterization, we will map independent draws from what we will call the orthogonal-triangular-block parameterization conditional on the exogeneity 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. As in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#), the key will be to properly account for the volume element in order to characterize the proposal. This proposal will be again used in an importance sampling algorithm.

2.4 The Orthogonal-Triangular-Block Parameterization and the Mapping

Let $\tilde{\Lambda}_0$ be an $\tilde{n} \times \tilde{n}$ matrix, $\tilde{\Lambda}_+$ be an $\tilde{m} \times \tilde{n}$ matrix, \mathbf{Q}_1 be an $n \times n$ orthogonal matrix, and \mathbf{Q}_2 be a $k \times k$ orthogonal matrix. The matrix $\tilde{\Lambda}_0$ is restricted to be upper-triangular with positive diagonal. The matrix $\tilde{\Lambda}'_+ = [\tilde{\Lambda}'_1 \cdots \tilde{\Lambda}'_p \mathbf{d}']$, where $\tilde{\Lambda}_i$ is $\tilde{n} \times \tilde{n}$ for $1 \leq i \leq p$ and \mathbf{d} is $1 \times \tilde{n}$, is restricted so that the lower left-hand $k \times n$ block of $\tilde{\Lambda}_i$ is zero for $1 \leq i \leq p$. We label the zero restrictions on $\tilde{\Lambda}_0$ and $\tilde{\Lambda}_+$ the triangular-block restrictions, and we call $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ such that the triangular-block restrictions hold the triangular-block parameters.

Given any values of the triangular-block parameters $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ and the orthogonal matrices $(\mathbf{Q}_1, \mathbf{Q}_2)$, we can map $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2)$ into Proxy-SVAR structural parameters $(\tilde{\mathcal{A}}_0, \tilde{\mathcal{A}}_+)$ by¹⁰

$$(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2) \xrightarrow{f} (\underbrace{\tilde{\Lambda}_0 \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2)}_{\tilde{\mathcal{A}}_0}, \underbrace{\tilde{\Lambda}_+ \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2)}_{\tilde{\mathcal{A}}_+}).$$

¹⁰The function f can be defined for all $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2)$ but will not be one-to-one over this larger set.

It is easy to verify that $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ will satisfy the block restrictions (i.e., they are Proxy-SVAR structural parameters).

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

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

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

Just as a standard SVAR can alternatively be written in the orthogonal reduced-form parameterization, the triangular-block parameters $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+)$ together with the orthogonal matrices $(\mathbf{Q}_1, \mathbf{Q}_2)$ define an equivalent parameterization of the Proxy-SVAR characterized by Equation (2) and the block restrictions. We call this alternative parameterization the orthogonal triangular-block parameterization of a Proxy-SVAR and we write the latter as follows

$$\tilde{\mathbf{y}}'_t \tilde{\mathbf{\Lambda}}_0 = \tilde{\mathbf{x}}'_t \tilde{\mathbf{\Lambda}}_+ + \tilde{\mathbf{u}}'_t \text{ for } 1 \leq t \leq T, \quad (5)$$

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

3 The Algorithm

In this section, we present an algorithm to make independent draws from the desired restricted normal-generalized-normal posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions. We achieve this goal by first independently drawing triangular-block parameters, $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+)$, using Waggoner and Zha's (2003) Gibbs sampler. Then, we show that the exogeneity restrictions are linear restrictions on the columns of the orthogonal matrix \mathbf{Q}_1 . This will allow us to use the ideas in Arias, Rubio-Ramírez and Waggoner (2018) to draw the orthogonal matrices $(\mathbf{Q}_1, \mathbf{Q}_2)$, conditional on each draw of the triangular-block parameters, such that the exogeneity restrictions hold. Then, we use f to map triangular-block parameters plus the orthogonal matrices, $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{Q}_1, \mathbf{Q}_2)$, into Proxy-SVAR structural parameters, $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$. While these independent draws are not from the desired restricted normal-generalized-normal posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity

restrictions, we will be able to numerically compute the density associated with the implied distribution. Hence, we can use those draws as an intermediate step in an importance sampler to draw from the desired posterior distribution.

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 normal-generalized-normal posterior distribution over the triangular-block parameters characterized by $NGN(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$.¹¹ This Gibbs sampler can be used to draw from a normal-generalized-normal distribution subject to linear restrictions, as long as the restrictions do not involve cross-equation restrictions and the matrices $\hat{\Phi}$, $\hat{\Psi}$, and $\hat{\Omega}$ are block diagonal.¹² Since the triangular and block restrictions on $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ are exogeneity restrictions that do not involve cross-equation restrictions, and $\hat{\Phi}$, $\hat{\Psi}$, and $\hat{\Omega}$ can be chosen to be block diagonal, the conditions for using the Gibbs sampler are satisfied. Furthermore, because $\tilde{\Lambda}_0$ 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.4, we describe how to adapt their paper to our purposes.

Often, it suffices to choose $(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ to be equal to $(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})$, the parameters associated with the desired restricted normal-generalized-normal posterior distribution over the structural parameterization of the Proxy-SVAR conditional on the exogeneity restrictions. However, sometimes this can lead to small effective sample sizes in our importance sampler. In Appendix A.5, we describe a more tailored choice of $(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ that can avoid this loss of efficiency.

3.2 Exogeneity Restrictions in the Orthogonal-Triangular-Block Parameterization

Let $\mathbf{J} = [\mathbf{0}_{k \times n} \quad \mathbf{I}_k]$ and $\mathbf{L} = [\mathbf{I}_n \quad \mathbf{0}_{n \times k}]$. Because of the arguments made in Section 2.2, if $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ are Proxy-SVAR structural parameters, the exogeneity restrictions are of the form

$$\mathbf{J}(\tilde{\mathbf{A}}_0^{-1})' \mathbf{e}_{\tilde{n},j} = \mathbf{0}_{k \times 1} \text{ for } 1 \leq j \leq n - k. \quad (6)$$

¹¹Here, by a restricted normal-generalized-normal distribution over the triangular-block parameters we mean a normal-generalized-normal distribution over $\mathbb{R}^{\tilde{n}^2 + \tilde{m}\tilde{n}}$ conditional on the triangular-block restrictions.

¹²The 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 all normal-generalized-normal distributions $NGN(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ with $\hat{\nu} \geq \tilde{n}$, $\hat{\Phi}$ block diagonal with \tilde{n} symmetric and positive definite $\tilde{n} \times \tilde{n}$ blocks, $\hat{\Psi}$ block diagonal with \tilde{n} arbitrary $\tilde{m} \times \tilde{n}$ blocks, and $\hat{\Omega}$ block diagonal with \tilde{n} symmetric and positive definite $\tilde{m} \times \tilde{m}$ blocks, conditional on the linear non-cross-equation restrictions. In Appendix A.4 we outline this Gibbs sampler in our context.

The index j stops at $n - k$ because there are no exogeneity restrictions for $n - k < j \leq \tilde{n}$. In terms of the orthogonal-triangular-block parameterization, this is equivalent to

$$\mathbf{J}(\tilde{\Lambda}_0^{-1})' \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2) \mathbf{e}_{\tilde{n},j} = \underbrace{\mathbf{J}(\tilde{\Lambda}_0^{-1})' \mathbf{L}' \mathbf{Q}_1 \mathbf{e}_{n,j}}_{\mathbf{G}(\tilde{\Lambda}_0)} = \mathbf{0}_{k \times 1} \text{ for } 1 \leq j \leq n - k, \quad (7)$$

where we have used the fact that $\tilde{\mathbf{A}}_0 = \tilde{\Lambda}_0 \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2)$. Thus, conditional on a draw of triangular-block parameters $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$, the exogeneity restrictions are linear restrictions on the columns of \mathbf{Q}_1 . We will denote the number of exogeneity restrictions on the j^{th} column of \mathbf{Q}_1 by \tilde{z}_j , which is k if $1 \leq j \leq n - k$ and is zero if $n - k < j \leq n$. As in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#), we will use this fact to draw the orthogonal matrices \mathbf{Q}_1 . As will become clear below, drawing the orthogonal matrix \mathbf{Q}_2 is simpler because its columns are unrelated to the exogeneity restrictions.

3.3 An Algorithm

Sections 3.1 and 3.2 suggest that we can devise an algorithm to make independent draws from a distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions.

Algorithm 1. *The following algorithm makes independent draws from a distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions.*

1. Draw triangular-block parameters $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ independently from the restricted $\text{NGN}(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ distribution using [Waggoner and Zha's \(2003\)](#) Gibbs sampler.
2. For $1 \leq j \leq n$, draw $\mathbf{x}_{1,j} \in \mathbb{R}^{n+1-j-\tilde{z}_j}$ independently from a standard normal distribution and set $\mathbf{w}_{1,j} = \mathbf{x}_{1,j} / \|\mathbf{x}_{1,j}\|$.
3. For $1 \leq j \leq k$, draw $\mathbf{x}_{2,j} \in \mathbb{R}^{k+1-j}$ independently from a standard normal distribution and set $\mathbf{w}_{2,j} = \mathbf{x}_{2,j} / \|\mathbf{x}_{2,j}\|$.
4. Define $\mathbf{Q}_1 = [\mathbf{q}_{1,1} \cdots \mathbf{q}_{1,n}]$ recursively by $\mathbf{q}_{1,j} = \mathbf{K}_{1,j} \mathbf{w}_{1,j}$ for any matrix $\mathbf{K}_{1,j}$ whose columns form an orthonormal basis for the null space of the $(j - 1 + \tilde{z}_j) \times n$ matrix:

$$\mathbf{M}_{1,j} = \begin{cases} \left[\begin{array}{ccc} \mathbf{q}_{1,1} & \cdots & \mathbf{q}_{1,j-1} & \mathbf{G}(\tilde{\Lambda}_0)' \end{array} \right]' & \text{for } 1 \leq j \leq n - k \\ \left[\begin{array}{ccc} \mathbf{q}_{1,1} & \cdots & \mathbf{q}_{1,j-1} \end{array} \right]' & \text{for } n - k + 1 \leq j \leq n. \end{cases}$$

5. Define $\mathbf{Q}_2 = [\mathbf{q}_{2,1} \cdots \mathbf{q}_{2,k}]$ recursively by $\mathbf{q}_{2,j} = \mathbf{K}_{2,j} \mathbf{w}_{2,j}$ for any matrix $\mathbf{K}_{2,j}$ whose columns form an

orthonormal basis for the null space of the $(j - 1) \times k$ matrix:

$$\mathbf{M}_{2,j} = \begin{bmatrix} \mathbf{q}_{2,1} & \cdots & \mathbf{q}_{2,j-1} \end{bmatrix}' \text{ for } 1 \leq j \leq k.$$

6. Set $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) = f(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{Q}_1, \mathbf{Q}_2)$.

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

In order for this algorithm to work, it must be the case that $\mathbf{M}_{i,j}$ is of full row rank; otherwise the number of columns in $\mathbf{K}_{i,j}$ will not equal the dimension of $\mathbf{w}_{i,j}$ and so the product $\mathbf{K}_{i,j}\mathbf{w}_{i,j}$ will not be defined. When $i = 1$ and $j > n - k$ or when $i = 2$ and $1 \leq j \leq k$, the matrix $\mathbf{M}_{i,j}$ will clearly be of full row rank. However, when $i = 1$ and $j \leq n - k$, the matrix $\mathbf{M}_{1,j}$ will be of full row rank if and only if $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ is of full row rank. This is because, by construction, the $\mathbf{q}_{1,1}, \dots, \mathbf{q}_{1,j-1}$ are perpendicular to $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$. Since the probability of drawing a $\mathbf{\Lambda}_0$ such that the $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ is not of full row rank is zero, we can assume without loss of generality that $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ is of full row rank.

When the exogeneity restrictions hold, the relevance condition is equivalent to $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ being of full row rank. To see this, note that $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)\mathbf{Q}_1 = \mathbf{J}(\tilde{\mathbf{A}}_0^{-1})'\mathbf{L}'$, so that by properties of the rank $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ is of full row rank if and only if $\mathbf{J}(\tilde{\mathbf{A}}_0^{-1})'\mathbf{L}'$ is of full row rank. If the exogeneity restrictions hold, then

$$\mathbb{E}[\mathbf{m}_t \boldsymbol{\varepsilon}_t'] = - \left(\mathbf{A}_0^{-1} \boldsymbol{\Gamma}_{0,1} \boldsymbol{\Gamma}_{0,2}^{-1} \right)' = \mathbf{J}(\tilde{\mathbf{A}}_0^{-1})'\mathbf{L}' = [\mathbf{0}_{k \times (n-k)} \quad \mathbf{V}], \quad (8)$$

where the $k \times k$ matrix \mathbf{V} is the covariance matrix of the k proxy variables and the last k structural shocks. So, the relevance condition, which requires \mathbf{V} to be non-singular, holds if and only if $\mathbf{G}(\tilde{\mathbf{\Lambda}}_0)$ is of full row rank. 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 γ , as explained in Section 2.3.

It is also the case that the matrix $\mathbf{K}_{i,j}$ is not unique. If the columns of $\mathbf{K}_{i,j}$ form an orthonormal basis for the null space of $\mathbf{M}_{i,j}$, then so will the columns of $\mathbf{K}_{i,j}\mathbf{X}$ for any orthogonal matrix \mathbf{X} . The particular choice of $\mathbf{K}_{i,j}$ does not make a material difference in the output of Algorithm 1, but in the next section, when we compute the density over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions implied by Algorithm 1, we will need to be able to define the function $\mathbf{K}_{i,j}$ so that it is continuously differentiable over a suitable large set. In Appendix A.2, we describe a particular choice that will work.

The independent draws of the Proxy-SVAR structural parameters conditional on the exogeneity restrictions produced by Algorithm 1 will not be from the desired restricted normal-generalized-normal posterior distribution

over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions. The density implied by Algorithm 1 will be analyzed in Section 3.4 to follow. It is important to clarify that our algorithm can handle cases in which a researcher wants to consider k instruments that are correlated with \tilde{k} shocks, with $\tilde{k} > k$. In such cases, Equations (6) and (7) will only hold for $1 < j < \tilde{k}$. 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 shocks unrestricted.

3.4 The Density Implied by Algorithm 1

Step 1 of Algorithm 1 independently draws triangular-block parameters $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ from a restricted normal-generalized-normal. Step 2 draws $\mathbf{w}_{1,j}$ from the uniform distribution on the unit sphere in $\mathbb{R}^{n+1-j-\tilde{z}_j}$. Step 3 draws $\mathbf{w}_{2,j}$ from the uniform distribution on the unit sphere in \mathbb{R}^{k+1-j} . Hence, the density over $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{w}_{1,1}, \dots, \mathbf{w}_{1,n}, \mathbf{w}_{2,1}, \dots, \mathbf{w}_{2,k})$ is proportional to a restricted normal-generalized-normal. Steps 4 and 5 map $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{w}_{1,1}, \dots, \mathbf{w}_{1,n}, \mathbf{w}_{2,1}, \dots, \mathbf{w}_{2,k})$ to $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2)$, we denote this mapping by g . Finally, Step 6 maps $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2)$ into Proxy-SVAR structural parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ using the function f . The composite mapping $f \circ g$ implied by Steps 1 through Step 6 together with Theorem 3 in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#) will be used to compute the density implied by Algorithm 1.

Since as noted before the matrix $\mathbf{K}_{i,j}$ is not unique, the function g is not uniquely defined. In Appendix A.2 we will show that g can be defined so that it is one-to-one and continuously differentiable except on a closed set of measure zero. This is what is needed to apply Theorem 3 in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#).

If \mathcal{Z} denotes the set of all Proxy-SVAR structural parameters that satisfy the exogeneity restrictions and the relevance condition, then by Theorem 3 in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#), the density over the structural parameterization of the Proxy-SVAR conditional on the exogeneity restrictions implied by Algorithm 1 is proportional to

$$NGN_{(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})}(\tilde{\Lambda}_0, \tilde{\Lambda}_+) v_{(f \circ g)^{-1}|_{\mathcal{Z}}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+),$$

where $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2) = f^{-1}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$.

3.5 An Importance Sampler

The results above show that Algorithm 1 generates independent draws from a distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions that is not equal to the desired restricted normal-generalized-normal posterior distribution over the Proxy-SVAR structural parameterization conditional

on the exogeneity restrictions. However, because Section 3.4 shows how to numerically evaluate the former distribution, we can use such a distribution as a proposal in the following importance sampler algorithm to accomplish our objective.

Algorithm 2. *The following algorithm independently draws from the restricted $NGN(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})$ posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions and the relevance condition.*

1. *Use Algorithm 1 to independently draw Proxy-SVAR structural parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ that satisfy the exogeneity restrictions and the relevance condition.*
2. *Set its importance weight to*

$$\frac{NGN_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)}{NGN_{(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)v_{(f \circ g)^{-1}|\mathcal{Z}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)},$$

where $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+, \mathbf{Q}_1, \mathbf{Q}_2) = f^{-1}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ and \mathcal{Z} denotes the set of all Proxy-SVAR structural parameters that satisfy the exogeneity restrictions and the relevance condition.

3. *Return to Step 1 until the required number of draws has been obtained.*
4. *Re-sample with replacement using the importance weights.*

Step 2 is the crucial one. The re-sampling Step 4 allows us to have unweighted and independent draws. Given the desired number of independent draws, the researcher should require enough draws from Steps 1-3 so that the effective sample size is at least as large as the number of desired independent draws. We define the effective sample size as

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

where w_i is the weight associated with the i^{th} draw and N is the total number of draws obtained in Steps 1-3.

Algorithm 2 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 such parameterization and the orthogonal triangular-block parameterization. It is also important to note that computing the volume element $v_{(f \circ g)^{-1}|\mathcal{Z}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ in Step 2 is the most expensive part in implementing Algorithm 2. The rest of Algorithm 2 is quite fast.¹³

¹³The reader should note that Algorithm 2 must be used with a normalization. Typically, SVARs are normalized by restricting the sign of the 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 set the importance weight to zero when the normalization is not satisfied, we do not explicitly state this in the algorithm.

3.6 Drawing from Any Posterior Distribution

Algorithm 2 independently draws from the restricted $NGN(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})$ posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions and the relevance condition. As mentioned several times already, the algorithm can be modified to independently draw from any desired restricted posterior distribution. When that it is the case we will need to modify Step 2 to include the density associated with desired restricted posterior distribution instead of $NGN_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$.

4 The Need for Additional Restrictions

We have described an algorithm that allows us to independently draw from the desired restricted normal-generalized-normal posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions. Next, we show that the exogeneity restrictions only allow us to categorize the Proxy-SVAR 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, we have an identification problem of the Proxy-SVAR shocks that are correlated with the proxies; unless $k = 1$. The same problem occurs within the Proxy-SVAR shocks that are not correlated with the proxies.

Proposition 2. *Let $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ be Proxy-SVAR structural parameters that also satisfy the exogeneity restrictions and the relevance condition, then $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent if and only if there exists a matrix $\mathbf{Q} \in \mathcal{X} \subset \mathcal{Q} \subset \mathcal{O}(\tilde{n})$ such that $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$ and $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$, where \mathcal{X} is defined by*

$$\mathcal{X} = \{\mathbf{Q} \in \mathcal{Q} \mid \mathbf{Q} = \text{diag}(\mathbf{Q}_3, \mathbf{Q}_4, \mathbf{Q}_5), \mathbf{Q}_3 \in \mathcal{O}(n-k), \mathbf{Q}_4 \in \mathcal{O}(k), \text{ and } \mathbf{Q}_5 \in \mathcal{O}(k)\}.$$

Proof. See Appendix A.3. □

Proposition 2 tells us that we need additional identification restrictions to identify the structural shocks within the set of structural shocks that are correlated with the proxies. In addition, the proposition makes clear that if we want to identify structural shocks within the set of the structural shocks that are not correlated with the proxies, then we also need additional identification restrictions. This is easy to see because \mathbf{Q}_3 will rotate the columns of the Proxy-SVAR structural parameters associated with the structural shocks that are not correlated with the proxies while \mathbf{Q}_4 will rotate the Proxy-SVAR equations associated with the structural

shocks that are correlated with the proxies.¹⁴ The additional restrictions can be sign and/or zero restrictions. Since imposing additional sign restrictions is straightforward, we will first focus on additional zero restrictions by showing how to modify Algorithm 1 to incorporate them. Then, we will show how to modify Algorithm 2 to consider both additional sign and zero restrictions.

Often, one is interested only in 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 exactly identify the structural shock correlated with the proxy. The following corollary of Proposition 2 formalizes this result.

Corollary 2. *Let $k = 1$ and let $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ be Proxy-SVAR structural parameters that also satisfy the exogeneity restrictions and the relevance condition. Then, the last column of $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ is identified up to a sign.*

Proof. If $k = 1$, then we have $\mathbf{Q}_4 = 1$ and the n^{th} column of \mathbf{Q} is equal to $(0 \ \dots \ 0 \ \pm 1)'$. □

Notwithstanding, the theory and techniques that we develop apply to additional sign and zero restrictions on any function $\mathbf{F}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ from the set of Proxy-SVAR structural parameters to the set of $\tilde{r} \times \tilde{n}$ matrices that satisfies the condition $\mathbf{F}(\tilde{\mathbf{A}}_0 \tilde{\mathbf{Q}}, \tilde{\mathbf{A}}_+ \tilde{\mathbf{Q}}) = \mathbf{F}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \tilde{\mathbf{Q}}$, for every $\tilde{\mathbf{Q}} \in \mathcal{O}(\tilde{n})$.¹⁵

4.1 Additional Zero Restrictions

To set the notation, let \mathbf{Z}_j be a $z_j \times \tilde{r}$ matrix of full row rank, where $0 \leq z_j \leq n - k - j$ for $1 \leq j \leq n - k$ and $0 \leq z_j \leq n - j$ for $n - k + 1 \leq j \leq n$. In particular, we assume that the additional zero restrictions in the Proxy-SVAR structural parameterization are of the form $\mathbf{Z}_j \mathbf{F}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \mathbf{e}_{\tilde{n},j} = \mathbf{0}_{z_j,1}$ for $1 \leq j \leq n$. The fact that $0 \leq z_j \leq n - k - j$ for $1 \leq j \leq n - k$ reflects that we have already imposed the exogeneity restrictions on the first $n - k$ structural shocks. Note that we are identifying only $\boldsymbol{\varepsilon}_t$. In principle, one could also use the additional zero restrictions to identify $\boldsymbol{\nu}_t$, but we do not believe that is of interest.

From the definition of f and the fact that

$$\mathbf{F}(\tilde{\mathbf{A}}_0 \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2), \tilde{\mathbf{A}}_+ \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2)) = \mathbf{F}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2),$$

the zero restrictions in the orthogonal-triangular-block parameterization are

$$\mathbf{Z}_j \mathbf{F}(f(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{Q}_1, \mathbf{Q}_2)) \mathbf{e}_{\tilde{n},j} = \mathbf{Z}_j \mathbf{F}(f(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{I}_n, \mathbf{I}_k)) \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2) \mathbf{e}_{\tilde{n},j} = \mathbf{0}_{z_j \times 1} \text{ for } 1 \leq j \leq n.$$

¹⁴Note that \mathbf{Q}_3 and \mathbf{Q}_4 are related to \mathbf{Q}_1 introduced in Proposition 1 while \mathbf{Q}_5 is related to \mathbf{Q}_2 of the same proposition.

¹⁵In addition, a regularity condition on \mathbf{F} is needed. For instance, it suffices to assume that \mathbf{F} is differentiable and that its derivative is of full row rank.

This last equation can be written as

$$\underbrace{\mathbf{Z}_j \mathbf{F}(f(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{I}_n, \mathbf{I}_k)) \mathbf{L}' \mathbf{Q}_1}_{\mathbf{G}_{1,j}(\tilde{\Lambda}_0, \tilde{\Lambda}_+)} e_{n,j} = \mathbf{0}_{z_j \times 1} \text{ for } 1 \leq j \leq n.$$

This allows us to modify Algorithm 1 to consider the additional zero restrictions as follows.

Algorithm 3. *The following algorithm makes independent draws from a distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions and the additional zero restrictions.*

1. Draw triangular-block parameters $(\tilde{\Lambda}_0, \tilde{\Lambda}_+)$ independently from the restricted $NGN(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ distribution using Waggoner and Zha's (2003) Gibbs sampler.
2. For $1 \leq j \leq n$, draw $\mathbf{x}_{1,j} \in \mathbb{R}^{n+1-j-\tilde{z}_j-z_j}$ independently from a standard normal distribution and set $\mathbf{w}_{1,j} = \mathbf{x}_{1,j} / \|\mathbf{x}_{1,j}\|$.
3. For $1 \leq j \leq k$, draw $\mathbf{x}_{2,j} \in \mathbb{R}^{k+1-j}$ independently from a standard normal distribution and set $\mathbf{w}_{2,j} = \mathbf{x}_{2,j} / \|\mathbf{x}_{2,j}\|$.
4. Define $\mathbf{Q}_1 = [\mathbf{q}_{1,1} \cdots \mathbf{q}_{1,n}]$ recursively by $\mathbf{q}_{1,j} = \mathbf{K}_{1,j} \mathbf{w}_{1,j}$ for any matrix $\mathbf{K}_{1,j}$ whose columns form an orthonormal basis for the null space of the $(j-1+\tilde{z}_j+z_j) \times n$ matrix:

$$\mathbf{M}_{1,j} = \begin{cases} \left[\begin{array}{cccc} \mathbf{q}_{1,1} & \cdots & \mathbf{q}_{1,j-1} & \mathbf{G}(\tilde{\Lambda}_0)' & \mathbf{G}_{1,j}(\tilde{\Lambda}_0, \tilde{\Lambda}_+)' \end{array} \right]' & \text{for } 1 \leq j \leq n-k \\ \left[\begin{array}{cccc} \mathbf{q}_{1,1} & \cdots & \mathbf{q}_{1,j-1} & \mathbf{G}_{1,j}(\tilde{\Lambda}_0, \tilde{\Lambda}_+)' \end{array} \right]' & \text{for } n-k+1 \leq j \leq n. \end{cases}$$

5. Define $\mathbf{Q}_2 = [\mathbf{q}_{2,1} \cdots \mathbf{q}_{2,k}]$ recursively by $\mathbf{q}_{2,j} = \mathbf{K}_{2,j} \mathbf{w}_{2,j}$ for any matrix $\mathbf{K}_{2,j}$ whose columns form an orthonormal basis for the null space of the $(j-1) \times k$ matrix:

$$\mathbf{M}_{2,j} = \left[\begin{array}{ccc} \mathbf{q}_{2,1} & \cdots & \mathbf{q}_{2,j-1} \end{array} \right]' \text{ for } 1 \leq j \leq k.$$

6. Set $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) = f(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2)$.
7. Return to Step 1 until the required number of draws has been obtained.

As was the case with Algorithm 1, the independent draws of the Proxy-SVAR structural parameters produced by Algorithm 3 will not be from the desired restricted normal-generalized-normal posterior distribution over the structural parameterization of a Proxy-SVAR conditional on the exogeneity restrictions and the additional zero restrictions. If \mathcal{X} denotes the set of all Proxy-SVAR structural parameters that satisfy the exogeneity restrictions, the relevance condition, and the additional zero restrictions, because of the results in Section 3.4,

the density over the structural parameterization implied by Algorithm 3 is proportional to

$$NGN_{(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) v_{(f \circ g)^{-1}|\mathcal{X}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+),$$

where $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+, \mathbf{Q}_1, \mathbf{Q}_2) = f^{-1}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ and g has been adapted to consider the new matrices $\mathbf{M}_{1,j}$ and $\mathbf{M}_{2,j}$.

Thus, we can use this density as our new proposal density. Thus, we next show how to adapt the importance sampler described in Algorithm 2 to accomplish our objective of making independent draws from the desired restricted normal-generalized-normal posterior distribution conditional on the exogeneity restrictions, the relevance condition, and the additional sign and zero restrictions.

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

4.2 Additional Sign and Zero Restrictions

We assume that the additional sign restrictions in the Proxy-SVAR structural parameterization are of the form $\tilde{\mathbf{F}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) > \mathbf{0}_{s,1}$, where $\tilde{\mathbf{F}}$ is continuous and $0 \leq s$ is the number of sign restrictions. As explained in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#), because $\tilde{\mathbf{F}}$ is continuous, the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions and the additional sign and zero restrictions will be open in the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions and the additional zero restrictions. Thus, if the additional sign restrictions are non-degenerate, so that there is at least one value of the Proxy-SVAR structural parameters satisfying the exogeneity restrictions and the additional sign and zero restrictions, then the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions and the additional sign and zero restrictions will be of positive volume measure in the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions and the additional zero restrictions.¹⁶ This justifies algorithms of the type described below to accomplish our objective of making independent draws from the desired restricted normal-generalized-normal posterior distribution conditional on the exogeneity restrictions, the relevance condition, and the additional sign and zero restrictions. The next algorithm is a modification of Algorithm 2 to consider additional sign and zero restrictions when using Algorithm 3 instead of Algorithm 1 in Step 1.

¹⁶The volume measure is defined in [Arias, Rubio-Ramírez, and Waggoner \(2018\)](#).

Algorithm 4. *The following algorithm independently draws from the desired restricted $NGN(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})$ posterior distribution over the Proxy-SVAR structural parameterization conditional on the exogeneity restrictions, the relevance condition, and the additional sign and zero restrictions.*

1. *Use Algorithm 3 to independently draw Proxy-SVAR structural parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ that satisfy the exogeneity restrictions, the relevance condition, and the additional zero restrictions.*
2. *If $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ satisfies the sign restrictions, set its importance weight to*

$$\frac{NGN_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)}{NGN_{(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega})}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)v_{(f \circ g)^{-1}|\mathcal{X}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)},$$

where $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+, \mathbf{Q}_1, \mathbf{Q}_2) = f^{-1}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ and where \mathcal{X} denotes the set of all Proxy-SVAR structural parameters that satisfy the exogeneity restrictions, the relevance condition, and the additional zero restrictions.

3. *Return to Step 1 until the required number of draws has been obtained.*
4. *Re-sample with replacement using the importance weights.*

As was the case with Algorithm 2, computing the volume element $v_{(f \circ g)^{-1}|\mathcal{X}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ in Step 2 is the most expensive part in implementing Algorithm 4. The rest of Algorithm 4 is quite fast. But the reader should note that we do not need to compute the volume element for all the draws, only for those that satisfy the sign restrictions. Finally, as before we also need the relevance condition to hold.

5 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 Algorithm 2 and simply use Algorithm 1. The reader should be aware of at least one dangerous feature of the posterior 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. Algorithm 2 does not suffer from this problem.¹⁷

To illustrate that Algorithm 1 is not invariant to a reordering of the instruments let us consider a Proxy-SVAR with three variables, one lag, and no constant. Suppose further that Researcher 1 wants to identify two structural shocks using two external instruments, i.e., $k = 2$. Let us assume that Researcher 1 orders the variables as follows $\tilde{\mathbf{y}}'_t = [\mathbf{y}'_t, \mathbf{m}'_t]$ and imposes the exogeneity restrictions on the Proxy-SVAR structural parameters.

¹⁷For ease of exposition we compare Algorithms 1 and 2, but it should be clear that the same applies to a comparison between Algorithm 3 and 4.

Let $\mathcal{Z}_{\mathcal{R}_1}$ denote the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions imposed by Researcher 1 and assume that she uses Algorithm 1. As a consequence, Researcher 1 draws from the density $NGN_{(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})}(\tilde{\Lambda}_0, \tilde{\Lambda}_+)v_{(f \circ g)^{-1}|\mathcal{Z}_{\mathcal{R}_1}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ over the Proxy-SVAR structural parameters, where $(\tilde{\Lambda}_0, \tilde{\Lambda}_+, \mathbf{Q}_1, \mathbf{Q}_2) = f^{-1}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$. For simplicity, let $\hat{\nu} = \tilde{n}$, $\hat{\Phi} = \mathbf{I}_{\tilde{n}}$, $\hat{\Psi} = \mathbf{0}_{\tilde{n}}$, and $\hat{\Omega} = \mathbf{I}_{\tilde{n}}$. Now assume that there is a second researcher called Researcher 2 who wants to identify the same shocks than Researcher 1 by imposing the same exogeneity restrictions. However, in contrast to Researcher 1, Researcher 2 switches the order of the instruments and consequently sets $\tilde{\mathbf{y}}'_{\mathcal{R}_2,t} = [\mathbf{y}'_t, \mathbf{m}'_t \mathbf{P}_m]$, where

$$\mathbf{P}_m = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Let $\mathcal{Z}_{\mathcal{R}_2}$ denote the set of all Proxy-SVAR structural parameters satisfying the exogeneity restrictions imposed by Researcher 2. If $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \in \mathcal{Z}_{\mathcal{R}_1}$, then it is easy to see that $(\tilde{\mathbf{P}}\tilde{\mathbf{A}}_0\tilde{\mathbf{P}}', \tilde{\mathbf{P}}\tilde{\mathbf{A}}_+\tilde{\mathbf{P}}') \in \mathcal{Z}_{\mathcal{R}_2}$, where

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n \times k} \\ \mathbf{0}_{k \times n} & \mathbf{P}_m \end{bmatrix}.$$

Furthermore, since $\tilde{\mathbf{y}}'_{\mathcal{R}_2,t} = \tilde{\mathbf{y}}'_t \tilde{\mathbf{P}}'$, $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ and $(\tilde{\mathbf{P}}\tilde{\mathbf{A}}_0\tilde{\mathbf{P}}', \tilde{\mathbf{P}}\tilde{\mathbf{A}}_+\tilde{\mathbf{P}}')$ are observationally equivalent. To illustrate that the ordering of the instruments affects inference, we make ten draws from $\mathcal{Z}_{\mathcal{R}_1}$ using Algorithm 1. Then, for each draw we use $\tilde{\mathbf{P}}$ to compute $(\tilde{\mathbf{P}}\tilde{\mathbf{A}}_0\tilde{\mathbf{P}}', \tilde{\mathbf{P}}\tilde{\mathbf{A}}_+\tilde{\mathbf{P}}')$. For each draw and its respective transformation we compute the ratio

$$\frac{v_{(f \circ g)^{-1}|\mathcal{Z}_{\mathcal{R}_1}}(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)}{v_{(f \circ g)^{-1}|\mathcal{Z}_{\mathcal{R}_2}}(\tilde{\mathbf{P}}\tilde{\mathbf{A}}_0\tilde{\mathbf{P}}', \tilde{\mathbf{P}}\tilde{\mathbf{A}}_+\tilde{\mathbf{P}}')}.$$

The results are reported in Table 1. If the order of the instruments did not matter, then all the entries of the table would be equal to 1, which is clearly not the case.

Table 1: Ratio of densities

Draw	1	2	3	4	5	6	7	8	9	10
	0.64	0.49	2.67	0.16	1.17	1.35	0.92	0.64	0.28	0.06

Ratio of densities for ten draws of the structural parameters using a different ordering for the instruments. The ratio is computed with the density over $\mathcal{Z}_{\mathcal{R}_1}$ on the numerator and the density over $\mathcal{Z}_{\mathcal{R}_2}$ on the denominator.

This is a disturbing result because it means that Bayesian inference based on Algorithm 1 hinges on an arbitrary decision regarding the order in which a researcher sets her instruments in \mathbf{m}_t . A similar argument can be made regarding the order of the variables within \mathbf{y}_t .

6 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 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.6](#). 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.¹⁸

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 $\nu = \tilde{n} = 7$, $\Phi^{-1} = \mathbf{0}_{\tilde{n}, \tilde{n}}$, and $\Psi = \mathbf{0}_{\tilde{m}\tilde{n}, \tilde{n}^2}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\Phi} = \tilde{\Phi}$, $\hat{\Phi} = \tilde{\Phi}$ and $\hat{\Omega}^{-1} = \tilde{\Omega}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We also set $\gamma = 0.2$.¹⁹

Let $\boldsymbol{\varepsilon}_t^{TFP}$ be a vector containing the consumption and investment TFP shocks, i.e. $\boldsymbol{\varepsilon}_t^{TFP} = [\boldsymbol{\varepsilon}_{C,t} \ \boldsymbol{\varepsilon}_{I,t}]'$, and let $\boldsymbol{\varepsilon}_t^O$ be a vector containing all other structural shocks. The first set of identification assumptions are

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

where $\mathbf{m}'_t = [\mathbf{m}_{C,t} \ \mathbf{m}_{I,t}]$ are the proxies for the consumption and investment TFP shocks. As mentioned in [Section 4](#), 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 sign restrictions $\mathbb{E}[\mathbf{m}_{C,t} \boldsymbol{\varepsilon}_{C,t}] > 0$, $\mathbb{E}[\mathbf{m}_{I,t} \boldsymbol{\varepsilon}_{I,t}] > 0$, $\mathbb{E}[\mathbf{m}_{C,t} \boldsymbol{\varepsilon}_{C,t}] > \mathbb{E}[\mathbf{m}_{C,t} \boldsymbol{\varepsilon}_{I,t}]$ and $\mathbb{E}[\mathbf{m}_{I,t} \boldsymbol{\varepsilon}_{I,t}] > \mathbb{E}[\mathbf{m}_{I,t} \boldsymbol{\varepsilon}_{C,t}]$ on the entries of \mathbf{V} . If we order the two structural shocks of interest last, this implies setting $s = 4$,

$$\mathbf{S}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{S}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

¹⁸We downloaded the proxies from Kurt Lunsford's website at <https://sites.google.com/site/kurtglunsford/research>.

¹⁹Clearly, 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.

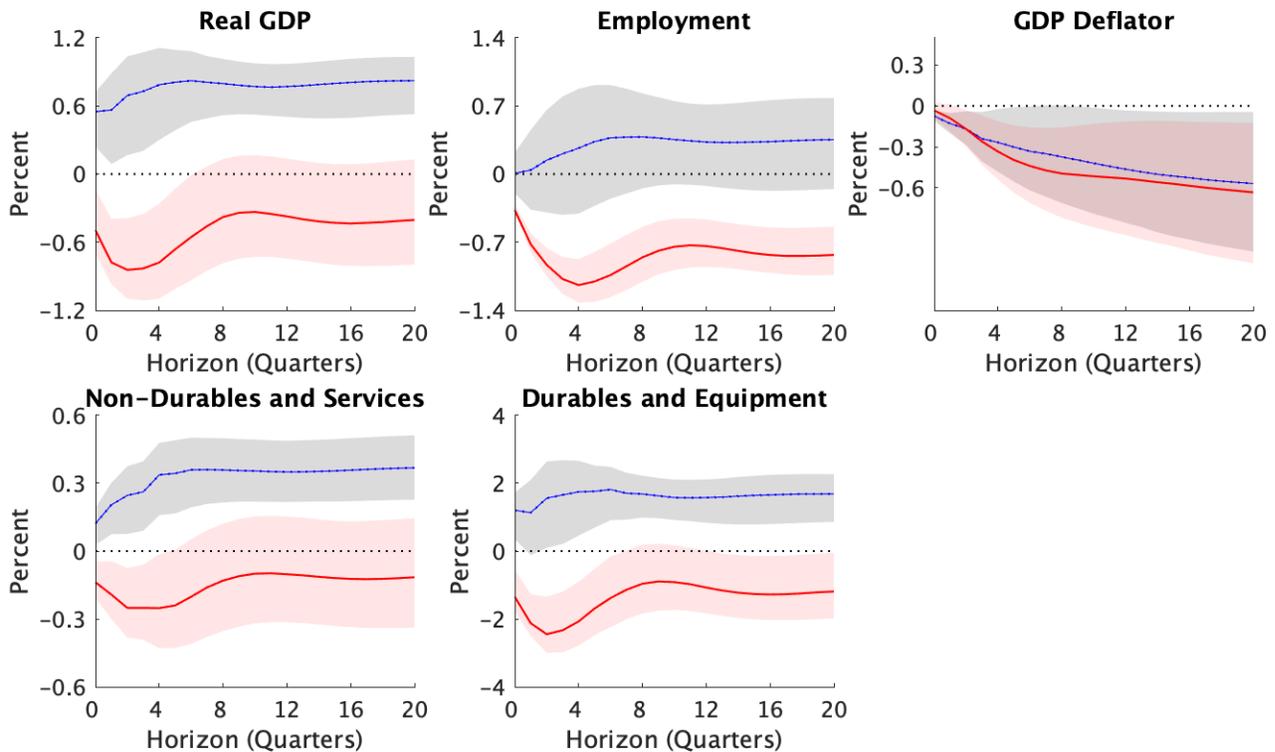


Figure 1: IRFs to a positive one standard deviation consumption and investment TFP shocks. The blue solid-dotted curves represent the point-wise posterior medians and the gray shaded areas represent the 68 percent equal-tailed point-wise probability bands to a consumption TFP shock. The red solid curves 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. The figure is based on 10,000 independent effective draws obtained using Algorithm 4.

and

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

Since in this application there are sign and exogeneity restrictions involved, we use Algorithm 4. More specifically, we obtain an effective sample size of 10,000 draws of the Proxy-SVAR structural parameters satisfying the sign and exogeneity restrictions. Figure 1 shows the IRFs to a positive one standard deviation consumption and investment TFP shocks, respectively.²⁰ 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.

²⁰While Lunsford (2016) reports the IRFs of the endogenous variables in the SVAR, we report the cumulative IRFs for easy of exposition.

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 \(2012\)](#).

Finally, 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.7](#). In [Appendix A.8](#) we show that these tests imply that the finite variance requirement holds for the application analyzed in this section.

7 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. The study under analysis works in the frequentist paradigm and its 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 Bayesian approach will basically replicate [Mertens and Montiel-Olea's \(2018\)](#) first finding. 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 as described in Section 4.

Our analysis shows that, although an average marginal tax rate shock causes larger effects on economic activity than an average tax rate shock, the latter do have positive effects. Thus, it is too early to rule out the income effects of unanticipated average tax rate changes. In contrast, our analysis shows that [Mertens and Montiel-Olea's \(2018\)](#) third finding is quite robust to using structural shocks instead of counterfactual tax experiments. The fact that we corroborate their third finding but question their second one illustrates that our approach is not only potentially useful for researchers skeptical of counterfactuals, but also for those embracing them. The former can use our method to parse out the margins through which fiscal policy operates relying on fundamental shocks as defined in [Ramey \(2016\)](#). The latter can use it to explore the robustness of their results to fundamental structural shocks.

In this application we use the dataset built by [Mertens and Montiel-Olea \(2018\)](#), which we downloaded from Karel Mertens's website.²¹ A detailed description of the dataset can be found in Appendix A of [Mertens and Montiel-Olea \(2018\)](#).

7.1 Macroeconomic Responses to Marginal Tax Rates

Let's begin by revisiting the first finding. 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. 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.²² 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.²³ Accordingly, the identification of the AMTR shock is achieved by assuming that the proxy is only correlated with the AMTR.²⁴ Following [Mertens and Montiel-Olea \(2018\)](#), 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 $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$.

²¹Link to the dataset: https://karelmertens.com.files.wordpress.com/2018/01/data_mmo.xlsx.

²²Net-of-tax rate is defined as 1 minus the AMTR.

²³The net-of-tax rate is based on [Barro and Redlick \(2011\)](#); for additional details on the construction of the endogenous variables and the proxy, we refer the reader to [Mertens and Montiel-Olea's \(2018\)](#) paper.

²⁴The sign of the AMTR shock is pinned-down by assuming that the IRF of AMTR is negative in response to a negative AMTR shock.

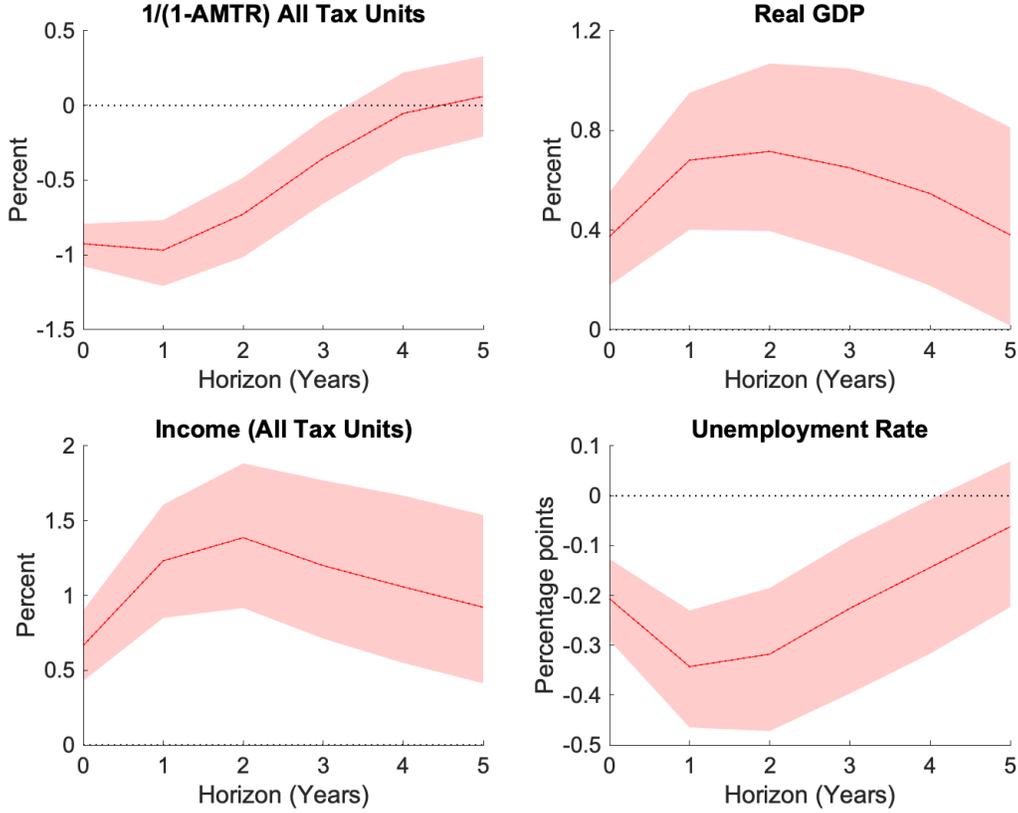


Figure 2: IRFs to a positive AMTR shock (rate cut). The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent equal-tailed point-wise probability bands. The figure is based on 10,000 independent effective draws obtained using Algorithm 2.

We set $\nu = \tilde{n}$, $\Phi = \mathbf{0}_{\tilde{n}, \tilde{n}}$, $\Psi = \mathbf{0}_{\tilde{m}\tilde{n}, \tilde{n}^2}$ and $\Omega^{-1} = \mathbf{0}_{\tilde{m}\tilde{n}, \tilde{m}\tilde{n}}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\Phi} = \tilde{\Phi}$, $\hat{\Psi} = \tilde{\Psi}$ and $\hat{\Omega}^{-1} = \tilde{\Omega}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We also set $\gamma = 0.2$. We use Algorithm 2 to obtain an effective sample size of 10,000 draws of Proxy-SVAR structural parameters satisfying the exogeneity restrictions.

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. The shock lowers the tax rate for at least 4 years. Clearly, the positive and sizable IRFs of real GDP—about 0.7 percent after one year—and the negative and sizable IRFs of the unemployment rate—about negative 0.35 percentage points after one year—coincide with a positive and sizable response of income peaking at nearly 1.5 percent two years after the shock. Therefore, our results clearly align with those reported in Figure 5 of Mertens and Montiel-Olea (2018). In Appendix A.9 we show that Koopman, Shephard and Creal’s (2009) tests validate the finite variance requirement for the reliable use of Algorithm 2.

The results reported above imply that AMTR shocks have sizable effects on economic activity. Even so, as

highlighted by [Mertens and Montiel-Olea \(2018\)](#), these results leave important questions unanswered. Chief among them are whether the stimulate effects reported above operate through substitution effects rather than income effects and whether changes in tax policy targeting top incomes have larger effects than changes in fiscal policy targeting lower incomes. We turn to this in the next two sections.

7.2 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 7.1 by adding the log ATR as an endogenous variable and they jointly identify two personal income tax rate shocks using two proxies.²⁵ 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 in two steps. First, [Mertens and Montiel-Olea \(2018\)](#) assume that the proxies are only correlated with the tax rate shocks. As shown in Proposition 2, this would only separate the tax rates shocks from the rest of the structural shocks but would not allow the researcher to identify the two tax rate shocks individually. Thus, they rely on two counterfactual tax experiments to establish causal effects. In their first counterfactual, they consider an unanticipated change in marginal tax rate that does not have a direct effect on the average tax rate. [Mertens and Montiel-Olea \(2018\)](#) argue that the Tax Reform Act of 1986 is the closest historical equivalent for this counterfactual. We will refer to this counterfactual as the AMTR counterfactual. In their second counterfactual, they consider an unanticipated change in the average tax rate that does not have a direct effect on the marginal tax rate. [Mertens and Montiel-Olea \(2018\)](#) argue that many of the postwar tax policy reforms are of this type. We will refer to this counterfactual as the ATR counterfactual. When analyzing [Mertens and Montiel-Olea’s \(2018\)](#) results we will consider the AMTR counterfactual to analyze the effects of marginal tax rate cuts and we will consider the ATR counterfactual to analyze the effects of average tax rate cuts.

Turning to the specification of the proxy-SVAR. In this application $T = 65$, $n = 10$, $k = 2$, $p = 2$, $\tilde{e} = 2$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$. We set $\nu = \tilde{n}$, $\Phi = \mathbf{0}_{\tilde{n}, \tilde{n}}$, $\Psi = \mathbf{0}_{\tilde{m}\tilde{n}, \tilde{n}^2}$ and $\Omega^{-1} = \mathbf{0}_{\tilde{m}\tilde{n}, \tilde{m}\tilde{n}}$ to characterize our prior over the Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\Phi} = \hat{\Phi}_0 \neq \tilde{\Phi}$, $\hat{\Psi} = \tilde{\Psi}$ and $\hat{\Omega}^{-1} = \tilde{\Omega}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We choose $\hat{\Phi}_0$ to maximize the efficiency of the importance sampler.²⁶ We also set $\gamma = 0.2$. We use Algorithm 4 to obtain an effective sample size of

²⁵ATR is defined as total revenue and contributions as a ratio of the [Piketty and Saez \(2003\)](#) measure of aggregate market income.

²⁶If we set $\hat{\Phi} = \tilde{\Phi}$ the algorithm becomes very inefficient. The basic description of the approach used for the selection of $\hat{\Phi}_0$ is described in Appendix A.5.

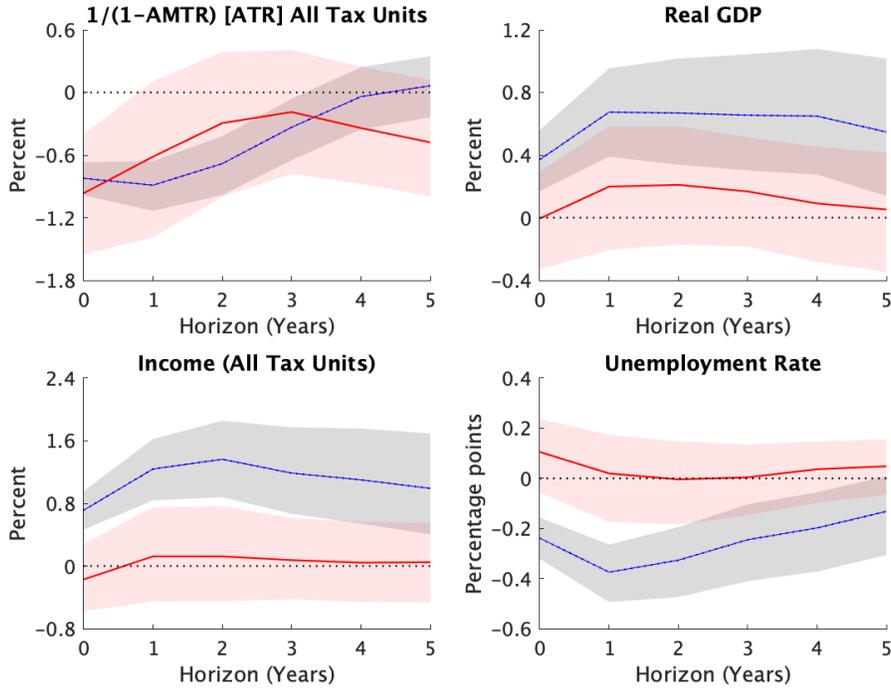
10,000 draws.

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. Essentially, Figure 3a shows that Mertens and Montiel-Olea’s (2018) findings regarding the effects of their counterfactuals can be supported when implemented from a Bayesian perspective. As the reader can see, the panel closely resemble the IRFs reported in Panels (B) and (C) of Figure 10 of Mertens and Montiel-Olea (2018) and it is very easy to conclude that income, GDP, and the unemployment rate only react to AMTR cuts. Overall, this figure justifies the following claims: “*There is, on the other hand, no evidence for any effect on incomes when average tax rates decline but marginal rates do not*” (Mertens and Montiel-Olea, 2018, page 3), 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 35). In other words, Figure 3a justifies to conclude that substitution effects rather than income effects are important for the transmission of unanticipated personal income tax policy changes in the U.S. economy post World War II.

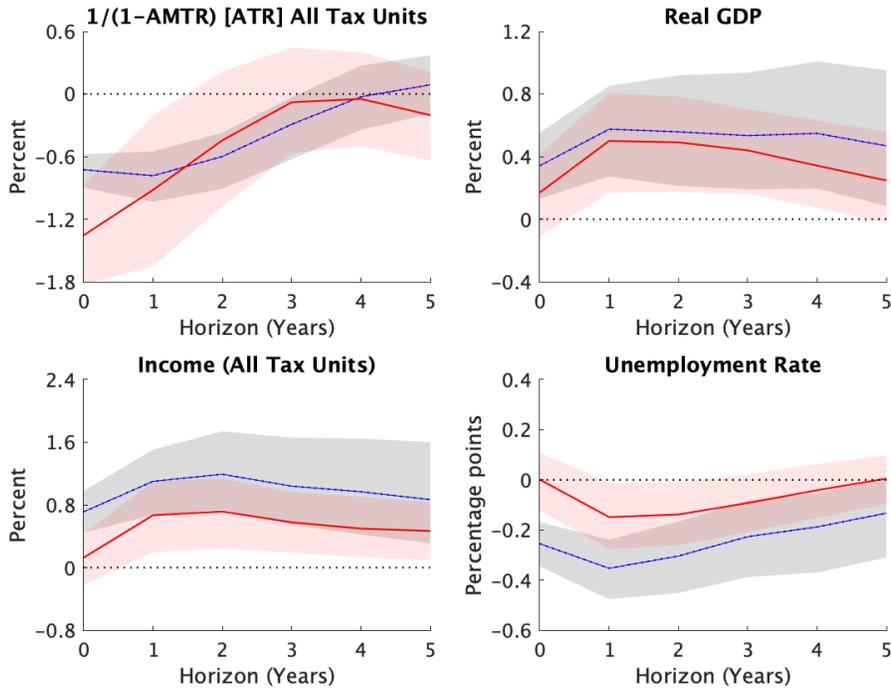
Researchers familiar with SVARs analysis may instead want to report the causal effects of each of the two structural shocks which could naturally be labeled marginal tax rate shock and average tax rate shock. For this reason, we now complement Mertens and Montiel-Olea’s (2018) results by identifying 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{F} and matrices S_j very similar to the ones described in Section 6. 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 shocks are also positive and similar. Turning to the unemployment rate, the IRF to both shocks are broadly similar and mostly negative. The differences between



(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 counterfactuals. 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 shocks. The figure is based on 10,000 independent effective draws obtained using Algorithm 4.

the results reported in Figures 3a and 3b are confirmed when analyzing Table 2.

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 a 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 same 10,000 independent effective draws obtained using Algorithm 4 used in Figure 3.

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 AMTR (ATR) to an AMTR (ATR) shock. Accordingly, these elasticities can be interpreted as the percent change in income and real GDP with respect to a 1 percentage point change in the pertinent tax rate induced by a counterfactual or shock. The reader can see that, when using 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. Although about three times lower than the ones associated with the AMTR, the short-run elasticities of income and real GDP to ATR are positive.

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. In Appendix [A.10](#) we show that [Koopman, Shephard and Creal’s \(2009\)](#) tests validate the finite variance requirement for the reliable use of Algorithm 4.

7.3 Marginal Rates 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 7.1 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.²⁷ 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 7.2, [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 first 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. We will refer to this counterfactual as the top 1 percent counterfactual. In their second 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. We will refer to this counterfactual as the bottom 99 percent counterfactual. When analyzing [Mertens and Montiel-Olea’s \(2018\)](#) results we will consider the top 1 percent counterfactual to analyze the effects of marginal tax rate cuts for taxpayers at the top 1 percent of the income distribution and we will consider the bottom 99 percent counterfactual to analyze the effects of marginal tax rate cuts for taxpayers at the bottom 99 percent of the income distribution.

Before moving to the results, we note that in this application $T = 65$, $n = 11$, $k = 2$, $p = 2$, $\tilde{e} = 4$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$. We set $\nu = \tilde{n}$, $\Phi = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\Psi = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\Omega^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the

²⁷In 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 \(2012\)](#).

Proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\Phi} = \hat{\Phi}_0 \neq \tilde{\Phi}$, $\hat{\Psi} = \tilde{\Psi}$ and $\hat{\Omega}^{-1} = \tilde{\Omega}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We choose $\hat{\Phi}_0$ to maximize the efficiency of the importance sampler.²⁸ We also set $\gamma = 0.2$. We use Algorithm 4 to obtain an effective sample size of 10,000.

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’s (2018) exercise from a Bayesian perspective. As the reader can see, the IRFs associated with the top 1 percent counterfactual resemble the IRFs reported in Figure XI of 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. Furthermore, these positive effects disappear after five years. Overall, this figure supports Mertens and Montiel-Olea’s (2018) conclusions regarding the effects of exogenous tax rate cuts for the top 1 percent. 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 from a Bayesian perspective there is 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 once we focus on the identification of fundamental structural shocks as done in Section 7.1.

In particular, we use a set of sign restrictions to identify top 1 percent and bottom 99 percent AMTR shocks analogous to the one used to study AMTR and ATR shocks. When using such an approach instead of the counterfactuals we are able to confirm Mertens and Montiel-Olea’s (2018) conclusions from a Bayesian perspective. But, before discussing the results, let us describe our sign restrictions.

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*

²⁸If we set $\hat{\Phi} = \tilde{\Phi}$ the algorithm becomes very inefficient. The basic description of the approach used for the selection of $\hat{\Phi}_0$ is described in Appendix A.5.

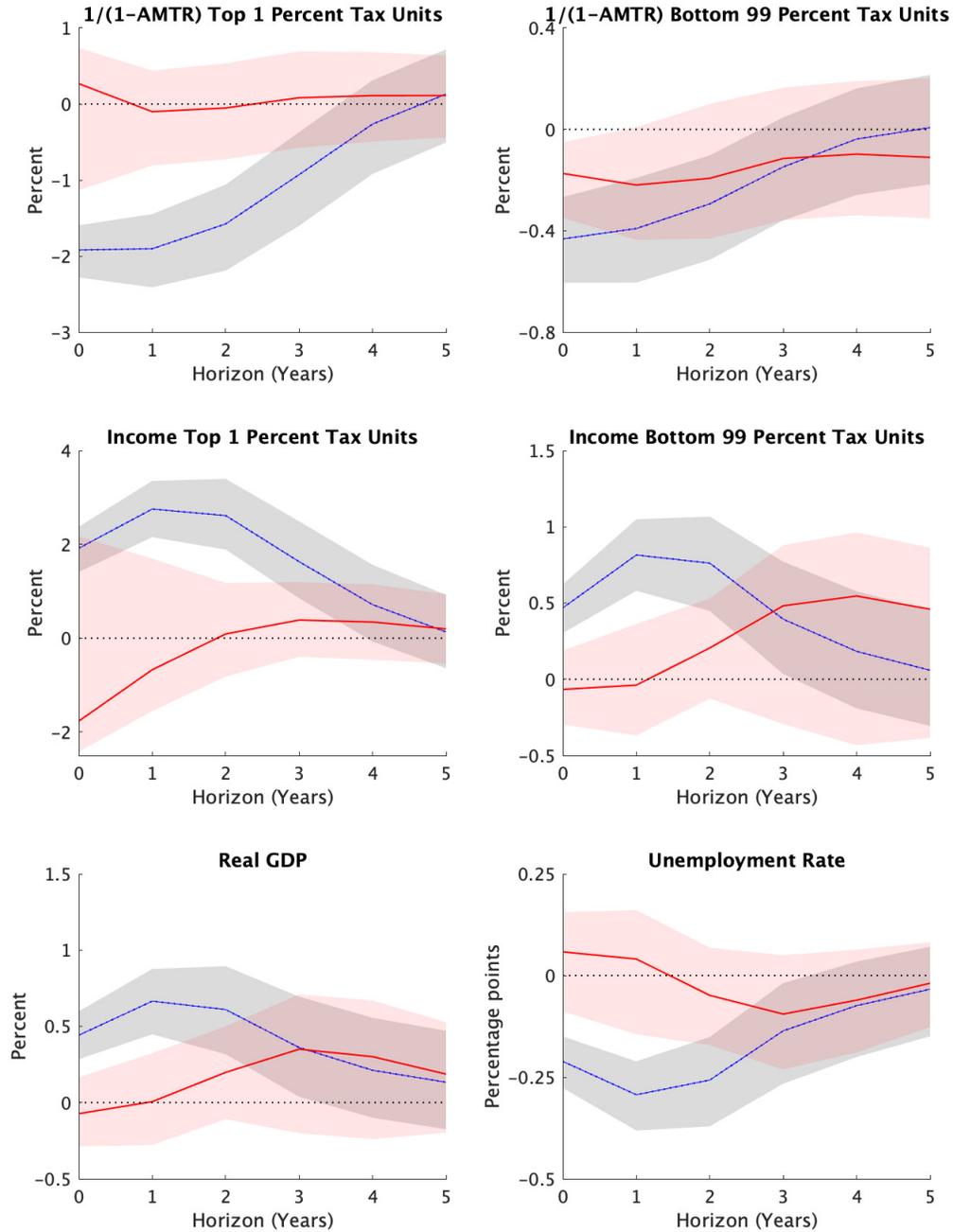


Figure 4: IRFs to counterfactuals. The solid curves (blue for the tax cut for the top 1 percent policy counterfactual and red for the tax cut for the bottom 99 percent policy counterfactual) represent the point-wise posterior medians, and the shaded areas (gray for the tax cut for the top 1 percent policy 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 counterfactuals. The figure is based on 10,000 independent draws obtained using Algorithm 4.

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

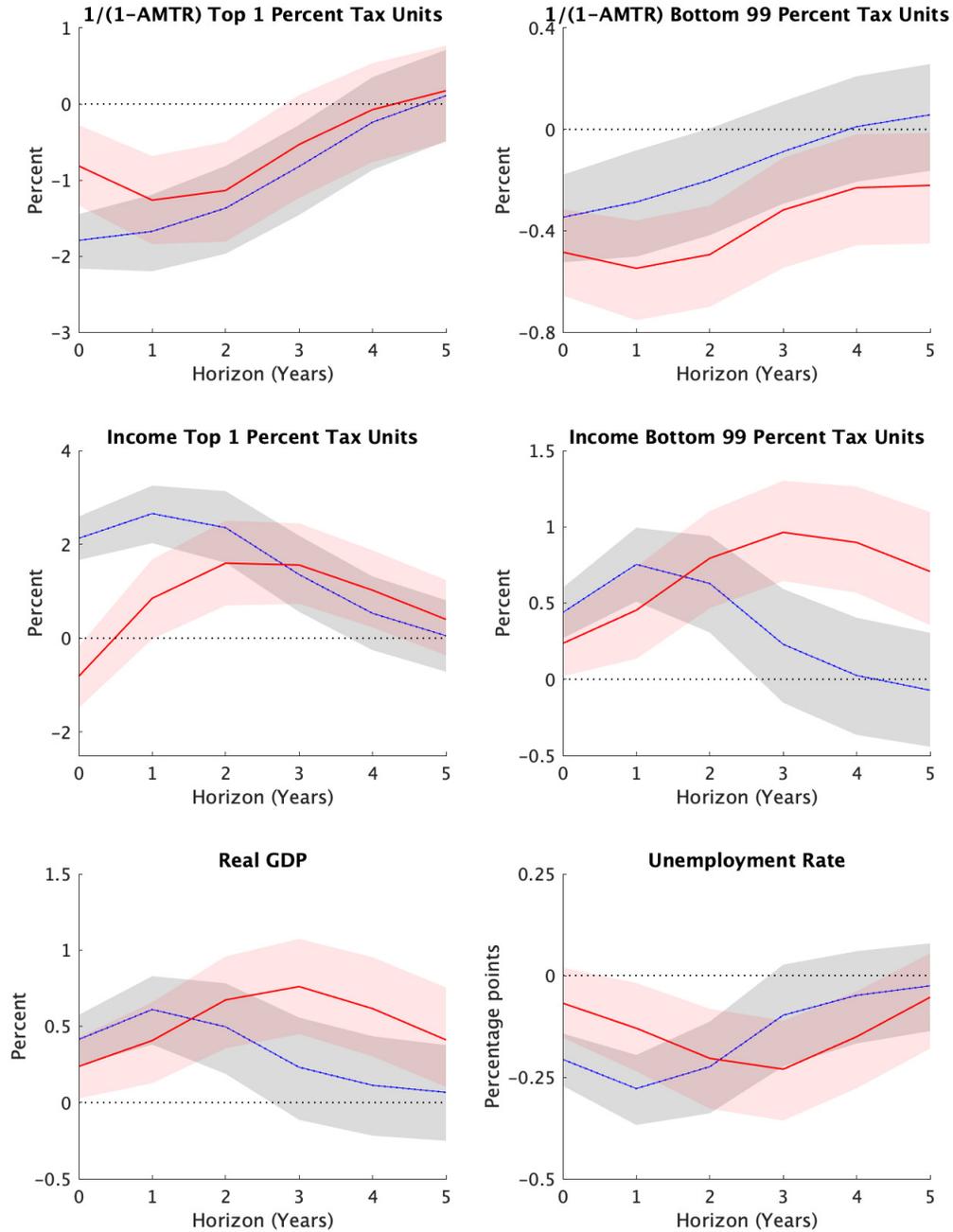


Figure 5: IRFs to structural shocks. The solid curves (blue for the 1 percent AMTR shock and red for the 99 percent AMTR shock) represent the point-wise posterior medians, and the shaded areas (gray for the 1 percent AMTR shock and red for the 99 percent AMTR shock) represent the 68 percent equal-tailed point-wise probability bands. The IRFs are with respect to a one standard deviation shocks. The figure is based on 10,000 independent draws obtained using Algorithm 4.

bottom 99 percent.

As in Section 7.2, the implementation of our sign restrictions needs a function \tilde{F} and matrices S_j very similar to the ones described in Section 6. In the interest of space, we do not describe them.

Figure 5 shows the IRFs to a 1 percent and a bottom 99 percent AMTR shock, respectively. 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 distribution will be quite robust to any linear combination of shocks. In Appendix A.11 we show that Koopman, Shephard and Creal’s (2009) tests validate the finite variance requirement for the reliable use of Algorithm 4.

8 Conclusion

This paper develops efficient algorithms to independently draw from the normal-generalized-normal family of conjugate posterior distributions over the structural parameterization of a Bayesian Proxy-SVAR. In addition, our approach expands the type of identification schemes that can be considered under the frequentist paradigm (e.g. Montiel-Olea, Stock and Watson, 2016). More specifically, influential papers using the frequentist paradigm rely on counterfactuals when more than one instrument is used to identify more than one structural shock. In contrast, our Bayesian 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.
- 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.
- Gleser, L. J. (1992). The Importance of Assessing Measurement Reliability in Multivariate Regression. *Journal of the American Statistical Association* 87(419), 696–707.
- Jarociński, M. and P. Karadi (2018, February). Deconstructing Monetary Policy Surprises: The Role of Information Shocks. (2133).
- Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the Assumptions Behind Importance Sampling. *Journal of Econometrics* 149(1), 2–11.
- Leeper, E. M., C. A. Sims, and T. Zha (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity* 1996(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.
- 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.
- Ramey, V. A. (2016). Macroeconomic Shocks and their Propagation. *Handbook of Macroeconomics* 2, 71–162.

- Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica* 39, 577–591.
- Rubio-Ramírez, J., D. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies* 77(2), 665–696.
- 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.
- 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 Proof of Proposition 1

Proof. It is well known that in the class of Gaussian linear models considered in this paper $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent if and only if they have the same reduced-form parameterization. They will have the same reduced-form parameterization if and only if

$$(\bar{\mathbf{A}}_0 \bar{\mathbf{A}}_0')^{-1} = (\hat{\mathbf{A}}_0 \hat{\mathbf{A}}_0')^{-1} \quad (9)$$

$$\bar{\mathbf{A}}_+ \bar{\mathbf{A}}_0^{-1} = \hat{\mathbf{A}}_+ \hat{\mathbf{A}}_0^{-1}. \quad (10)$$

First, assume that $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent so that Equations (9) and (10) hold. We show that there is $\mathbf{Q} \in \mathcal{Q}$ such that $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$ and $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$. Let $\mathbf{Q} = \hat{\mathbf{A}}_0^{-1} \bar{\mathbf{A}}_0$. It follows directly from the definition of \mathbf{Q} that $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$, and Equation (10) implies $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$. We now show that $\mathbf{Q} \in \mathcal{Q}$. It follows from Equation (9) that $\mathbf{Q}' = \mathbf{Q}^{-1}$. Hence, \mathbf{Q} is orthogonal. Because the lower left-hand $k \times n$ block of both $\bar{\mathbf{A}}_0$ and $\hat{\mathbf{A}}_0$ are zero, the same will be true for both $\hat{\mathbf{A}}_0^{-1}$ and \mathbf{Q} . Let

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{W} \\ \mathbf{0}_{k \times n} & \mathbf{Q}_2 \end{bmatrix},$$

where \mathbf{Q}_1 is $n \times n$, \mathbf{Q}_2 is $k \times k$, and \mathbf{W} is $n \times k$. Because $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_{\bar{n}}$, it must be the case that $\mathbf{W}'\mathbf{Q}_1 = \mathbf{0}_{k \times n}$, $\mathbf{Q}_1'\mathbf{Q}_1 = \mathbf{I}_n$ and $\mathbf{W}'\mathbf{W} + \mathbf{Q}_2'\mathbf{Q}_2 = \mathbf{I}_k$. But $\mathbf{W}'\mathbf{Q}_1 = \mathbf{0}_{k \times n}$, if and only if $\mathbf{W} = \mathbf{0}_{n \times k}$. So $\mathbf{Q} \in \mathcal{Q}$.

Now, assume that $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$ and $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$ for some matrix $\mathbf{Q} \in \mathcal{Q}$. It is easy to see by direct substitution that Equations (9) and (10) hold, which implies that $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent. \square

A.2 The Function $K_{i,j}$

In Steps 1, 2, and 3 of Algorithm 1 draws of $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ were obtained.²⁹ For each draw, Steps 4 and 5 of Algorithm 1 recursively defined matrices $\mathbf{M}_{i,j}$ and $\mathbf{K}_{i,j}$, though the $\mathbf{K}_{i,j}$ were not uniquely defined. Using the $\mathbf{K}_{i,j}$, we can define a mapping g of $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ to $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{q}_{i,j})$, where $\mathbf{q}_{i,j} = \mathbf{K}_{i,j} \mathbf{w}_{i,j}$. In this appendix we show that there is an open set $U \subset \mathbb{R}^{\bar{n}^2 + \bar{n}\bar{m} + \sum_{j=1}^n (n+1-j-\bar{z}_j) + \sum_{j=1}^k (k+1-j)}$ such that g can be uniquely defined over U so that it is continuously differentiable. Furthermore, U will contain almost all draws of $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$.

²⁹In this appendix it will always be the case that $i = 1, 2$ and $1 \leq j \leq n$ when $i = 1$ and $1 \leq j \leq k$ when $i = 2$.

To define the $\mathbf{K}_{i,j}$ we need $m_i \times n_{i,j}$ reference matrices $\mathbf{W}_{i,j}$ that are of full column rank, where

$$m_i = \begin{cases} n & \text{if } i = 1 \\ k & \text{if } i = 2 \end{cases} \quad \text{and} \quad n_{i,j} = \begin{cases} n+1-j-k & \text{if } i = 1 \text{ and } j \leq n-k \\ n+1-j & \text{if } i = 1 \text{ and } j > n-k \\ k+1-j & \text{if } i = 2 \end{cases}$$

We recursively define $U_{i,j}$, $\tilde{\mathbf{M}}_{i,j}$, and $\mathbf{K}_{i,j}$. Let $U_{i,0} = \mathbb{R}^{\tilde{n}^2 + \tilde{n}\tilde{m} + \sum_{j=1}^n (n+1-j-\tilde{z}_j) + \sum_{j=1}^k (k+1-j)}$. For $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ in $U_{i,j-1}$, define

$$\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j}) = \begin{cases} \begin{bmatrix} \mathbf{K}_{i,1}\mathbf{w}_{i,1} & \cdots & \mathbf{K}_{i,j-1}\mathbf{w}_{i,j-1} & \mathbf{G}(\tilde{\mathbf{\Lambda}}_0)' & \mathbf{W}_{i,j} \end{bmatrix} & i = 1 \text{ and } j \leq n-k \\ \begin{bmatrix} \mathbf{K}_{i,1}\mathbf{w}_{i,1} & \cdots & \mathbf{K}_{i,j-1}\mathbf{w}_{i,j-1} & \mathbf{W}_{i,j} \end{bmatrix} & \text{otherwise} \end{cases}.$$

Next, let

$$U_{i,j} = \left\{ (\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j}) \in U_{i,j-1} \mid \det(\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})) \neq 0 \right\}.$$

Because the determinant function is continuous, $U_{i,j}$ will be open. Let

$$\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j}) = \mathbf{Q}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})\mathbf{R}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$$

be the QR-decomposition of $\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ normalized so that the diagonal of $\mathbf{R}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ is positive and define $\mathbf{K}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ to be the last $n_{i,j}$ columns of $\mathbf{Q}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$. The QR-decomposition of a square non-singular matrix, normalized so that the diagonal of the triangular component is positive, is a continuously differentiable function. Thus, $\mathbf{K}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$ is defined and continuously differentiable over $U_{i,j}$. Let U be the intersection of all the $U_{i,j}$. Note that U is open and g can be defined over U so that it is continuously differentiable.

All that remains to be shown is that U contains almost all draws of $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$. It suffices to show that $U_{i,j}^c$, the complement of $U_{i,j}$ in $U_{i,j-1}$, is of measure zero. We will show that $U_{i,j}^c$ is a union of manifolds of strictly lower dimension and thus is of measure zero. For each ℓ with $0 < \ell < m_i$ and each set of ℓ columns of $\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$, consider the set of all $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j}) \in U_{i,j-1}$ such that the given set of ℓ columns are linearly independent and the remaining columns are in the span of the ℓ linearly independent columns. Any element of $U_{i,j}^c$ will be in at least one such set. Furthermore, sets of these forms will be manifolds of strictly lower dimension.³⁰ This completes the proof.

³⁰For any set of ℓ columns of $\tilde{\mathbf{M}}_{i,j}(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \mathbf{w}_{i,j})$, we can project each of the remaining columns onto perpendicular component of the span of the ℓ selected columns. As long as the selected columns are linearly independent, this mapping will be continuously

A.3 Proof of Proposition 2

Proof. By Proposition 1, the Proxy-SVAR structural parameters $(\bar{\mathbf{A}}_0, \bar{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ are observationally equivalent if and only if there exists $\mathbf{Q} \in \mathcal{Q}$ such that $\bar{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 \mathbf{Q}$ and $\bar{\mathbf{A}}_+ = \hat{\mathbf{A}}_+ \mathbf{Q}$. So to prove Proposition 2, it suffices to prove that if $\mathbf{Q} \in \mathcal{Q}$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ satisfy the exogeneity restrictions and the relevance condition, then $(\hat{\mathbf{A}}_0 \mathbf{Q}, \hat{\mathbf{A}}_+ \mathbf{Q})$ will satisfy the exogeneity restrictions and relevance condition if and only if $\mathbf{Q} \in \mathcal{X} \subset \mathcal{Q}$. Recall from Sections 3.2 and 3.3, the Proxy-SVAR parameters $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ will satisfy the relevance condition if and only if $\mathbf{J}(\tilde{\mathbf{A}}_0^{-1})' \mathbf{L}'$ is of full row rank and will satisfy the exogeneity restrictions if and only if $\mathbf{J}(\tilde{\mathbf{A}}_0^{-1})' \mathbf{L}' \mathbf{K}' = \mathbf{0}_{k \times n-k}$, where $\mathbf{K} = [\mathbf{I}_{n-k} \ \mathbf{0}_{n-k \times k}]$.

We will maintain the assumption that $\mathbf{Q} \in \mathcal{Q}$ and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ satisfy the exogeneity restrictions and the relevance condition. First assume that $\mathbf{Q} \in \mathcal{X} \subset \mathcal{Q}$, so that $\mathbf{Q} = \text{diag}(\mathbf{Q}_3, \mathbf{Q}_4, \mathbf{Q}_5)$, where $\mathbf{Q}_3 \in (n-k)$, $\mathbf{Q}_4 \in (k)$, and $\mathbf{Q}_5 \in (k)$. Since $\mathbf{J}((\hat{\mathbf{A}}_0 \mathbf{Q})^{-1})' \mathbf{L}' = \mathbf{J}(\hat{\mathbf{A}}_0^{-1})' \mathbf{L}' \text{diag}(\mathbf{Q}_3, \mathbf{Q}_4)$, $\mathbf{J}((\hat{\mathbf{A}}_0 \mathbf{Q})^{-1})' \mathbf{L}' \mathbf{K}' = \mathbf{J}(\hat{\mathbf{A}}_0^{-1})' \mathbf{L}' \mathbf{K}' \mathbf{Q}_3$, and $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ satisfy the relevance condition and the exogeneity restriction, so will $(\hat{\mathbf{A}}_0 \mathbf{Q}, \hat{\mathbf{A}}_+ \mathbf{Q})$.

Now assuming that the parameters $(\hat{\mathbf{A}}_0 \mathbf{Q}, \hat{\mathbf{A}}_+ \mathbf{Q})$ satisfy the relevance condition and the exogeneity restriction, we show that $\mathbf{Q} \in \mathcal{X}$. Because $\mathbf{Q} \in \mathcal{Q}$, it is of the form $\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2)$, where $\mathbf{Q}_1 \in \mathcal{O}(n)$ and $\mathbf{Q}_2 \in \mathcal{O}(k)$. Let

$$\mathbf{Q}_1 = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix},$$

where \mathbf{Q}_{11} is $(n-k) \times (n-k)$, \mathbf{Q}_{12} is $(n-k) \times k$, \mathbf{Q}_{21} is $k \times (n-k)$, and \mathbf{Q}_{22} is $k \times k$. If \mathbf{Q}_{21} is zero, then it will follow that $\mathbf{Q} \in \mathcal{X}$. Because $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ satisfies both the relevance condition and the exogeneity restrictions, the columns of \mathbf{K}' form a basis for the null space of $\mathbf{J}(\hat{\mathbf{A}}_0^{-1})' \mathbf{L}'$. Because both $(\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_+)$ and $(\hat{\mathbf{A}}_0 \mathbf{Q}, \hat{\mathbf{A}}_+ \mathbf{Q})$ satisfy the exogeneity restrictions, then

$$\mathbf{0}_{k \times n-k} = \mathbf{J}((\hat{\mathbf{A}}_0 \mathbf{Q})^{-1})' \mathbf{L}' \mathbf{K}' = \mathbf{J}(\hat{\mathbf{A}}_0^{-1})' \mathbf{L}' \left(\mathbf{K}' \mathbf{Q}_{11} + \begin{bmatrix} \mathbf{0}_{n-k, n-k} \\ \mathbf{Q}_{21} \end{bmatrix} \right) = \mathbf{J}(\hat{\mathbf{A}}_0^{-1})' \mathbf{L}' \begin{bmatrix} \mathbf{0}_{n-k, n-k} \\ \mathbf{Q}_{21} \end{bmatrix}.$$

Thus \mathbf{Q}_{21} must be zero as desired. □

A.4 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

differentiable and the remaining columns will be in the span of the selected columns if and only if this mapping is zero. This mapping along with standard results from multivariate calculus is enough to imply that the sets defined above are manifolds of strictly lower dimension.

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 j^{th} 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

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

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

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

$$\begin{aligned} \nu &= T + \tilde{n} \\ \mathbf{H}_j &= (\mathbf{V}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \mathbf{V}_j)^{-1} \\ \mathbf{P}_j &= \mathbf{H}_j \mathbf{V}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \hat{\boldsymbol{\Psi}}_j \mathbf{U}_j \\ \mathbf{S}_j &= (\mathbf{U}_j' \hat{\boldsymbol{\Phi}}_j \mathbf{U}_j + \mathbf{U}_j' \hat{\boldsymbol{\Psi}}_j' \hat{\boldsymbol{\Omega}}_j^{-1} \hat{\boldsymbol{\Psi}}_j \mathbf{U}_j - \mathbf{P}_j' \mathbf{H}_j^{-1} \mathbf{P}_j)^{-1}. \end{aligned}$$

A.5 Proposal Normal-Generalized-Normal Parameters

As mentioned in Section 3, while oftentimes it suffices to choose $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ to be equal to $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, 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 squared 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.6 Data Appendix for Section 6

Here we describe the data used in Section 6 in more details. 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.7 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.8 Tests for the Analysis in Section 6

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 6. 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 6.

A.9 Tests for the Analysis in Section 7.1

Table A.2 shows that Koopman, Shephard and Creal’s (2009) tests indicate that the importance weights used in Section 7.1 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 7.1.

A.10 Tests for the Analysis in Section 7.2

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 7.2. Table A.5 shows that Koopman, Shephard and Creal’s (2009) tests for the Proxy-SVAR identified

with the proxy and sign restrictions described in Section 7.2.

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 7.2 for the AMTR counterfactual.

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 7.2 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 7.2 for the Proxy-SVAR identified with proxy and sign restrictions.

The tables indicate that the importance weights have finite variance.

A.11 Tests for the Analysis in Section 7.3

Tables A.6 and A.7 present Koopman, Shephard and Creal's (2009) tests for the case in which the bottom and top tax rates cut shocks are identified using the counterfactuals in Mertens and Montiel-Olea (2018) described in Section 7.3. Table A.8 presents Koopman, Shephard and Creal's (2009) tests for the case in which the bottom and top tax rates cut shocks are identified using the less restrictive scheme described in Section 7.3.

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 7.3 for the Top 1 percent counterfactual.

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 7.3 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 7.3 for the Proxy-SVAR identified using the less restrictive identification scheme.

The tables indicate that the importance weights have finite variance.