FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Inference in Bayesian Proxy-SVARs

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

Working Paper 2018-16
December 2018

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. Importantly, our techniques can handle the case of set identification and hence they can be used to relax the additional exclusion restrictions unrelated to the external instruments often imposed to facilitate inference when more than one instrument are used to identify more than one equation, as in Mertens and Montiel-Olea (2018).

JEL classification: C15, C32

Key words: SVARs, external instruments, importance sampler

https://doi.org/10.29338/wp2018-16

The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. Any remaining errors are the authors' responsibility. No statements here should be treated as legal advice.

Please address questions regarding content to Jonas E. Arias, Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574, jonas.arias@phil.frb.org; Juan F. Rubio-Ramírez, Department of Economics, Emory University, Rich Memorial Building, Room 306, Atlanta, GA 30322-2240 and Federal Reserve Bank of Atlanta and BBVA Research, juan.rubio-ramirez@emory.edu; or Daniel F. Waggoner, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8278, daniel.f.waggoner@atl.frb.org.

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's website at www.frbatlanta.org. Click "Publications" and then "Working Papers." To receive e-mail notifications about new papers, use frbatlanta.org/forms/subscribe.

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 such front, less is known about conducting Bayesian inference in this class of structural time series models; exceptions are Bahaj (2014), Drautzburg (2016), and Caldara and Herbst (2016).

In this paper we contribute to this line of research by developing efficient algorithms to independently draw from the family of restricted normal-generalized-normal posterior distributions over the structural parameterization of a proxy-SVAR conditional on exogeneity restrictions. These restrictions require that the correlation between the proxies and some of the structural shocks be equal to zero. The fact that we can draw independently opens the door to use the Bayesian paradigm in larger models.

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 the normal-generalized-normal 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 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—that makes it possible to sample independently from the structural parameterization of the proxy-SVAR even in the presence of the zero restrictions ingrained in this framework.

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

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

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 adapting the volume elements used in Arias, Rubio-Ramírez and Waggoner (2018), so that they come from the normal-generalized-normal 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 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

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

one structural equation—can provide critical insights for a few but highly influential studies identifying two structural equations 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 in order to separately identify the two structural equations of interest they are limited to consider a narrow class of additional zero restrictions that are hard to justify. We will make this clear by revisiting Mertens and Montiel-Olea (2018). This paper relies on proxy-SVARs to study the effects of exogenous changes in marginal and average personal income tax rates. One of their main conclusions is that, even though the response of reported income to exogenous changes in marginal tax rates is strong and significant, the response to exogenous changes in average tax rates is not statistically significant at any horizon. We will show that the identification scheme underlying this result exactly identifies the equations associated with both tax shocks by imposing an additional zero restriction on the systematic component of tax policies. We argue that their identification scheme is hard to justify; as a result, we substitute it for a set of less questionable sign restrictions. Once this is done, we find that both substitution and wealth 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). 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. But this is not the only difference with our approach. More importantly, they ignore the volume element when transforming the draws into the structural proxy-SVAR parameterization. As explained in Arias, Rubio-Ramírez, and Waggoner (2018), this implies that they are not drawing from a normal-generalized-normal posterior distribution over the structural parameterization of the proxy-SVAR conditional on the exogeneity restrictions and that the posterior distribution from which they are drawing depends on the zero restrictions—making comparisons across identification schemes impossible.

Finally, let's relate our paper to Caldara and Herbst (2016). As in our approach, this paper draws from the structural parameterization of the proxy-SVAR. Nevertheless, while our approach imposes a conjugate prior on the structural parameterization, they work with non-conjugate prior densities, which complicates inference. In particular, the posterior draws are not independent and their Metropolis-Hastings sampler could become computationally inefficient compared with ours in large models. One advantage of Caldara and Herbst's (2016)

approach relative to ours is that they can impose any desired prior on the strength of the proxy. Even so, researchers using our approach can also explore the 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 5 and 6 present two applications. Section 7 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 that we would like to independently draw from. 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 $\boldsymbol{y}_t$ be a $n \times 1$ vector of endogenous variables, $\boldsymbol{m}_t$ be a $k \times 1$ vector of instruments (also called proxies), $\tilde{\boldsymbol{y}}_t' = [\boldsymbol{y}_t' \;\; \boldsymbol{m}_t']$, and $\tilde{n} = n + k$. If these are governed by a SVAR, then

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

where $\tilde{\boldsymbol{A}}_i$ is an $\tilde{n} \times \tilde{n}$ matrix for $0 \leq i \leq p$ with $\tilde{\boldsymbol{A}}_0$ invertible, $\tilde{\boldsymbol{c}}$ is a $1 \times \tilde{n}$ row vector, and $\tilde{\boldsymbol{\varepsilon}}_t$ is conditionally standard normal with mean zero and identity variance-covariance matrix.[3] If $\tilde{\boldsymbol{x}}_t' = [\tilde{\boldsymbol{y}}_{t-1}' \;\; \cdots \;\; \tilde{\boldsymbol{y}}_{t-p}' \;\; 1]$ and

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

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

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

$\tilde{\boldsymbol{A}}'_+ = [\tilde{\boldsymbol{A}}'_1 \cdots \tilde{\boldsymbol{A}}'_p \;\; \tilde{\boldsymbol{c}}']$, Equation (1) can be more compactly written as

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

Let $\tilde{\boldsymbol{\varepsilon}}'_t = [\boldsymbol{\varepsilon}'_t \;\; \boldsymbol{\nu}'_t]$, 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 $\boldsymbol{y}_t$ evolves according to

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

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

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

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

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 $\boldsymbol{0}_{k \times n}$ is a $k \times n$ matrix of zeros. We call the zero restrictions on $\tilde{\boldsymbol{A}}_0$ and $\tilde{\boldsymbol{A}}_+$ the block restrictions. We call Equation (2) plus the block restrictions the structural parameterization of the proxy-SVAR and $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ such that the block restrictions hold the proxy-SVAR structural parameters. Hence, for the set of proxy-SVAR structural parameters the block restrictions always hold. Mertens and Ravn (2013) and Stock and Watson (2018) also model $\boldsymbol{y}_t$ and $\boldsymbol{m}_t$ as a joint system, but the structure is not exactly the same.

2.2 The Identification Problem in a Proxy-SVAR

Following Rothenberg (1971) the proxy-SVAR structural parameters $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent if and only if they imply the same joint distribution of $\tilde{\boldsymbol{y}}_1, \cdots, \tilde{\boldsymbol{y}}_T$. Proposition 1 extends an insight of Rubio-Ramírez, Waggoner, and Zha (2010) to the case of proxy-SVARs. Let $\text{diag}(X_1, \cdots, X_j)$ denote a block diagonal matrix with the matrices $X_1, \cdots, X_j$ along the diagonal.

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

orthogonal matrices and $\mathcal{Q}$ is defined by

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

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 $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ are proxy-SVAR parameters, then so are $(\tilde{\boldsymbol{A}}_0\boldsymbol{Q}, \tilde{\boldsymbol{A}}_+\boldsymbol{Q})$ for all $\boldsymbol{Q} \in \mathcal{Q}$. By Proposition 1, $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ and $(\tilde{\boldsymbol{A}}_0\boldsymbol{Q}, \tilde{\boldsymbol{A}}_+\boldsymbol{Q})$ 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.[4] 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.[5]

More specifically, the identification problem in proxy-SVARs is typically solved by assuming that the k proxies are correlated with k structural shocks in $\boldsymbol{\varepsilon}_t$ and uncorrelated with the remaining structural shocks. Without loss of generality let the structural shocks correlated with the proxies be the last k elements of $\boldsymbol{\varepsilon}_t$ and the structural shocks uncorrelated with the proxies be the first $n-k$ elements of $\boldsymbol{\varepsilon}_t$. We now show that the latter restrictions—which are known in the literature as exogeneity restrictions—are zero restrictions on a non-linear function of the proxy-SVAR structural parameters. To see this, first note that for the proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ it is the case that

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

[4] A proxy-SVAR equation is identified if for any two sets of observationally equivalent proxy-SVAR parameters, the parameters in that equation are identical.

[5] As 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.

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

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

It follows that

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

Thus, the identifying restrictions imply that the first $n-k$ rows of matrix $\boldsymbol{A}_0^{-1}\boldsymbol{\Gamma}_{0,1}\boldsymbol{\Gamma}_{0,2}^{-1}$ must be zero. Taking transposes, this implies that the lower left-hand $k \times (n-k)$ block of $(\tilde{\boldsymbol{A}}_0^{-1})'$ 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 $(\tilde{\boldsymbol{A}}_0^{-1})'$. 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 $-\boldsymbol{A}_0^{-1}\boldsymbol{\Gamma}_{0,1}\boldsymbol{\Gamma}_{0,2}^{-1}$, to be non-singular. Following the literature, we refer to this as the relevance condition.

2.3 Priors, Posteriors, and a Useful Parameterization and Mapping

We will use a restricted normal-generalized-normal distribution over the structural parameterization of the proxy-SVAR as our prior distribution.[6] Hence, the proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ have a prior density proportional to $NGN_{(\nu,\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Omega})}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, where

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

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 $\boldsymbol{\Phi}$, an $\tilde{m}\tilde{n} \times \tilde{n}^2$ matrix $\boldsymbol{\Psi}$, and an $\tilde{m}\tilde{n} \times \tilde{m}\tilde{n}$ symmetric and positive definite matrix $\boldsymbol{\Omega}$.[7] The normal-generalized-normal distribution over the structural parameterization is a conjugate family of distributions commonly used in the literature. For instance, the Sims-Zha prior (see Sims and Zha, 1998) belongs to this family.

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

[6] By 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, $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$.

[7] Arias, Rubio-Ramírez, and Waggoner (2018) assumed a Kronecker structure for $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Omega}$.

by Equation (4). More specifically, such posterior distribution is proportional to $NGN(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\nu} = T+\nu$, $\tilde{\boldsymbol{\Omega}} = (\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{\Omega}^{-1})^{-1}$, $\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Omega}}(\boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} + \boldsymbol{\Omega}^{-1}\boldsymbol{\Psi})$, and $\tilde{\boldsymbol{\Phi}} = \boldsymbol{I}_{\tilde{n}} \otimes \tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}} + \boldsymbol{\Phi} + \boldsymbol{\Psi}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi} - \tilde{\boldsymbol{\Psi}}'\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\boldsymbol{\Psi}}$, $\tilde{\boldsymbol{Y}} = [\tilde{\boldsymbol{y}}_1 \ \cdots \ \tilde{\boldsymbol{y}}_T]'$, and $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1 \ \cdots \ \tilde{\boldsymbol{x}}_T]'$.

Arias, Rubio-Ramírez, and Waggoner (2018) showed how to independently draw from a normal-generalized-normal posterior distribution over the structural parameterization of a SVAR conditional on sign and zero restrictions. Since a proxy-SVAR identified with exogeneity restrictions can be represented by the SVAR in Equation (2) identified with the zero restrictions on $\tilde{\boldsymbol{A}}_0$, $\tilde{\boldsymbol{A}}_+$, and $(\tilde{\boldsymbol{A}}_0^{-1})'$ 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 $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$, while the maximum number of restrictions that the aforementioned algorithm can handle on the jth column of the structural parameters is $\tilde{n} - j$. So unless $p = 0$, an uninteresting case, the maximum will be exceeded for the nth column, if not before. In this paper we show 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 normal-generalized-normal 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 restricted normal-generalized-normal 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 $\tilde{\boldsymbol{\Lambda}}_0$ be an $\tilde{n} \times \tilde{n}$ matrix, $\tilde{\boldsymbol{\Lambda}}_+$ be an $\tilde{m} \times \tilde{n}$ matrix, $\boldsymbol{Q}_1$ be an $n \times n$ orthogonal matrix, and $\boldsymbol{Q}_2$ be a $k \times k$ orthogonal matrix. The matrix $\tilde{\boldsymbol{\Lambda}}_0$ is restricted to be upper-triangular with positive diagonal. The matrix $\tilde{\boldsymbol{\Lambda}}'_+ = [\tilde{\boldsymbol{\Lambda}}'_1 \cdots \tilde{\boldsymbol{\Lambda}}'_p \;\; \boldsymbol{d}']$, where $\tilde{\boldsymbol{\Lambda}}_i$ is $\tilde{n} \times \tilde{n}$ for $1 \leq i \leq p$ and $\boldsymbol{d}$ is $1 \times \tilde{n}$, is restricted so that the lower left-hand $k \times n$ block of $\tilde{\boldsymbol{\Lambda}}_i$ is zero for $1 \leq i \leq p$. We label the zero restrictions on $\tilde{\boldsymbol{\Lambda}}_0$ and $\tilde{\boldsymbol{\Lambda}}_+$ the triangular-block restrictions, and we call $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ such that the triangular-block restrictions hold the triangular-block parameters.

Given any values of the triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ and the orthogonal matrices $(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$, we can map $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ into proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ by[8]

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

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

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

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

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

Just as a standard SVAR can alternatively be written in the orthogonal reduced-form parameterization, the triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ together with the orthogonal matrices $(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ 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

$$\tilde{\boldsymbol{y}}'_t \tilde{\boldsymbol{\Lambda}}_0 = \tilde{\boldsymbol{x}}'_t \tilde{\boldsymbol{\Lambda}}_+ + \tilde{\boldsymbol{u}}'_t \text{ for } 1 \leq t \leq T, \tag{5}$$

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

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

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, $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, 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 $\boldsymbol{Q}_1$. This will allow us to use the ideas in Arias, Rubio-Ramírez and Waggoner (2018) to draw the orthogonal matrices $(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$, conditional on each draw of the triangular-block parameters, such that the exogeneity restrictions hold. Then, we use f to map triangular-block parameters plus the orthogonal matrices, $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$, into proxy-SVAR structural parameters, $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$. 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 $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$.[9] 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 $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ are block diagonal.[10] Since the triangular and block restrictions on $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ are exclusion restrictions that do not involve cross-equation restrictions, and $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Omega}}$ can be chosen to be block diagonal, the conditions for using the Gibbs sampler are satisfied. Furthermore, because $\tilde{\boldsymbol{\Lambda}}_0$ is restricted to be upper-triangular, it follows from Theorem 2 of Waggoner and Zha (2003) that the Gibbs sampler draws will be independent. In Appendix A.4, we describe how to adapt their paper to our purposes.

Often, it suffices to choose $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ to be equal to $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$, the parameters associated with the

[9] Here, 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.

[10] The Gibbs sampler of Waggoner and Zha (2003) was developed to draw from the posterior distribution of a structural VAR with linear non-cross-equation restrictions using a certain class of normal priors. The class of posterior distributions that can be obtained with this class of priors is the set of all normal-generalized-normal distributions $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ with $\hat{\nu} \geq \tilde{n}$, $\hat{\boldsymbol{\Phi}}$ block diagonal with $\tilde{n}$ symmetric and positive definite $\tilde{n} \times \tilde{n}$ blocks, $\hat{\boldsymbol{\Psi}}$ block diagonal with $\tilde{n}$ arbitrary $\tilde{m} \times \tilde{n}$ blocks, and $\hat{\boldsymbol{\Omega}}$ 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.

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 $(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ that can avoid this loss of efficiency.

3.2 Exogeneity Restrictions in the Orthogonal-Triangular-Block Parameterization

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

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

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

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

where we have used the fact that $\tilde{\boldsymbol{A}}_0 = \tilde{\boldsymbol{\Lambda}}_0 \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Thus, conditional on a draw of triangular-block parameters $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$, the exogeneity restrictions are linear restrictions on the columns of $\boldsymbol{Q}_1$. We will denote the number of exogeneity restrictions on the jth column of $\boldsymbol{Q}_1$ 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 $\boldsymbol{Q}_1$. As will become clear below, drawing the orthogonal matrix $\boldsymbol{Q}_2$ 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 $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ independently from the restricted $NGN(\hat{\nu}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}})$ distribution using Waggoner and Zha's (2003) Gibbs sampler.*

2. *For* $1 \leq j \leq n$, *draw* $\boldsymbol{x}_{1,j} \in \mathbb{R}^{n+1-j-\tilde{z}_j}$ *independently from a standard normal distribution and set* $\boldsymbol{w}_{1,j} = \boldsymbol{x}_{1,j} / \parallel \boldsymbol{x}_{1,j} \parallel$.
3. *For* $1 \leq j \leq k$, *draw* $\boldsymbol{x}_{2,j} \in \mathbb{R}^{k+1-j}$ *independently from a standard normal distribution and set* $\boldsymbol{w}_{2,j} = \boldsymbol{x}_{2,j} / \parallel \boldsymbol{x}_{2,j} \parallel$.
4. *Define* $\boldsymbol{Q}_1 = [\boldsymbol{q}_{1,1} \ \cdots \ \boldsymbol{q}_{1,n}]$ *recursively by* $\boldsymbol{q}_{1,j} = \boldsymbol{K}_{1,j}\boldsymbol{w}_{1,j}$ *for any matrix* $\boldsymbol{K}_{1,j}$ *whose columns form an orthonormal basis for the null space of the* $(j-1+\tilde{z}_j) \times n$ *matrix:*

$$\boldsymbol{M}_{1,j} = \begin{cases} \begin{bmatrix} \boldsymbol{q}_{1,1} & \cdots & \boldsymbol{q}_{1,j-1} & \boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)' \end{bmatrix}' & \text{for } 1 \leq j \leq n-k \\ \begin{bmatrix} \boldsymbol{q}_{1,1} & \cdots & \boldsymbol{q}_{1,j-1} \end{bmatrix}' & \text{for } n-k+1 \leq j \leq n. \end{cases}$$

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

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

6. *Set* $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+) = f(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$.
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 $\boldsymbol{M}_{i,j}$ is of full row rank; otherwise the number of columns in $\boldsymbol{K}_{i,j}$ will not equal the dimension of $\boldsymbol{w}_{i,j}$ and so the product $\boldsymbol{K}_{i,j}\boldsymbol{w}_{i,j}$ will not be defined. When $i = 1$ and $j > n-k$ or when $i = 2$ and $1 \leq j \leq k$, the matrix $\boldsymbol{M}_{i,j}$ will clearly be of full row rank. However, when $i = 1$ and $j \leq n-k$, the matrix $\boldsymbol{M}_{1,j}$ will be of full row rank if and only if $\boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)$ is of full row rank. This is because, by construction, the $\boldsymbol{q}_{1,1}, \cdots, \boldsymbol{q}_{1,j-1}$ are perpendicular to $\boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)$. Since the probability of drawing a $\boldsymbol{\Lambda}_0$ such that the $\boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)$ is not of full row rank is zero, we can assume without loss of generality that $\boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)$ is of full row rank.

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

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

where the $k \times k$ matrix $\boldsymbol{V}$ is the covariance matrix of the k proxy variables and the last k structural shocks. So, the relevance condition, which requires $\boldsymbol{V}$ to be non-singular, holds if and only if $\boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)$ 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.

It is also the case that the matrix $\boldsymbol{K}_{i,j}$ is not unique. If the columns of $\boldsymbol{K}_{i,j}$ form an orthonormal basis for the null space of $\boldsymbol{M}_{i,j}$, then so will the columns of $\boldsymbol{K}_{i,j}\boldsymbol{X}$ for any orthogonal matrix $\boldsymbol{X}$. The particular choice of $\boldsymbol{K}_{i,j}$ 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 $\boldsymbol{K}_{i,j}$ 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 $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+)$ from a restricted normal-generalized-normal. Step 2 draws $\boldsymbol{w}_{1,j}$ from the uniform distribution on the unit sphere in $\mathbb{R}^{n+1-j-\tilde{z}_j}$. Step 3 draws $\boldsymbol{w}_{2,j}$ from the uniform distribution on the unit sphere in $\mathbb{R}^{k+1-j}$. Hence, the density over $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{1,1}, \cdots, \boldsymbol{w}_{1,n}, \boldsymbol{w}_{2,1}, \cdots, \boldsymbol{w}_{2,k})$ is proportional to a restricted normal-generalized-normal. Steps 4 and 5 map $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{1,1}, \cdots, \boldsymbol{w}_{1,n}, \boldsymbol{w}_{2,1}, \cdots, \boldsymbol{w}_{2,k})$ to $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$, we denote this mapping by g. Finally, Step 6 maps $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$ into proxy-SVAR structural parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ 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 $\boldsymbol{K}_{i,j}$ 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

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

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

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 $NGN(\tilde{\nu},\tilde{\mathbf{\Phi}},\tilde{\mathbf{\Psi}},\tilde{\mathbf{\Omega}})$ 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 $(\tilde{\boldsymbol{A}}_0,\tilde{\boldsymbol{A}}_+)$ that satisfy the exogeneity restrictions and the relevance condition.*
2. *Set its importance weight to*

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

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

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

where w_i is the weight associated with the ith 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 $v_{(f\circ g)^{-1}|_{\mathcal{Z}}}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ in Step 2 is the most expensive part in implementing Algorithm 2. The rest of Algorithm 2 is quite fast.[11]

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

$$\mathcal{X} = \{\boldsymbol{Q} \in \mathcal{Q} | \boldsymbol{Q} = \operatorname{diag}(\boldsymbol{Q}_3, \boldsymbol{Q}_4, \boldsymbol{Q}_5), \boldsymbol{Q}_3 \in \mathcal{O}(n-k), \boldsymbol{Q}_4 \in \mathcal{O}(k), \textit{and } \boldsymbol{Q}_5 \in \mathcal{O}(k)\}.$$

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

[11] The 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.

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 $\boldsymbol{Q}_3$ will rotate the columns of the proxy-SVAR structural parameters associated with the structural shocks that are not correlated with the proxies while $\boldsymbol{Q}_4$ will rotate the proxy-SVAR equations associated with the structural shocks that are correlated with the proxies.[12] 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 $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ be proxy-SVAR structural parameters that also satisfy the exogeneity restrictions and the relevance condition. Then, the last column of $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ is identified up to a sign.*

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

Notwithstanding, the theory and techniques that we develop apply to additional sign and zero restrictions on any function $\boldsymbol{F}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ from the set of proxy-SVAR structural parameters to the set of $\tilde{r} \times \tilde{n}$ matrices that satisfies the condition $\boldsymbol{F}(\tilde{\boldsymbol{A}}_0\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{A}}_+\tilde{\boldsymbol{Q}}) = \boldsymbol{F}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)\tilde{\boldsymbol{Q}}$, for every $\tilde{\boldsymbol{Q}} \in \mathcal{O}(\tilde{n})$.[13]

4.1 Additional Zero Restrictions

To set the notation, let $\boldsymbol{Z}_j$ 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 $\boldsymbol{Z}_j\boldsymbol{F}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)\boldsymbol{e}_{\tilde{n},j} = \boldsymbol{0}_{z_j,1}$ 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

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

[12] Note that $\boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ are related to $\boldsymbol{Q}_1$ introduced in Proposition 1 while $\boldsymbol{Q}_5$ is related to $\boldsymbol{Q}_2$ of the same proposition.

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

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

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

This last equation can be written as

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

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

$$\boldsymbol{M}_{1,j}=\begin{cases}\begin{bmatrix}\boldsymbol{q}_{1,1} & \cdots & \boldsymbol{q}_{1,j-1} & \boldsymbol{G}(\tilde{\boldsymbol{\Lambda}}_0)' & \boldsymbol{G}_{1,j}(\tilde{\boldsymbol{\Lambda}}_0,\tilde{\boldsymbol{\Lambda}}_+)'\end{bmatrix}' & \text{for } 1\leq j\leq n-k\\ \begin{bmatrix}\boldsymbol{q}_{1,1} & \cdots & \boldsymbol{q}_{1,j-1} & \boldsymbol{G}_{1,j}(\tilde{\boldsymbol{\Lambda}}_0,\tilde{\boldsymbol{\Lambda}}_+)'\end{bmatrix}' & \text{for } n-k+1\leq j\leq n.\end{cases}$$

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

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

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

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

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

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 $\boldsymbol{V}$ 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 $\tilde{\boldsymbol{F}}(\tilde{\boldsymbol{A}}_0,\tilde{\boldsymbol{A}}_+)>\boldsymbol{0}_{s,1}$, where $\tilde{\boldsymbol{F}}$ is continuous and $0\leq s$ is the number of sign restrictions. As explained in Arias, Rubio-Ramírez, and Waggoner (2018), because $\tilde{\boldsymbol{F}}$ 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.[14] 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

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

Algorithm 2 to consider additional sign and zero restrictions when using Algorithm 3 instead of Algorithm 1 in Step 1.

Algorithm 4. *The following algorithm independently draws from the desired restricted* $NGN(\tilde{\nu}, \tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}, \tilde{\mathbf{\Omega}})$ *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* $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ *that satisfy the exclusion restrictions, the relevance condition, and the additional zero restrictions.*
2. *If* $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ *satisfies the sign restrictions, set its importance weight to*

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

   where $(\tilde{\mathbf{\Lambda}}_0, \tilde{\mathbf{\Lambda}}_+, \boldsymbol{Q}_1, \boldsymbol{Q}_2) = f^{-1}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ *and where* $\mathcal{X}$ *denotes the set of all proxy-SVAR structural parameters that satisfy the exclusion 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 $v_{(f\circ g)^{-1}|_{\mathcal{X}}}(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ 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 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.[15]

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

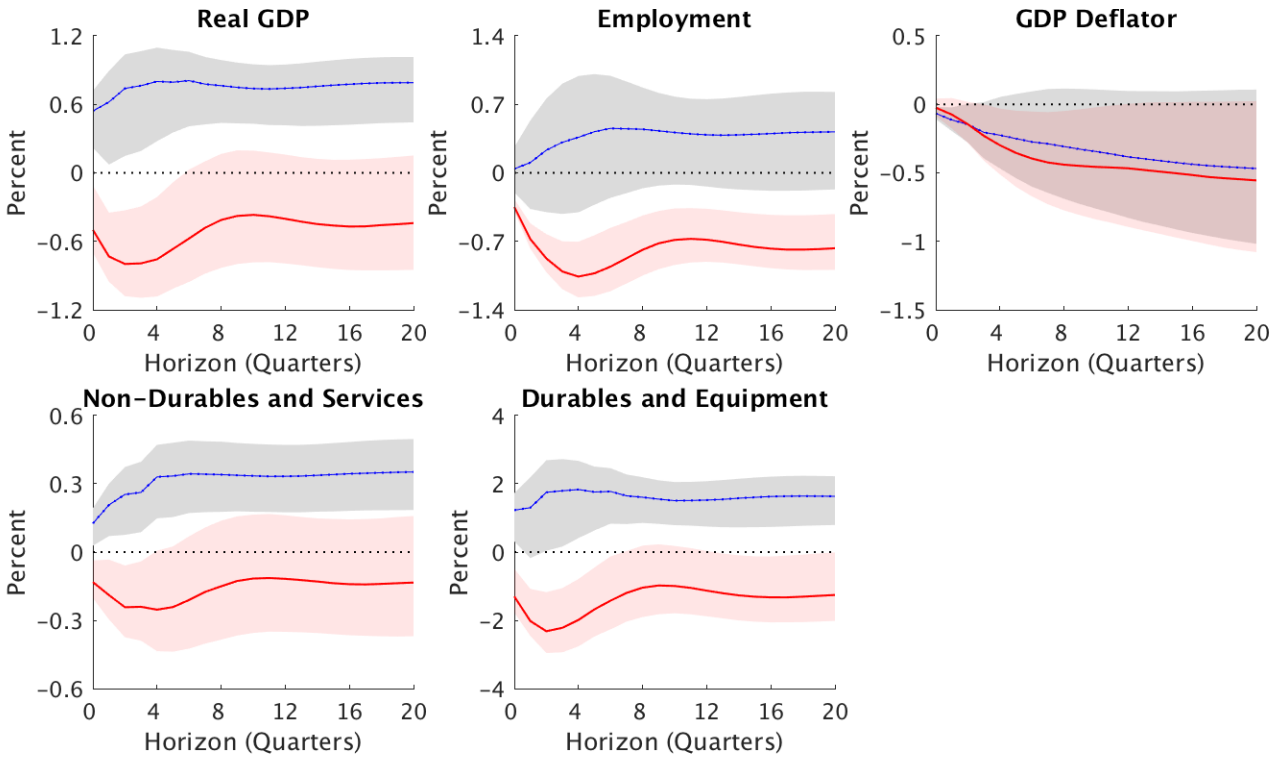


Figure 1: IRFs to a positive one standard deviation consumption and investment TFP shocks. The blue solid-dotted curves represent the point-wise posterior medians and the gray shaded areas represent the 68 percent equal-tailed point-wise probability bands to a consumption TFP shock. The red solid curves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

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

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

where $\boldsymbol{m}_t' = [\boldsymbol{m}_{C,t} \quad \boldsymbol{m}_{I,t}]$ are the proxies for the consumption and investment TFP shocks. As mentioned in

[15] We downloaded the proxies from Kurt Lunsford's website at https://sites.google.com/site/kurtglunsford/research.

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 $\mathbb{E}\left[\boldsymbol{m}_{C,t}\boldsymbol{\varepsilon}_{C,t}\right] > 0$, $\mathbb{E}\left[\boldsymbol{m}_{I,t}\boldsymbol{\varepsilon}_{I,t}\right] > 0$, $\mathbb{E}\left[\boldsymbol{m}_{C,t}\boldsymbol{\varepsilon}_{C,t}\right] > \mathbb{E}\left[\boldsymbol{m}_{C,t}\boldsymbol{\varepsilon}_{I,t}\right]$ and $\mathbb{E}\left[\boldsymbol{m}_{I,t}\boldsymbol{\varepsilon}_{I,t}\right] > \mathbb{E}\left[\boldsymbol{m}_{I,t}\boldsymbol{\varepsilon}_{C,t}\right]$ on the entries of $\boldsymbol{V}$. If we order the two structural shocks of interest last, this implies setting $s = 4$,

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

and

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

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.[16] 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. 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

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

Appendix A.8 we show that these tests imply that the finite variance requirement holds for the application analyzed in this section.

6 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 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, while income responds to AMTR shocks, it does not react to average tax rate (ATR) shocks, indicating that substitution effects instead of wealth effects are crucial for the transmission of tax rate shocks. Third, negative AMTR shocks for taxpayers at the top 1 percent of the income distribution have positive short-run effects on income and economic activity but zero long-run effects (i.e., 4 to 5 years after the shock). In contrast, negative AMTR shocks for tax payers at the bottom 99 percent of the income distribution have zero short-run effects on income and economic activity but positive long-run effects.

Our Bayesian approach will basically replicate Mertens and Montiel-Olea's (2018) first finding. When analyzing their second finding, we will show that the identification scheme in Mertens and Montiel-Olea (2018) imposes a zero restriction on $\boldsymbol{A}_0$. More specifically, they exactly identify both the AMTR and ATR shocks by imposing a zero restriction on the systematic component of tax policies. It is important to note that their zero restriction is linked to a particular Cholesky factorization and, hence, the order of the variables affects the identification. Each ordering implies a different zero restriction on the systematic component of tax policy and, therefore, a different identification of the AMTR and ATR shocks. So, Mertens and Montiel-Olea (2018) are in fact using two different identification schemes. Strangely, when analyzing the IRFs for the AMTR shocks Mertens and Montiel-Olea (2018) use a different identification than when analyzing the IRFs for the ATR shocks. In any case, we think that both of their identification schemes are hard to justify because of the zero restriction on the systematic component of tax policies. Our methods will allow us to substitute this zero restriction for a set of less restrictive sign restrictions to be described below. Accordingly, we set identify the AMTR and ATR shocks. Once this is done, and contrary to what Mertens and Montiel-Olea (2018) report, we find that both substitution and wealth effects seem to play a relevant role for the transmission of tax rate shocks. Because the IRFs for AMTR shocks implied by both orderings (identification schemes) are in

agreement, one could be tempted to conclude that Mertens and Montiel-Olea's (2018) results are robust; in fact, Mertens and Montiel-Olea (2018) seem to fall into this trap by comparing Panels (A) and (B) in Figure 10 of their paper and highlighting that the results are similar. An additional insight of our exercise is that just checking two possible identification schemes is a dangerous practice that can give the researcher a false sense of robustness on her results.

When analyzing Mertens and Montiel-Olea's (2018) third finding, we first show that they place a zero restriction on $\boldsymbol{A}_0$ that is analogous to the one used when disentangling the effects of AMTR relative to ATR shocks. This implies that they exactly identify both the AMTR shocks to the top 1 percent and the AMTR shocks to the bottom 99 percent by imposing a zero restriction on the systematic component of tax policies. Second, we show that in this case—even when one decides to ignore the robustness issues raised above and focuses on two orderings—the IRFs are more sensitive to the ordering of the variables. As in Mertens and Montiel-Olea's (2018) second finding, when reporting results for the shocks to the AMTR of the top 1 percent, they choose a different ordering than when reporting results for the AMTR shocks to the bottom 99 percent. We find this confusing since in fact their results are not comparable because they come from two different identification schemes. In any case, we again substitute the zero restriction for a set of less restrictive sign restrictions to be described below, and we show that our results are very much in line with the results that Mertens and Montiel-Olea (2018) seem to favor when analyzing their results.

In this application we use the dataset built by Mertens and Montiel-Olea (2018), which we downloaded from Karel Mertens's website.[17] A detailed description of the dataset can be found in Appendix A of Mertens and Montiel-Olea (2018).

6.1 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.[18] 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

[17] Link to the dataset: `https://karelmertenscom.files.wordpress.com/2018/01/data_mmo.xlsx`.
[18] Net-of-tax rate is defined as 1 minus the AMTR.

tax rates that the authors reasonably consider to be contemporaneously exogenous changes in the AMTR.[19] Accordingly, the identification of the AMTR shock is achieved by assuming that the proxy is only correlated with the AMTR.[20] 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 $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$.

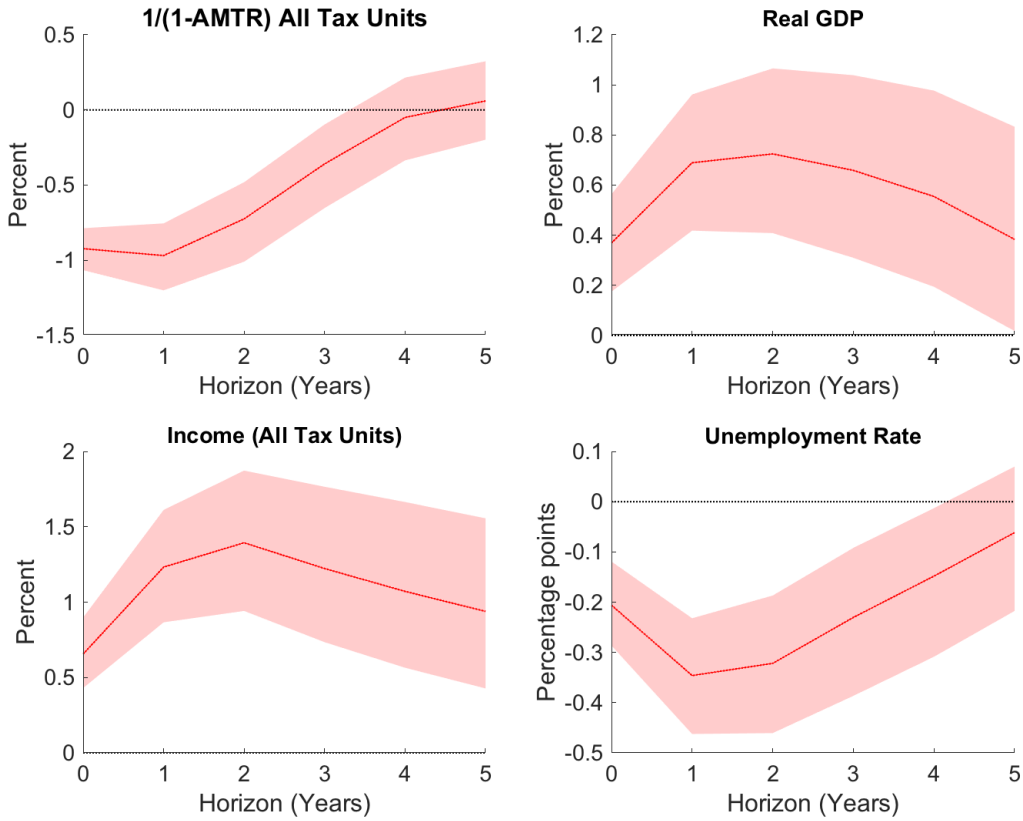


Figure 2: IRFs to a positive AMTR shock (rate cut). The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent equal-tailed point-wise probability bands. The figure is based on 10,000 independent effective draws obtained using Algorithm 2.

We set $\nu = \tilde{n}$, $\mathbf{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\mathbf{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\mathbf{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\mathbf{\Phi}} = \tilde{\mathbf{\Phi}}$, $\hat{\mathbf{\Psi}} = \tilde{\mathbf{\Psi}}$ and $\hat{\mathbf{\Omega}}^{-1} = \tilde{\mathbf{\Omega}}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization.[21] We use Algorithm 2 to obtain an effective sample size of

[19] The 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.

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

[21] In addition, we impose a lower bound on the reliability of the instruments tailored to maximize the efficiency of the importance sampler. In particular, we assume that the minimum eigenvalue of the reliability matrix $\left(\mathbf{\Gamma}_{0,2}^{-1\prime}\mathbf{\Gamma}_{0,2}^{-1} + \boldsymbol{V}\boldsymbol{V}'\right)^{-1}\boldsymbol{V}\boldsymbol{V}'$ implied by the posterior distribution of the structural parameters must be greater or equal than 0.2. This implies that at least 20 percent of the variance of any linear combination of the instruments must be related to the underlying structural shocks of interest. See Mertens and Ravn (2013) and Gleser (1992) for details on the reliability matrix. Without this threshold our algorithm becomes inefficient in the subsequent variant of this application, i.e. Section 6.2. For consistency, we use the same lower bound throughout Section 6 although it only becomes crucial for the efficiency of the algorithm in Section 6.2. We think that this issue is due to the fact that

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 and the negative and sizable IRFs of the unemployment rate coincide with a positive and sizable response of income. Therefore, our results clearly align with those reported in Figure 5 of Mertens and Montiel-Olea (2018).

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.

6.2 Average versus Marginal Tax Rates

The results shown in the previous section imply that AMTR shocks have sizable effects on economic activity. For example, a negative one standard deviation AMTR shock boosts real GDP by about 0.7 percent after one year. Next, Mertens and Montiel-Olea (2018) compare the effects of these shocks with those of ATR shocks in order to understand whether tax policy operates through substitution effects associated with AMTR shocks or through income effects associated with ATR shocks, which ultimately can provide new insights for assessing the economic consequences of tax reforms.

To address this issue, Mertens and Montiel-Olea (2018) expand the SVAR used in Section 6.1 by adding the log ATR as an endogenous variable.[22] They identify AMTR and ATR shocks using two proxies. 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 rates shocks individually. Thus, they also use a particular Cholesky decomposition of a sub-matrix of $(\boldsymbol{A}_0^{-1})'$ to isolate one shock from the other. In Appendix A.10 we show how such a Cholesky decomposition imposes a zero restriction on $\boldsymbol{A}_0$ (and thereby on $\tilde{\boldsymbol{A}}_0$) which implies that the AMTR cannot react contemporaneously to the ATR if the AMTR is ordered first, while the ATR cannot react contemporaneously to the AMTR if ATR ordered first. Evidently, this additional zero restriction is a restriction on the systematic component of tax policy and it will exactly identify both the AMTR and

the two tax instruments to be used in this application have many zeros.

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

ATR shocks. When analyzing Mertens and Montiel-Olea's (2018) results contrasting the effects of AMTR and ATR we will order AMTR first because that is the identification scheme under which they conduct such comparison.

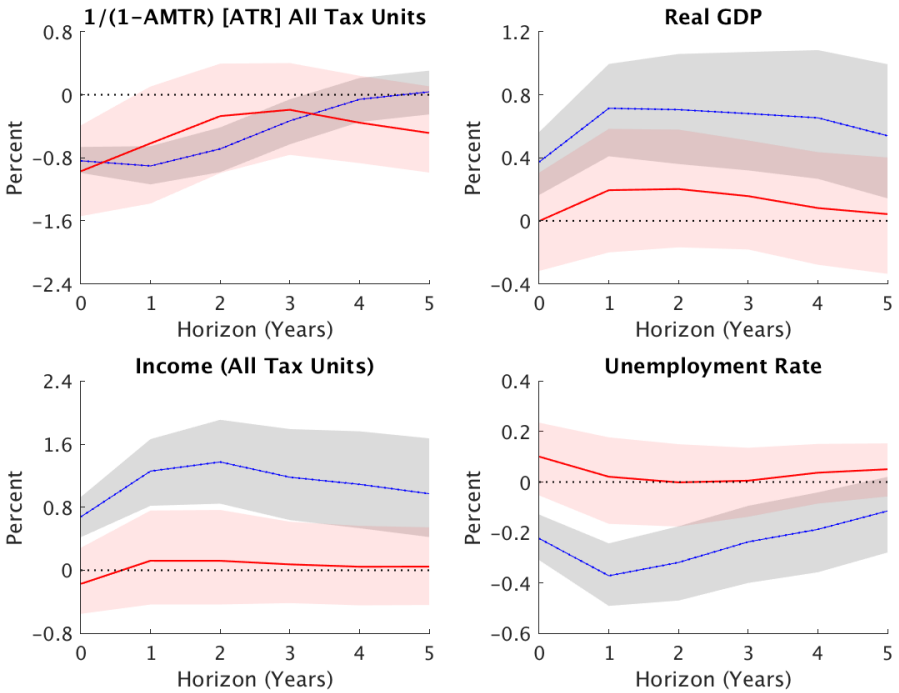
Thus, in this application we have $T = 65$, $n = 10$, $k = 2$, $p = 2$, $\tilde{e} = 2$, and $\tilde{m} = p\tilde{n} + 1 + \tilde{e}$. We set $\nu = \tilde{n}$, $\mathbf{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\mathbf{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\mathbf{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\mathbf{\Phi}} = \hat{\mathbf{\Phi}}_0 \neq \tilde{\mathbf{\Phi}}$, $\hat{\mathbf{\Psi}} = \tilde{\mathbf{\Psi}}$ and $\hat{\mathbf{\Omega}}^{-1} = \tilde{\mathbf{\Omega}}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We choose $\hat{\mathbf{\Phi}}_0$ to maximize the efficiency of the importance sampler.[23] We use Algorithm 4 to obtain an effective sample size of 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 a negative one standard deviation AMTR (gray) and ATR (red) shock when the zero restriction is used. Essentially, this panel replicates Mertens and Montiel-Olea's (2018) exercise from a Bayesian perspective. As the reader can see, the panel closely resembles the IRFs reported in Panels (A) 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 shocks. 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).

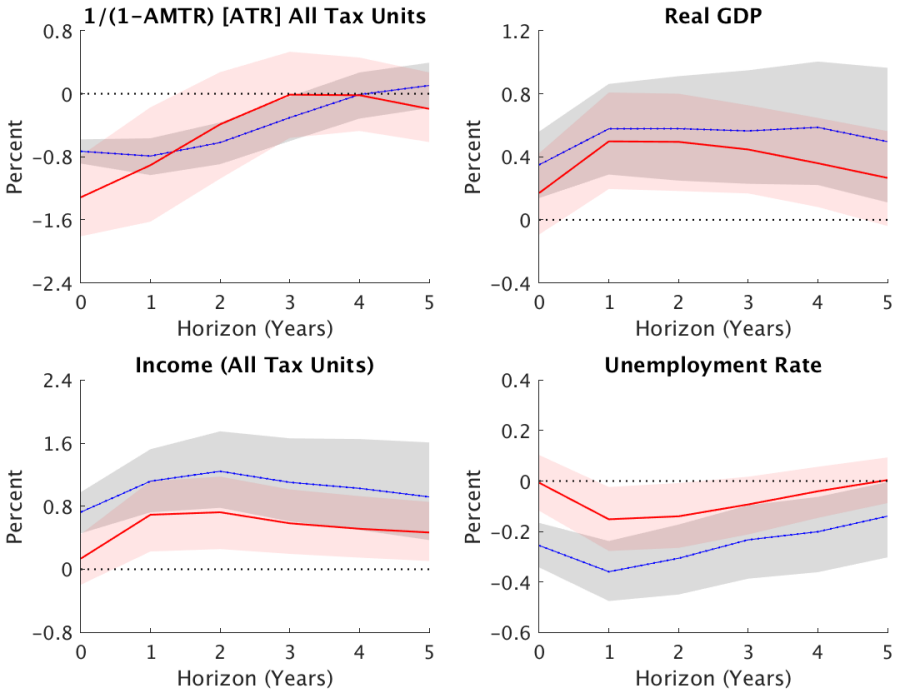
But one may think that the zero restriction imposed by Mertens and Montiel-Olea (2018) is too restrictive. We now check how robust the results are to this zero restriction by substituting it with a set of less restrictive sign restrictions. As we will see below, the results obtained using these sign restrictions 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.*

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



(a) Mertens and Montiel-Olea (2018)



(b) Sign Restrictions

Figure 3: IRFs to a one standard deviation to the AMTR and ATR 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 figure is based on 10,000 independent effective draws obtained using Algorithm 4.

The implementation of our sign restrictions needs a function $\tilde{\boldsymbol{F}}$ and matrices $\boldsymbol{S}_j$ very similar to the ones described in Section 5. In the interest of space, we do not describe them. It is important to emphasize that, while Mertens and Montiel-Olea (2018) exactly identify the two shocks, our sign restrictions set identify them.

Figure 3b uses the set of signs restrictions described above instead of the zero restriction implied by the Cholesky decomposition. The dynamic responses of income, real GDP, and the unemployment rate change substantially relative to those shown in Figure 3a. 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. When looking at real GDP and the unemployment rate we also observe differences in the results. The 68 percent equal-tailed point-wise probability bands for the IRF of real GDP to both shocks are similar when sign restrictions are used. Turning to the unemployment rate, the IRF of the unemployment rate to an ATR shock is now mostly negative. The differences in the results are confirmed when analyzing Table 1.

Table 1: Short-run elasticities of income (Inc) and real GDP to tax shocks

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	1.52	0.12	0.85	0.20
68% Prob. Interval	[0.97; 2.15]	[-0.46; 1.06]	[0.48; 1.28]	[-0.20; 0.83]
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	0.52	0.79	0.38
68% Prob. Interval	[0.92; 2.32]	[0.16; 1.15]	[0.37; 1.30]	[0.13; 0.83]

NOTE: 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 shocks when the zero restrictions are used and when they are substituted by the set of sign restrictions. The short-run elasticities are measured by the ratio between the IRF of income (real GDP) one period after the shock 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 following an AMTR or ATR shock that decreases the corresponding tax rate by about 1 percentage point. The reader can see that, when using the Cholesky decomposition, the 68 percent posterior probability intervals for the short-run elasticities of income and real

GDP to ATR shocks 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 lower than the ones corresponding to the AMTR shocks, the short-run elasticities of income and real GDP to ATR shocks are clearly positive.

Comparing Figures 3a and 3b and reading the results in Table 1, it becomes clear that when using our less restrictive identification scheme it is very difficult to claim that "*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 2) 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 35). It is true that there may be other restrictions consistent with the results in Mertens and Montiel-Olea (2018). Nevertheless, we think that our results make apparent that further analysis is needed before ruling out the income effects of exogenous changes in average tax cut rates.

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.

6.3 Marginal Rates Cuts for the Top and Bottom of the Income Distribution

We now turn to the third result highlighted by Mertens and Montiel-Olea (2018). According to their findings, negative AMTR shocks for taxpayers at the top 1 percent of the income distribution have positive short-run effects on income and economic activity but zero long-run effects (i.e. 4 to 5 years after the shock); in contrast, negative AMTR shocks for tax payers at the bottom 99 percent of the income distribution have zero short-run effects on income and economic activity but positive long-run effects.

As was the case when addressing the effects of AMTR relative to ATR rate cuts, Mertens and Montiel-Olea (2018) modify the SVAR used in Section 6.1 by introducing additional variables germane to the question under study. Specifically, they replace the negative of the aggregate log net-of-tax rate with the negative of the log net-of-tax rate for the top 1 percent and bottom 99 percent of the income distribution, and the aggregate log income level with the log income levels for the top 1 percent and bottom 99 percent of the income distribution. In addition, they modify the reduced-form specification by including a linear and a quadratic trend to capture longer trends in income inequality following Saez (2004) and Saez, Slemrod, and Giertz (2012).

Mertens and Montiel-Olea (2018) parse out the effects of average marginal personal tax rates shocks at the top 1 percent of the income distribution relative to shocks at the bottom 99 percent by using two newly built disaggregated measures of exogenous variation in tax rates across the income distribution as proxies for top

and bottom marginal tax rate shocks. As was the case in Section 6.2, Mertens and Montiel-Olea (2018) exactly identify both tax rate shocks by combining the exogeneity restrictions with an additional zero restriction on $\boldsymbol{A}_0$ (and hence on $\tilde{\boldsymbol{A}}_0$) imposed by means of a Cholesky decomposition of a sub-matrix of $(\boldsymbol{A}_0^{-1})'$. This implies that the AMTR of the top 1 percent cannot react contemporaneously to the AMTR of the bottom 99 percent (or vice-versa) depending on the ordering. Consequently, this additional zero restriction is again a restriction on the systematic component of tax policy. When assessing the effects of an AMTR shock for the top 1 percent they choose the order in which the AMTR for the bottom 99 percent cannot react contemporaneously to the AMTR for the top 1 percent. We will refer to this identification scheme as Case I. When assessing the effects of an AMTR for the bottom 99 percent, Mertens and Montiel-Olea (2018) choose the order in which the AMTR for the top 1 percent cannot react contemporaneously to the AMTR for the bottom 99 percent. We will refer to this identification scheme as Case II.

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 $\nu = \tilde{n}$, $\boldsymbol{\Phi} = \mathbf{0}_{\tilde{n},\tilde{n}}$, $\boldsymbol{\Psi} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{n}^2}$ and $\boldsymbol{\Omega}^{-1} = \mathbf{0}_{\tilde{m}\tilde{n},\tilde{m}\tilde{n}}$ to characterize our prior over the proxy-SVAR structural parameters, and we set $\hat{\nu} = \tilde{\nu}$, $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Phi}}_0 \neq \tilde{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \tilde{\boldsymbol{\Omega}}^{-1}$ to characterize our proposal over the orthogonal-triangular-block parameterization. We choose $\hat{\boldsymbol{\Phi}}_0$ to maximize the efficiency of the importance sampler.[24] We use Algorithm 4 to obtain an effective sample size of 10,000.

Figure 4 shows the IRFs to a one standard deviation shock to the top 1 percent and bottom 99 percent personal income marginal tax rate, respectively, for Case I. The solid-dotted blue line and the gray area show the point-wise posterior medians and the 68 percent equal-tailed point-wise posterior probability bands to an AMTR shock to the top 1 percent. These IRFs replicate the results in Figure 11 of Mertens and Montiel-Olea (2018). The solid red line and the red area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the bottom 99 percent; these IRFs are not reported by Mertens and Montiel-Olea (2018).

Figure 5 shows the IRFs to a one standard deviation shock to the top 1 percent and bottom 99 percent personal income marginal rate, respectively, for Case II. The red solid line and the red area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the bottom 99 percent. These IRFs correspond to the results in Figure 12 of Mertens and Montiel-Olea (2018). By comparing the latter figure with our results, it becomes evident that we find less support for Mertens and Montiel-Olea's (2018) conclusions under this identification scheme. The solid-dotted blue line and the gray area show the point-wise posterior medians and the 68 percent equal-tailed point-wise posterior probability

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

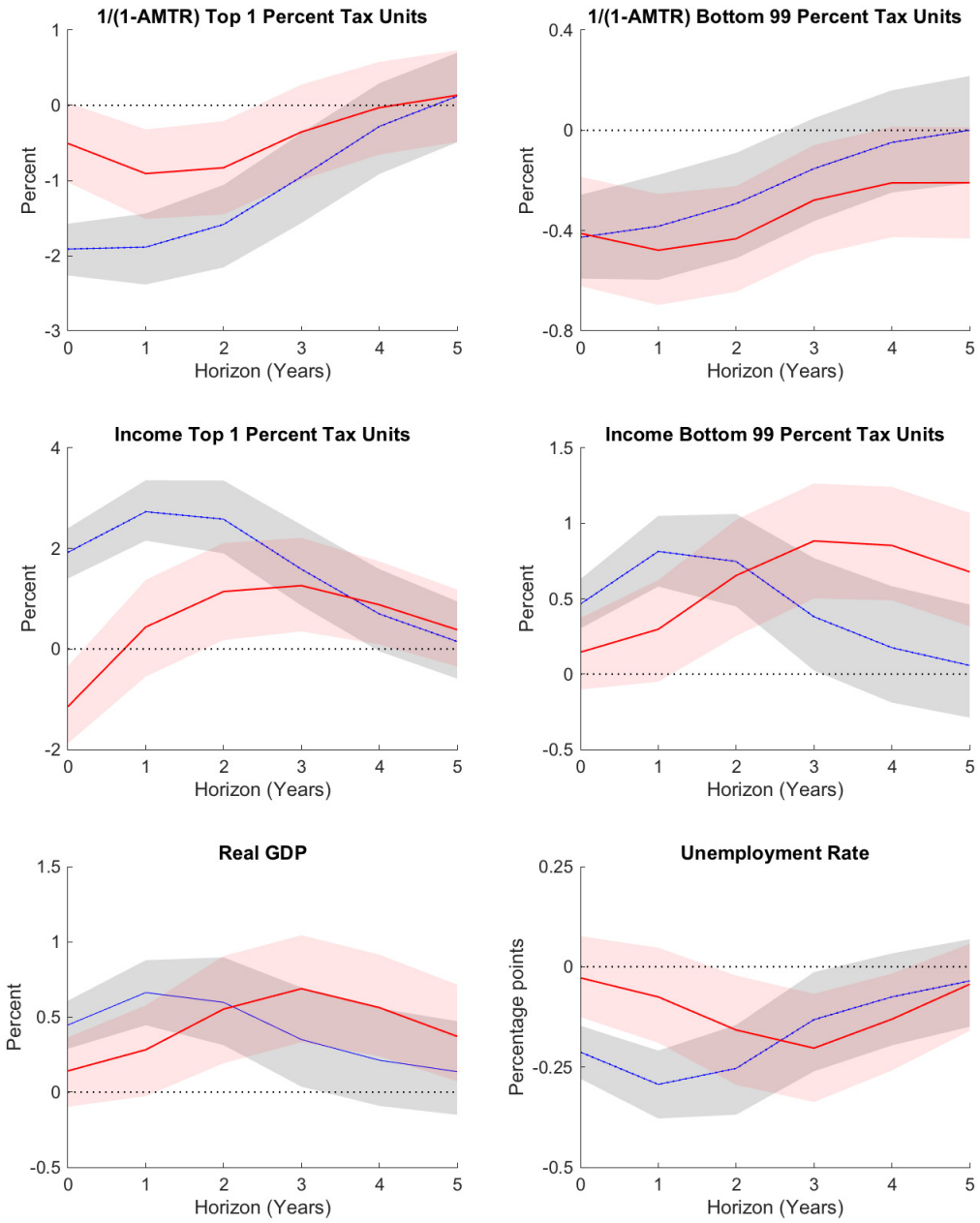


Figure 4: IRFs to a one standard deviation shock to the top 1 percent and bottom 99 percent personal income marginal rate, respectively, identified with the Case I scheme. The solid-dotted blue line and the gray area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the top 1 percent. The solid red line and the red area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the bottom 99 percent. The figure is based on 10,000 independent draws obtained using Algorithm 4.

bands to an AMTR shock to the top 1 percent; these IRFs are not reported by Mertens and Montiel-Olea (2018).

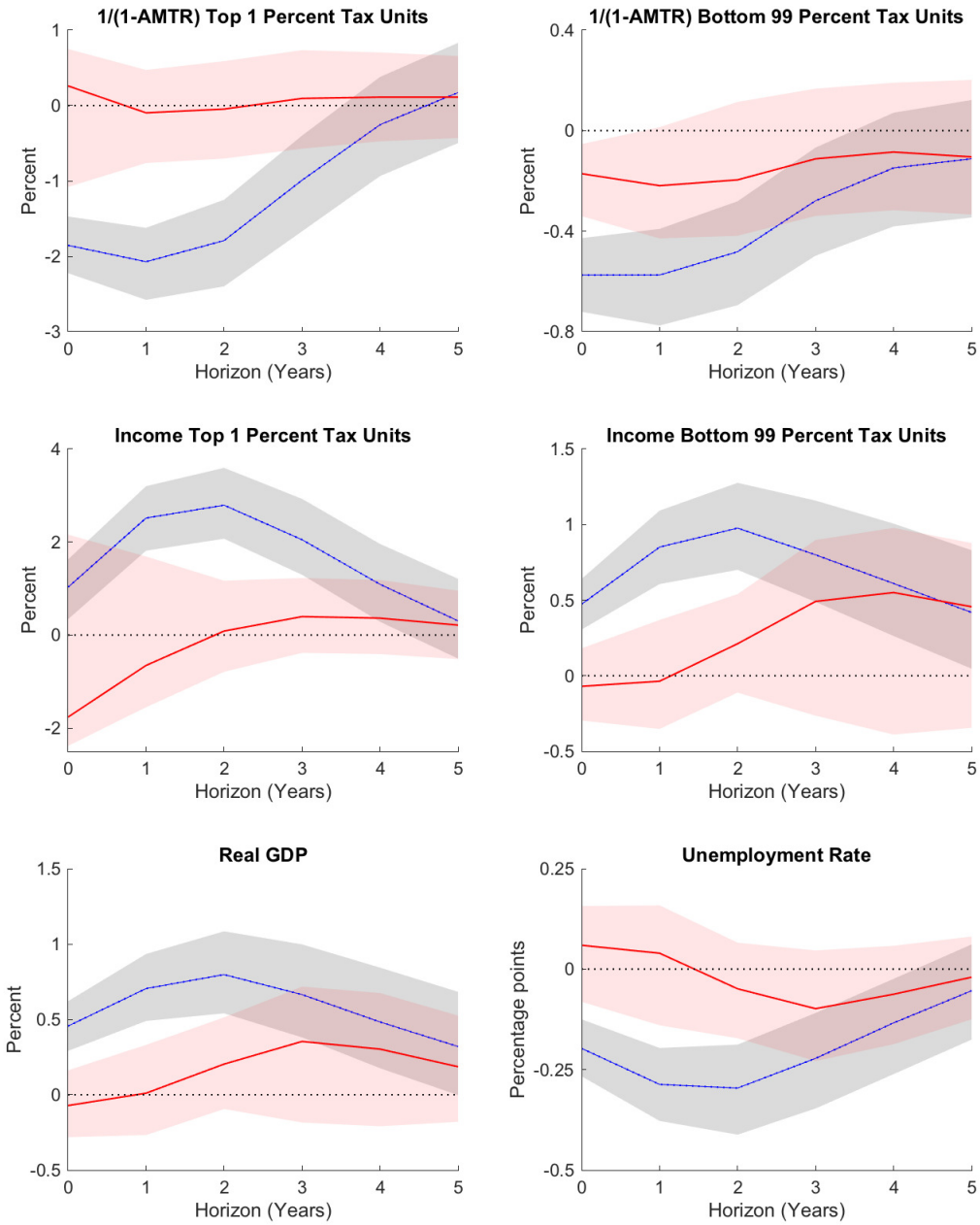


Figure 5: IRFs to a one standard deviation shock to the top 1 percent and bottom 99 percent personal income marginal rate, respectively, identified with the Case II scheme. The solid-dotted blue line and the gray area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the top 1 percent. The solid red line and the red area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the bottom 99 percent. The figure is based on 10,000 independent draws obtained using Algorithm 4.

As can be seen, the IRFs to a one standard deviation shock to the top 1 percent and bottom 99 percent marginal tax rates shown in Figure 4 endorse the view put forward by Mertens and Montiel-Olea (2018).

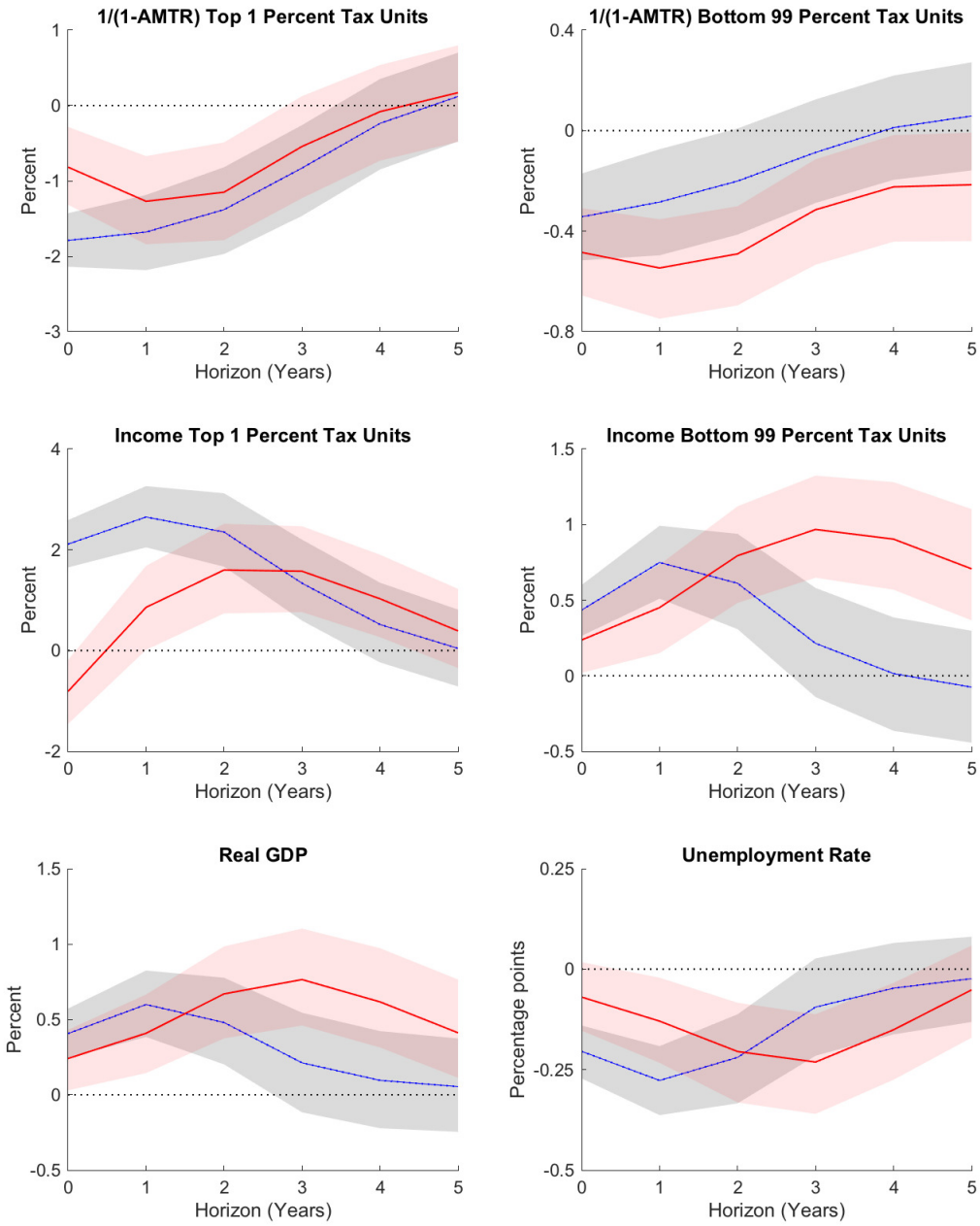


Figure 6: IRFs to a one standard deviation to the top 1 percent (gray) and bottom 99 percent (red) personal income marginal rate. Instruments plus sign restrictions for identifying AMTR shocks to the top 1 percent and bottom 99 percent of the income distribution. The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68 percent equal-tailed point-wise probability bands. The figure is based on 10,000 independent draws obtained using Algorithm 4.

Nevertheless, Figure 5 makes clear that the results on the effects of tax rate cuts shocks to the bottom 99 percent are not robust to using Case II. In particular, even though under Case I the 68 percent posterior probability IRFs for real GDP and the unemployment rate are above and below zero respectively, the 68

percent probability bands for these IRFs do contain zero under Case II. This implies that, while we are able to replicate Mertens and Montiel-Olea's (2018) findings under one ordering, we are not able to replicate them under the alternative order. The fact that a potentially influential result hinges on the ordering of an additional zero restriction makes it less appealing. Furthermore, note that under Case II the IRFs of income for the top 1 percent as well as the IRFs for the AMTR for the top 1 percent to a one standard deviation shock to the bottom 99 percent marginal tax rate are not well identified.

Interestingly, next we will show that the lack of robustness and identification vanishes once we replace Mertens and Montiel-Olea's (2018) zero restriction on tax policy with less restrictive sign restrictions analogous to those used to parse out AMTR from ATR shocks. More specifically, when using the less restrictive sign restrictions described below we are able to confirm Mertens and Montiel-Olea's (2018) conclusions.

Sign Restrictions for Identifying AMTR Shocks to the Top 1 and Bottom 99 percent. *(i) The proxy for the AMTR shock to the top 1 percent is positively correlated with the AMTR shock to the top 1 percent; (ii) The proxy for the AMTR shock to the bottom 99 percent is positively correlated with the AMTR shock to the bottom 99 percent; (iii) The covariance between the AMTR shock to the top 1 percent and the proxy for the AMTR shock to the top 1 percent is bigger than the covariance between the AMTR shock to the bottom 99 percent and the proxy for the AMTR shock to the top 1 percent; and (iv) The covariance between the AMTR shock to the bottom 99 percent and the proxy for the AMTR shock to the bottom 99 percent is bigger than the covariance between the AMTR shock to the top 1 percent and the proxy for the AMTR shock to the bottom 99 percent.*

Analogously to the case in Section 6.2, the implementation of our sign restrictions needs a function $\tilde{\boldsymbol{F}}$ and matrices $\boldsymbol{S}_j$ very similar to the ones described in Section 5. In the interest of space, we do not describe them.

Figure 6 shows the IRFs to a one standard deviation AMTR shock to the top 1 percent and bottom 99 percent, respectively, identified with our sign restrictions. The red solid line and the red area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the bottom 99 percent. The solid-dotted blue line and the gray area show the point-wise posterior medians and the 68 percent equal-tailed posterior probability bands to an AMTR shock to the top 1 percent. It is clear from the figure that Mertens and Montiel-Olea's (2018) conclusions regarding the effects of tax cut rates shocks at the top and bottom of the distribution can be supported on the grounds of the exogeneity restrictions and our sign restrictions.

In Appendix A.12 we show that Koopman, Shephard and Creal's (2009) tests validate the finite variance requirement for the reliable use of Algorithm 4.

7 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 additional and often questionable zero restrictions when more than one instrument is used to identify more than one structural shock. In contrast, our Bayesian approach allows researchers to consider less restrictive identification schemes.

A Appendix

A.1 Proof of Proposition 1

Proof. It is well known that in the class of Gaussian linear models considered in this paper $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ 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

$$\left(\bar{\boldsymbol{A}}_0 \bar{\boldsymbol{A}}_0'\right)^{-1} = \left(\hat{\boldsymbol{A}}_0 \hat{\boldsymbol{A}}_0'\right)^{-1} \tag{8}$$

$$\bar{\boldsymbol{A}}_+ \bar{\boldsymbol{A}}_0^{-1} = \hat{\boldsymbol{A}}_+ \hat{\boldsymbol{A}}_0^{-1}. \tag{9}$$

First, assume that $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent so that Equations (8) and (9) hold. We show that there is $\boldsymbol{Q} \in \mathcal{Q}$ such that $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0 \boldsymbol{Q}$ and $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+ \boldsymbol{Q}$. Let $\boldsymbol{Q} = \hat{\boldsymbol{A}}_0^{-1} \bar{\boldsymbol{A}}_0$. It follows directly from the definition of $\boldsymbol{Q}$ that $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0 \boldsymbol{Q}$, and Equation (9) implies $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+ \boldsymbol{Q}$. We now show that $\boldsymbol{Q} \in \mathcal{Q}$. It follows from Equation (8) that $\boldsymbol{Q}' = \boldsymbol{Q}^{-1}$. Hence, $\boldsymbol{Q}$ is orthogonal. Because the lower left-hand $k \times n$ block of both $\bar{\boldsymbol{A}}_0$ and $\hat{\boldsymbol{A}}_0$ are zero, the same will be true for both $\hat{\boldsymbol{A}}_0^{-1}$ and $\boldsymbol{Q}$. Let

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

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

Now, assume that $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0 \boldsymbol{Q}$ and $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+ \boldsymbol{Q}$ for some matrix $\boldsymbol{Q} \in \mathcal{Q}$. It is easy to see by direct substitution that Equations (8) and (9) hold, which implies that $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent. □

A.2 The Function $\boldsymbol{K}_{i,j}$

In Steps 1, 2, and 3 of Algorithm 1 draws of $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ were obtained.[25] For each draw, Steps 4 and 5 of Algorithm 1 recursively defined matrices $\boldsymbol{M}_{i,j}$ and $\boldsymbol{K}_{i,j}$, though the $\boldsymbol{K}_{i,j}$ were not uniquely defined. Using the $\boldsymbol{K}_{i,j}$, we can define a mapping g of $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ to $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{q}_{i,j})$, where $\boldsymbol{q}_{i,j} = \boldsymbol{K}_{i,j} \boldsymbol{w}_{i,j}$. In this appendix we show that there is an open set $U \subset \mathbb{R}^{\tilde{n}^2 + \tilde{n}\tilde{m} + \sum_{j=1}^{n}(n+1-j-\tilde{z}_j) + \sum_{j=1}^{k}(k+1-j)}$ such that g can be uniquely defined over U so that it is continuously differentiable. Furthermore, U will contain almost all draws of $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$.

[25] In 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 $\boldsymbol{K}_{i,j}$ we need $m_i \times n_{i,j}$ reference matrices $\boldsymbol{W}_{i,j}$ 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 $U_{i,j}$, $\tilde{\boldsymbol{M}}_{i,j}$, and $\boldsymbol{K}_{i,j}$. 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 $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ in $U_{i,j-1}$, define

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

Next, let

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

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

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

be the QR-decomposition of $\tilde{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ normalized so that the diagonal of $\boldsymbol{R}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ is positive and define $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ to be the last $n_{i,j}$ columns of $\boldsymbol{Q}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$. 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, $\boldsymbol{K}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$ 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 $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$. 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 $\tilde{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$, consider the set of all $(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j}) \in U_{i,j-1}$ 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.[26] This completes the proof.

[26] For any set of ℓ columns of $\tilde{\boldsymbol{M}}_{i,j}(\tilde{\boldsymbol{\Lambda}}_0, \tilde{\boldsymbol{\Lambda}}_+, \boldsymbol{w}_{i,j})$, 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 $(\bar{\boldsymbol{A}}_0, \bar{\boldsymbol{A}}_+)$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ are observationally equivalent if and only if there exists $\boldsymbol{Q} \in \mathcal{Q}$ such that $\bar{\boldsymbol{A}}_0 = \hat{\boldsymbol{A}}_0\boldsymbol{Q}$ and $\bar{\boldsymbol{A}}_+ = \hat{\boldsymbol{A}}_+\boldsymbol{Q}$. So to prove Proposition 2, it suffices to prove that if $\boldsymbol{Q} \in \mathcal{Q}$ and $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{A}}_+)$ satisfy the exogeneity restrictions and the relevance condition, then $(\hat{\boldsymbol{A}}_0\boldsymbol{Q}, \hat{\boldsymbol{A}}_+\boldsymbol{Q})$ will satisfy the exogeneity restrictions and relevance condition if and only if $\boldsymbol{Q} \in \mathcal{X} \subset \mathcal{Q}$. Recall from Sections 3.2 and 3.3, the proxy-SVAR parameters $(\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_+)$ will satisfy the relevance condition if and only if $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'$ is of full row rank and will satisfy the exogeneity restrictions if and only if $\boldsymbol{J}(\tilde{\boldsymbol{A}}_0^{-1})'\boldsymbol{L}'\boldsymbol{K}' = \boldsymbol{0}_{k\times n-k}$, where $\boldsymbol{K} = [\boldsymbol{I}_{n-k} \;\; \boldsymbol{0}_{n-k\times k}]$.

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

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

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

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

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

Thus $\boldsymbol{Q}_{21}$ 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 jth columns of $\tilde{\boldsymbol{\Lambda}}_0$ and $\tilde{\boldsymbol{\Lambda}}_+$, respectively, then it is assumed that the $\tilde{\boldsymbol{\lambda}}_{0,j}$ and $\tilde{\boldsymbol{\lambda}}_{+,j}$ that satisfy the restrictions are of the form

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

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

The Gibbs sampler is described in terms of a non-negative scalar T and matrices $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$, for $1 \leq j \leq \tilde{n}$. In that paper, the goal was to sample from a posterior and so T, $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$ were given in terms of restrictions, prior, and data. Our goal is to sample from a normal-generalized-normal distribution conditional on the above restrictions, so we will describe T, $\boldsymbol{H}_j$, $\boldsymbol{P}_j$ and $\boldsymbol{S}_j$ 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 assume that $\hat{\boldsymbol{\Phi}} = \text{diag}(\hat{\boldsymbol{\Phi}}_j)$, $\hat{\boldsymbol{\Psi}} = \text{diag}(\hat{\boldsymbol{\Psi}}_j)$, and $\hat{\boldsymbol{\Omega}} = \text{diag}(\hat{\boldsymbol{\Omega}}_j)$.

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

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

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

$$\boldsymbol{S}_j = \left(\frac{1}{T} (\boldsymbol{U}_j' \hat{\boldsymbol{\Phi}} \boldsymbol{U}_j + \boldsymbol{U}_j' \hat{\boldsymbol{\Psi}}' \hat{\boldsymbol{\Omega}}^{-1} \hat{\boldsymbol{\Psi}} \boldsymbol{U}_j - \boldsymbol{P}_j' \boldsymbol{H}_j^{-1} \boldsymbol{P}_j) \right)^{-1}$$

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 $(\tilde{\nu}, \tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Omega}})$.[27]

[27] In the applications of our paper that rely on this procedure (i.e. Sections 6.2 and 6.3), performing this optimization based on 1,000 draws of the structural parameters suffices to find an efficient choice of $\hat{\boldsymbol{\Phi}}$.

A.6 Data Appendix for Section 5

Here we describe the data used in Section 5 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 5

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 5. 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	-2.36	-2.06	-1.70	-0.86	-0.22
Score	-9.34	-8.07	-6.31	-3.77	-1.25
LR	0	0	0	0	0

Table A.1: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 5.

A.9 Tests for the Analysis in Section 6.1

Table A.2 shows that Koopman, Shephard and Creal's (2009) tests indicate that the importance weights used in Section 6.1 have finite variance.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-4979.60	-4237.73	-3478.55	-1716.32	-355.17
Score	-6.07	-5.58	-5.10	-3.99	-1.63
LR	0	0	0	0	0

Table A.2: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.1.

A.10 Understanding the Additional Zero Restrictions

The reduced-form parameterization implied by Equation (3) is

$$\boldsymbol{y}_t' = \boldsymbol{x}_t'\boldsymbol{B} + \boldsymbol{u}_t',$$

where $\boldsymbol{B} = \boldsymbol{A}_+\boldsymbol{A}_0^{-1}$, $\boldsymbol{u}_t' = \boldsymbol{\varepsilon}_t'\boldsymbol{A}_0^{-1}$, and $\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t'\right] = \boldsymbol{\Sigma} = \left(\boldsymbol{A}_0\boldsymbol{A}_0'\right)^{-1}$. Thus, we have that

$$\boldsymbol{u}_t = (\boldsymbol{A}_0^{-1})'\boldsymbol{\varepsilon}_t, \tag{10}$$

where $\boldsymbol{u}_t$ are the reduced-form innovations. We can rewrite Equation (10) as

$$\begin{bmatrix} \boldsymbol{u}_{1t} \\ \boldsymbol{u}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix} \text{ where } (\boldsymbol{A}_0^{-1})' = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix}$$

and $\boldsymbol{\varepsilon}_{1t}$ are the k structural shocks correlated with the proxies and $\boldsymbol{\varepsilon}_{2t}$ are the $n-k$ structural shocks that are not correlated with the proxies.[28] $\boldsymbol{B}_{11}$ is a $k \times k$ matrix, $\boldsymbol{B}_{12}$ is a $k \times (n-k)$ matrix, $\boldsymbol{B}_{21}$ is a $(n-k) \times (n-k)$ matrix, and $\boldsymbol{B}_{22}$ is a $(n-k) \times (n-k)$ matrix. Hence, we have

$$\boldsymbol{u}_{1t} = \boldsymbol{B}_{11}\boldsymbol{\varepsilon}_{1t} + \boldsymbol{B}_{12}\boldsymbol{\varepsilon}_{2t} \tag{11}$$

and

$$\boldsymbol{u}_{2t} = \boldsymbol{B}_{21}\boldsymbol{\varepsilon}_{1t} + \boldsymbol{B}_{22}\boldsymbol{\varepsilon}_{2t}. \tag{12}$$

From Equation (12), we have that

$$\boldsymbol{\varepsilon}_{2t} = -\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}\boldsymbol{\varepsilon}_{1t} + \boldsymbol{B}_{22}^{-1}\boldsymbol{u}_{2t}. \tag{13}$$

Then, Equations (11) and (13) imply

$$\boldsymbol{u}_{1t} = \boldsymbol{B}_{11}\boldsymbol{\varepsilon}_{1t} + \boldsymbol{B}_{12}\left(-\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}\boldsymbol{\varepsilon}_{1t} + \boldsymbol{B}_{22}^{-1}\boldsymbol{u}_{2t}\right) \tag{14}$$

or

$$\boldsymbol{u}_{1t} = \eta\boldsymbol{u}_{2t} + \boldsymbol{S}_1\boldsymbol{\varepsilon}_{1t}, \tag{15}$$

where $\boldsymbol{\eta} = \boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}$ and $\boldsymbol{S}_1 = \left(\boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}\right)$. Similarly,

$$\boldsymbol{u}_{2t} = \zeta\boldsymbol{u}_{1t} + \boldsymbol{S}_2\boldsymbol{\varepsilon}_{2t}, \tag{16}$$

[28] In Section 2.2 we correlated the last k structural shocks with the proxies. That was without loss of generality. We change the order here to better match the explanations in Mertens and Ravn (2013).

where $\zeta = \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}$ and $\boldsymbol{S}_2 = \left(\boldsymbol{B}_{22} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\right)$. Equations (15) and (16) replicate Equations (15) and (16) in Mertens and Ravn (2013). Using Equation (15) and (16), we get

$$\boldsymbol{u}_t = \begin{bmatrix} \boldsymbol{I}_k & -\eta \\ -\zeta & \boldsymbol{I}_{n-k} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2 \end{bmatrix} \varepsilon_t = \begin{bmatrix} \boldsymbol{I}_k + \eta\left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta & \left(\boldsymbol{I}_k - \eta\zeta\right)^{-1}\eta \\ \left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta & \boldsymbol{I}_{n-k} + \zeta\left(\boldsymbol{I}_k - \eta\zeta\right)^{-1}\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2 \end{bmatrix} \boldsymbol{\varepsilon}_t. \tag{17}$$

Hence, we have that

$$(\boldsymbol{A}_0^{-1})' = \begin{bmatrix} \boldsymbol{I}_k + \eta\left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta & \left(\boldsymbol{I}_k - \eta\zeta\right)^{-1}\eta \\ \left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta & \boldsymbol{I}_{n-k} + \zeta\left(\boldsymbol{I}_k - \eta\zeta\right)^{-1}\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2 \end{bmatrix}. \tag{18}$$

By letting the first column of $(\boldsymbol{A}_0^{-1})'$ be denoted by β_1 we see from Equation (18) that

$$\beta_1 = \begin{bmatrix} \boldsymbol{I}_k + \eta\left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta \\ \left(\boldsymbol{I}_{n-k} - \zeta\eta\right)^{-1}\zeta \end{bmatrix} \boldsymbol{S}_1. \tag{19}$$

Mertens and Ravn (2013) show that for any value of the reduced-form parameters one can solve for $\beta_1\boldsymbol{S}_1^{-1}$ and $\boldsymbol{S}_1\boldsymbol{S}_1'$, but not for $\boldsymbol{S}_1$. In particular, on page 1224 the authors write, "*Ideally one would like to identify $\boldsymbol{S}_1$ but this requires arbitrary assumptions on how personal income taxes respond contemporaneously to unanticipated changes in corporate taxes (beyond the indirect contemporaneous endogenous effects through $\boldsymbol{u}_{2,t}$), and vice versa. Fortunately, knowledge of $\beta_1\boldsymbol{S}_1^{-1}$ still permits economically meaningful structural responses to any linear combination of tax shocks. We report responses that result from a Cholesky decomposition of $\boldsymbol{S}_1\boldsymbol{S}_1'$, imposing that $\boldsymbol{S}_1$ is lower triangular.*" Hence, Mertens and Ravn (2013) seem to claim that imposing that $\boldsymbol{S}_1$ is lower triangular does not imply any additional identification restrictions.

Let's now see what the implications are of imposing that $\boldsymbol{S}_1$ is lower triangular for the matrix describing the contemporaneous relations among the variables, i.e., $\boldsymbol{A}_0$. First, letting $\boldsymbol{L}_1 = \boldsymbol{S}_1^{-1}$, note that

$$\boldsymbol{A}_0 = \begin{bmatrix} \boldsymbol{L}_1' & -\zeta'\boldsymbol{L}_1' \\ -\boldsymbol{\eta}'\boldsymbol{S}_2' & \left(\boldsymbol{S}_2^{-1}\right)' \end{bmatrix}, \tag{20}$$

where $\boldsymbol{L}_1 = \boldsymbol{S}_1^{-1}$. Since $\boldsymbol{S}_1$ is lower triangular, $\boldsymbol{L}_1$ is also lower triangular. Hence, Equation (20) implies that imposing that $\boldsymbol{S}_1$ is lower triangular is indeed imposing additional zero restrictions in $\boldsymbol{A}_0$. Thus, imposing that $\boldsymbol{S}_1$ is lower triangular does indeed imply some additional zero identification restrictions.

When identifying marginal and average tax shocks, Mertens and Montiel-Olea (2018) use two proxies to

identify two structural shocks. Thus, $\boldsymbol{L}_1' = \boldsymbol{U}_1$ is an upper triangular matrix of dimension 2×2. Therefore, we have

$$\boldsymbol{A}_0 = \begin{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} & -\zeta' \left(\boldsymbol{S}_2^{-1}\right)' \\ -\boldsymbol{\eta}' \boldsymbol{L}_1' & \left(\boldsymbol{S}_2^{-1}\right)' \end{bmatrix},$$

which implies that the systematic part of AMTR cannot react contemporaneously to ATR if AMTR is ordered first. If ATR is ordered first, ATR cannot react contemporaneously to AMTR.

A.11 Tests for the Analysis in Section 6.2

Table A.3 presents Koopman, Shephard and Creal's (2009) tests for the case in which the average and marginal tax rates are identified using the restrictions in Mertens and Montiel-Olea (2018) described in Section 6.2.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-17.50	-14.25	-11.04	-4.10	-0.48
Score	-10.63	-9.13	-7.71	-3.95	-0.92
LR	0	0	0	0	0

Table A.3: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.2 for the proxy-SVAR identified with proxy and zero restrictions.

Table A.4 shows that Koopman, Shephard and Creal's (2009) tests indicate that the importance weights used in Section 6.2 for the proxy-SVAR identified with proxy and sign restrictions have finite variance.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-10.23	-9.05	-7.43	-3.62	-0.90
Score	-6.18	-6.03	-5.47	-3.81	-1.72
LR	0	0	0	0	0

Table A.4: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.2 for the proxy-SVAR identified with proxy and sign restrictions.

A.12 Tests for the Analysis in Section 6.3

Table A.5 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 Case I scheme described in Section 6.3.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-34.51	-29.20	-23.77	-10.35	-1.64
Score	-11.94	-9.97	-8.23	-4.15	-0.92
LR	0	0	0	0	0

Table A.5: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.3 for the proxy-SVAR identified using the Case I scheme.

Table A.6 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 Case II scheme described in Section 6.3.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-32.99	-27.91	-22.22	-8.41	-0.72
Score	-11.58	-9.67	-7.89	-3.55	-0.54
LR	0	0	0	0	0

Table A.6: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.3 for the proxy-SVAR identified using the the Case II scheme.

Table A.7 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 6.3.

Threshold	Largest 50%	Largest 40%	Largest 30%	Largest 10%	Largest 1%
Wald	-28.67	-24.77	-20.30	-9.17	-2.03
Score	-10.85	-9.28	-7.52	-3.60	-1.09
LR	0	0	0	0	0

Table A.7: Wald, Score and Likelihood Ratio Tests for the Analysis in Section 6.3 for the proxy-SVAR identified using the less restrictive identification scheme.

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.

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*.

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

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.