

# DISCUSSION PAPER SERIES

DP14626

## **ROBUST BAYESIAN INFERENCE IN PROXY SVARS**

Raffaella Giacomini, Toru Kitagawa and Matthew  
Read

**MONETARY ECONOMICS AND FLUCTUATIONS**



# ROBUST BAYESIAN INFERENCE IN PROXY SVARS

*Raffaella Giacomini, Toru Kitagawa and Matthew Read*

Discussion Paper DP14626

Published 16 April 2020

Submitted 13 April 2020

Centre for Economic Policy Research  
33 Great Sutton Street, London EC1V 0DX, UK  
Tel: +44 (0)20 7183 8801  
[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Raffaella Giacomini, Toru Kitagawa and Matthew Read

# ROBUST BAYESIAN INFERENCE IN PROXY SVARS

## Abstract

We develop methods for robust Bayesian inference in structural vector autoregressions (SVARs) where the parameters of interest are set-identified using external instruments, or 'proxy SVARs'. Set-identification in these models typically occurs when there are multiple instruments for multiple structural shocks. Existing Bayesian approaches to inference in proxy SVARs require researchers to specify a single prior over the model's parameters, but, under set-identification, a component of the prior is never revised. We extend the robust Bayesian approach to inference in set-identified models proposed by Giacomini and Kitagawa (2018) – which allows researchers to relax potentially controversial point-identifying restrictions without having to specify an unrevisable prior – to proxy SVARs. We provide new results on the frequentist validity of the approach in proxy SVARs. We also explore the effect of instrument strength on inference about the identified set. We illustrate our approach by revisiting Mertens and Ravn (2013) and relaxing the assumption that they impose to obtain point identification.

JEL Classification: N/A

Keywords: N/A

Raffaella Giacomini - r.giacomini@ucl.ac.uk  
*UCL and CEPR*

Toru Kitagawa - t.kitagawa@ucl.ac.uk  
*UCL*

Matthew Read - matthew.read.16@ucl.ac.uk  
*UCL*

# Robust Bayesian Inference in Proxy SVARs\*

Raffaella Giacomini<sup>†</sup>, Toru Kitagawa<sup>‡</sup> and Matthew Read<sup>§</sup>

April 10, 2020

## Abstract

We develop methods for robust Bayesian inference in structural vector autoregressions (SVARs) where the parameters of interest are set-identified using external instruments, or ‘proxy SVARs’. Set-identification in these models typically occurs when there are multiple instruments for multiple structural shocks. Existing Bayesian approaches to inference in proxy SVARs require researchers to specify a single prior over the model’s parameters, but, under set-identification, a component of the prior is never revised. We extend the robust Bayesian approach to inference in set-identified models proposed by Giacomini and Kitagawa (2018) – which allows researchers to relax potentially controversial point-identifying restrictions without having to specify an unrevisable prior – to proxy SVARs. We provide new results on the frequentist validity of the approach in proxy SVARs. We also explore the effect of instrument strength on inference about the identified set. We illustrate our approach by revisiting Mertens and Ravn (2013) and relaxing the assumption that they impose to obtain point identification.

---

\*We thank Tetsuya Kaji, Frank Kleibergen, Morten Ravn, two referees and the editor, George Tauchen, for useful feedback. We gratefully acknowledge financial support from ERC grants (numbers 536284 and 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001).

<sup>†</sup>University College London, Department of Economics/Cemmap. Email: r.giacomini@ucl.ac.uk

<sup>‡</sup>University College London, Department of Economics/Cemmap. Email: t.kitagawa@ucl.ac.uk

<sup>§</sup>University College London, Department of Economics. Email: matthew.read.16@ucl.ac.uk

# 1 Introduction

Proxy structural vector autoregressions (SVARs) are an increasingly popular method for estimating the dynamic causal effects of macroeconomic shocks.<sup>1</sup> The key identifying assumption in the proxy SVAR is that there exists one or more variables external to the SVAR – ‘proxies’ or ‘external instruments’ – that are correlated with particular structural shocks (i.e., ‘relevant’) and uncorrelated with all other structural shocks (i.e., ‘exogenous’). The impulse responses to a single structural shock can be point-identified when a single proxy is correlated with that structural shock and uncorrelated with all other structural shocks (Stock 2008). Mertens and Ravn (2013) (henceforth MR13) develop a proxy SVAR with multiple proxies for multiple structural shocks and show that point identification of the impulse responses to these shocks requires zero restrictions on the structural parameters in addition to the zero restrictions implied by exogeneity of the proxies. Other papers that use multiple proxies to identify multiple structural shocks given additional point-identifying restrictions include Lunsford (2015) and Mertens and Montiel-Olea (2018). The additional restrictions required to achieve point identification may not always have a theoretically sound motivation. Consequently, there may be interest in assessing the robustness of the analysis to relaxing these additional restrictions, which would result in set identification.

The majority of the literature that makes use of proxy SVARs conducts inference in the frequentist setting. A notable exception is Arias, Rubio-Ramírez and Waggoner (2019) (henceforth ARW19), who develop algorithms for Bayesian inference that are applicable under set identification. Bayesian inference may be appealing because it allows the researcher to use prior information about the model’s parameters and, under set identification, it may be computationally more convenient than a frequentist approach. This is perhaps why, since Uhlig (2005), the dominant inferential approach in set-identified SVARs has been Bayesian.<sup>2</sup> However, under set identification, posterior inference is sensitive to the choice of prior over the set-identified parameters, even asymptotically (Poirier 1998), and Bayesian credible intervals do not asymptotically coincide with frequentist confidence intervals (Moon and Schorfheide 2012). Moreover, in the context of SVARs, Baumeister and Hamilton (2015) show that even priors that are ‘uniform’ over a set-identified parameter may be informative about the

---

<sup>1</sup>See, for example, Stock and Watson (2012, 2016, 2018), Mertens and Ravn (2013, 2014, 2019), Gertler and Karadi (2015), Lunsford (2015), Ramey (2016), Caldara and Kamps (2017), Mertens and Montiel-Olea (2018), Montiel-Olea, Stock and Watson (2018), Angelini and Fanelli (2019), Arias, Rubio-Ramírez and Waggoner (2019), Caldara and Herbst (2019), Jentsch and Lunsford (2019), Bahaj (forthcoming) and Drautzburg (forthcoming).

<sup>2</sup>Gafarov, Meier and Montiel-Olea (2018) and Granziera, Moon and Schorfheide (2018) develop frequentist inferential tools in set-identified SVARs. We are unaware of papers that conduct frequentist inference in set-identified proxy SVARs.

objects of interest, such as impulse responses.

To address these issues, Giacomini and Kitagawa (2018) (GK18) propose an approach to Bayesian inference in set-identified models that is robust to the choice of prior over the set-identified parameters. The approach considers the class of all priors over the model’s set-identified parameters that are consistent with the identifying restrictions. This generates a class of posteriors, which can be summarised by reporting the set of posterior means (an estimator of the identified set) and a robust credible region. GK18 provide conditions under which these quantities have valid frequentist interpretations and they apply their approach to SVARs in which the impulse responses are set-identified by imposing sign and zero restrictions.

In this paper we extend the approach of GK18 to set-identified proxy SVARs. Following MR13 and ARW19, we consider the case where there are  $k < n$  proxies that are correlated with  $k$  structural shocks (a ‘relevance’ condition) and are uncorrelated with the remaining  $n - k$  shocks (an ‘exogeneity’ condition), where  $n$  is the dimension of the SVAR. If  $n > 3$  and  $1 < k < n - 1$ , the impulse responses to all structural shocks are set-identified in the absence of further zero restrictions on the structural parameters. For other values of  $n$  and  $k$ , it may be the case that impulse responses to particular structural shocks are point-identified, while other impulse responses are set-identified. We focus on cases where the impulse responses of interest are set-identified.

This paper makes several new contributions relative to GK18. First, we provide conditions under which our procedure is guaranteed to have a valid frequentist interpretation in proxy SVARs. These results do not follow directly from those in GK18, and are tailored to the different structure of the problem in proxy SVARs. Second, we show that, in the presence of weak proxies, both the frequentist and the Bayesian approach no longer provide asymptotically valid inference about the identified set: the estimators of the bounds of the identified set are not consistent, they converge to non-degenerate and data-dependent distributions, and these distributions are different for the frequentist and the Bayesian approach (implying a failure of the Bernstein-von Mises property that we prove holds under strong proxies). Third, we show how to conduct posterior inference not only about the impulse responses, but also about the forecast error variance decomposition (FEVD), which is the relative contribution of a particular structural shock to the unexpected variation in a particular variable over some horizon.<sup>3</sup> Finally, we provide an algorithm for computing impulse responses to a unit shock (as opposed to a standard-deviation shock), which are often considered in the

---

<sup>3</sup>Plagborg-Møller and Wolf (2019) develop frequentist procedures for conducting inference about the FEVD in a general semiparametric moving average model when there are valid external instruments available. The setting that they consider allows for cases where the FEVD is set-identified.

proxy-SVAR literature.

As in ARW19, our algorithms allow for zero and sign restrictions on the covariances between the proxies and the structural shocks in addition to the zero restrictions implied by the exogeneity assumption. These types of restrictions are likely to be justifiable in applications, given that the proxies are typically constructed with the purpose of measuring a particular structural shock. An example of a zero restriction would be to assume that, among the  $k$  structural shocks that are assumed to be correlated with the  $k$  proxies, a particular structural shock is uncorrelated with a particular proxy. Examples of sign restrictions are when a particular proxy is positively correlated with a particular structural shock, or when the covariance between a particular proxy and a particular structural shock is larger than the covariance between that proxy and another structural shock.<sup>4</sup> Additionally, our algorithms allow for restrictions of the kind considered in GK18, including ‘short-run’ zero restrictions (as in Sims (1980) and Christiano, Eichenbaum and Evans (1999)), ‘long-run’ zero restrictions (as in Blanchard and Quah (1989)), sign restrictions on impulse responses (as in Uhlig (2005)), and zero or sign restrictions on the matrix determining the contemporaneous relationships among the endogenous variables (as in Arias, Caldara and Rubio-Ramírez (2019)). By extending and adapting the algorithms in GK18 to allow for identification using proxy variables alongside standard zero and sign restrictions, we provide a general and flexible tool for empirical researchers to relax potentially controversial point-identifying restrictions without having to adopt an unrevisable prior.

Some existing approaches to Bayesian inference in proxy SVARs place priors directly on the model’s structural parameters. For example, ARW19 place a normal-generalised-normal conjugate prior over the proxy SVAR’s structural parameters and propose algorithms for drawing from the resulting normal-generalised-normal posterior. More generally, Baumeister and Hamilton (2015, 2018, 2019) advocate placing priors on the structural parameters of an SVAR, because these parameters can have economic interpretations that facilitate prior elicitation. A problem with this approach in set-identified models is that the prior implicitly incorporates a component that is unrevisable by the data. Our approach overcomes this problem by decomposing the prior over the structural parameters into a revisable prior over reduced-form parameters and an unrevisable prior over the orthonormal matrix that maps VAR innovations into structural shocks (see, for example, Uhlig (2005)). We then allow

---

<sup>4</sup>The first type of sign restriction is considered by Ludvigson, Ma and Ng (2018) and Piffer and Podstawski (2018) in the frequentist setting and by Braun and Brüggemann (2017) and ARW19 in the Bayesian setting, while the second type is considered by Braun and Brüggemann (2017), Piffer and Podstawski (2018) and ARW19. Braun and Brüggemann (2017) and Piffer and Podstawski (2018) assume that the proxy is correlated with all structural shocks (i.e., there are no exogeneity restrictions); it would be straightforward to implement this setup under our approach.

for multiple priors for this matrix, which delivers inference that is robust to the choice of unrevisable prior. We see our approach as being complementary to existing Bayesian approaches. In particular, we suggest reporting output based on the multiple-prior robust Bayesian approach together with output from the single-prior Bayesian posterior to document the sensitivity of posterior inference to the choice of unrevisable prior.<sup>5</sup>

It is well-known that frequentist inference in the linear instrumental-variables model is non-standard when the instruments are weakly correlated with the included endogenous variables (e.g., Stock, Wright and Yogo 2002). Similar problems arise in the proxy SVAR when the proxies are weakly correlated with the structural shocks. In the case where there is one proxy for one structural shock, Lunsford (2015) shows that the estimator of the impulse response is inconsistent when the proxy is weak, and he derives a test for the presence of a weak proxy. Montiel-Olea, Stock and Watson (2018) show that standard asymptotic (delta-method) inference about the objects of interest in the proxy SVAR is invalid when the proxy is weak, and they derive a weak-instrument-robust confidence interval for the impulse response when there is one proxy for one shock. As noted in Caldara and Herbst (2019), from the standpoint of Bayesian inference, having a weak proxy does not invalidate posterior inference in the sense that one still obtains (numerical approximations of) the exact finite-sample posterior distributions of the objects of interest. However, practitioners may be interested in the asymptotic frequentist properties of Bayesian inferential procedures. For example, Bayesians may be better able to credibly communicate their results to frequentist audiences when the Bayesian inferential procedure is asymptotically equivalent to a frequentist procedure. Accordingly, we investigate the asymptotic properties of our robust Bayesian procedure in the presence of weak instruments. Using a simple analytical example, we show that our robust Bayesian procedure does not provide valid frequentist inference about the identified set under weak-proxy asymptotics, which contrasts with the results in Kline and Tamer (2016) and GK18. To the best of our knowledge, this is the first paper to provide formal results on the interplay between set identification and weak identification.

We illustrate our procedure by considering the analysis in MR13, which is also discussed in Jentsch and Lunsford (2019) and Mertens and Ravn (2019). MR13 use series of plausibly exogenous, unanticipated changes in average personal and corporate income tax rates in the United States as proxies for structural shocks to these tax rates to identify the effects of fiscal shocks on macroeconomic variables. Since there are two proxies for two structural shocks, the impulse responses to these shocks are set-identified in the absence of additional zero restrictions. MR13 impose a zero restriction in addition to those implied by exogene-

---

<sup>5</sup>An alternative approach is to consider variation in the prior within some neighbourhood around a benchmark prior, as in Giacomini, Kitagawa and Uhlig (2019).



ity of the proxies, which yields point identification. The additional restriction is a causal ordering, which restricts the direct contemporaneous response of one tax rate to the other. This assumption could be violated if, for instance, there are constraints that impinge on the ability of the government to change tax rates independently of one another. MR13 assess the robustness of the results to imposing the additional restriction by considering two alternative causal orderings of the tax rates within the proxy SVAR. Our approach extends and formalizes this robustness analysis by providing an estimator of the set of impulse responses compatible with relaxing the additional zero restriction and replacing it with a set of – arguably weaker – sign restrictions. We compare the results under our multiple-prior Bayesian approach to those obtained under a single prior to assess the role of prior choice in driving posterior inference.

The remainder of the paper is structured as follows. Section 2 describes our robust Bayesian inferential framework for set-identified proxy SVARs. Section 3 provides results on the frequentist properties of this approach and explores how weak proxies affect posterior inference asymptotically. Section 4 details the numerical algorithms used to implement the approach. Section 5 contains the empirical application and Section 6 concludes.

**Generic notation:** For the matrix  $\mathbf{X}$ ,  $\text{vec}(\mathbf{X})$  is the vectorisation of  $\mathbf{X}$  and  $\text{vech}(\mathbf{X})$  is the half-vectorisation of  $\mathbf{X}$  (when  $\mathbf{X}$  is symmetric).  $\mathbf{e}_{i,n}$  is the  $i$ th column of the  $n \times n$  identity matrix,  $\mathbf{I}_n$ .  $\mathbf{0}_{n \times m}$  is a  $n \times m$  matrix of zeros.  $\|\cdot\|$  is the Euclidean norm.  $\mathcal{S}^{n-1}$  is the unit sphere in  $\mathbb{R}^n$ .

## 2 Framework

### 2.1 The SVAR

Let  $\mathbf{y}_t$  be an  $n \times 1$  vector of endogenous variables following the SVAR( $p$ ) process:

$$\mathbf{A}_0 \mathbf{y}_t = \sum_{l=1}^p \mathbf{A}_l \mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (1)$$

where  $\mathbf{A}_0$  has positive diagonal elements (a sign normalization) and is invertible, and  $\boldsymbol{\varepsilon}_t$  are structural shocks with  $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{I}_n$ . The initial conditions  $(\mathbf{y}_{1-p}, \dots, \mathbf{y}_0)$  are given. We omit exogenous regressors (such as a constant) for simplicity of exposition, but these are straightforward to include. Letting  $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$  and  $\mathbf{A}_+ = (\mathbf{A}_1, \dots, \mathbf{A}_p)$ , we can rewrite the SVAR( $p$ ) as

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{A}_+ \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T. \quad (2)$$

$(\mathbf{A}_0, \mathbf{A}_+)$  are the structural parameters. The reduced-form VAR( $p$ ) representation is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u}_t, \quad t = 1, \dots, T, \quad (3)$$

where  $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_p)$ ,  $\mathbf{B}_l = \mathbf{A}_0^{-1}\mathbf{A}_l$  for  $l = 1, \dots, p$ , and  $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t$  with  $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$ .  $(\mathbf{B}, \boldsymbol{\Sigma})$  are the reduced-form parameters. We assume that  $\mathbf{B}$  is such that the VAR( $p$ ) can be inverted into an infinite-order vector moving average (VMA( $\infty$ )) model.<sup>6</sup>

To facilitate computing the identified set of the objects of interest, we reparameterize the model into its ‘orthogonal reduced form’:

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (4)$$

where  $\boldsymbol{\Sigma}_{tr}$  is the lower-triangular Cholesky factor of  $\boldsymbol{\Sigma}$  (i.e.,  $\boldsymbol{\Sigma}_{tr}\boldsymbol{\Sigma}_{tr}' = \boldsymbol{\Sigma}$ ) with diagonal elements normalized to be non-negative,  $\mathbf{Q} \in \mathcal{O}(n)$  is an  $n \times n$  orthonormal matrix and  $\mathcal{O}(n)$  is the set of all such matrices. The parameterizations are related through the mapping  $\mathbf{B} = \mathbf{A}_0^{-1}\mathbf{A}_+$ ,  $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$  and  $\mathbf{Q} = \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$ , or  $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$  and  $\mathbf{A}_+ = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{B}$ . The sign normalization that the diagonal elements of  $\mathbf{A}_0$  are nonnegative therefore corresponds to the restriction that  $\text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}$ .

The VMA( $\infty$ ) representation of the model is

$$\mathbf{y}_t = \sum_{h=0}^{\infty} \mathbf{C}_h \mathbf{u}_{t-h} = \sum_{h=0}^{\infty} \mathbf{C}_h \boldsymbol{\Sigma}_{tr} \mathbf{Q} \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (5)$$

where  $\mathbf{C}_h$  is the  $h$ th term in  $(\mathbf{I}_n - \sum_{l=1}^p \mathbf{B}_l L^l)^{-1}$  and  $L$  is the lag operator. The  $(i, j)$ th element of the matrix  $\mathbf{C}_h \boldsymbol{\Sigma}_{tr} \mathbf{Q}$ , which we denote by  $\eta_{i,j,h}$ , is the impulse response of the  $i$ th variable to the  $j$ th structural shock at the  $h$ th horizon:

$$\eta_{i,j,h} = \mathbf{e}'_{i,n} \mathbf{C}_h \boldsymbol{\Sigma}_{tr} \mathbf{Q} \mathbf{e}_{j,n} = \mathbf{c}'_{i,h} \mathbf{q}_j, \quad (6)$$

where  $\mathbf{c}'_{i,h} \equiv \mathbf{e}'_{i,n} \mathbf{C}_h \boldsymbol{\Sigma}_{tr}$  is the  $i$ th row of  $\mathbf{C}_h \boldsymbol{\Sigma}_{tr}$  and  $\mathbf{q}_j \equiv \mathbf{Q} \mathbf{e}_{j,n}$  is the  $j$ th column of  $\mathbf{Q}$ .

Another object that is also often of interest in analyses using (proxy) SVARs is the FEVD. Under quadratic loss, the optimal  $h$ -step-ahead forecast of  $\mathbf{y}_t$  given information available at time  $t$  is  $\mathbb{E}(\mathbf{y}_{t+h} | \mathcal{F}_t) = \sum_{k=0}^{\infty} \mathbf{C}_{h-k} \mathbf{u}_{t-k}$ . The  $h$ -step-ahead forecast error is then  $\mathbf{y}_{t+h} - \mathbb{E}(\mathbf{y}_{t+h} | \mathcal{F}_t) = \sum_{k=0}^{h-1} \mathbf{C}_k \mathbf{u}_{t+h-k} = \sum_{k=0}^{h-1} \mathbf{C}_k \boldsymbol{\Sigma}_{tr} \mathbf{Q} \boldsymbol{\varepsilon}_{t+h-k}$ . It follows that the forecast error variance of  $y_{i,t+h}$  is  $\text{var}(y_{i,t+h} | \mathcal{F}_t) = \sum_{k=0}^{h-1} \mathbf{c}'_{i,k} \mathbf{c}_{i,k}$ . The contribution of the  $j$ th structural shock to the forecast error variance of the  $i$ th variable at the  $h$ th horizon

<sup>6</sup>The VAR( $p$ ) is invertible into a VMA( $\infty$ ) process when the eigenvalues of the companion matrix lie inside the unit circle. See Hamilton (1994) or Kilian and Lütkepohl (2017).

is  $\text{var}(y_{i,t+h}|\mathcal{F}_t, \boldsymbol{\varepsilon}_{-j,t+1}, \dots, \boldsymbol{\varepsilon}_{-j,t+h}) = \sum_{k=0}^{h-1} \mathbf{c}'_{i,k} \mathbf{q}_j \mathbf{q}'_j \mathbf{c}_{i,k}$ , where  $\boldsymbol{\varepsilon}_{-j,t} = \{\varepsilon_{i,t} : i \neq j \wedge i = 1, \dots, n\}$ . The contribution of the  $j$ th structural shock to the forecast error variance of the  $i$ th variable at the  $h$ th horizon as a fraction of the total forecast error variance is then

$$\text{FEVD}_{i,j,h} = \frac{\sum_{k=0}^{h-1} \mathbf{c}'_{i,k} \mathbf{q}_j \mathbf{q}'_j \mathbf{c}_{i,k}}{\sum_{k=0}^{h-1} \mathbf{c}'_{i,k} \mathbf{c}_{i,k}}. \quad (7)$$

## 2.2 Identification Using Proxies

In the absence of identifying restrictions, the structural parameters – and any function of these parameters, such as the impulse responses or FEVD – are set-identified. Since any  $\mathbf{A}_0 = \mathbf{Q}' \boldsymbol{\Sigma}_{tr}^{-1}$  satisfies  $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1} (\mathbf{A}_0^{-1})'$ , the identified set for  $\mathbf{A}_0$  is  $\{\mathbf{A}_0 = \mathbf{Q}' \boldsymbol{\Sigma}_{tr}^{-1} : \mathbf{Q} \in \mathcal{O}(n)\}$ . Imposing identifying restrictions restricts  $\mathbf{Q}$  to lie in a subspace  $\mathcal{Q}$  of  $\mathcal{O}(n)$ , which shrinks the identified set.

The key identifying assumption in the proxy SVAR is that there are variables external to the SVAR that are correlated with particular structural shocks and uncorrelated with all other structural shocks. Let  $\boldsymbol{\varepsilon}_{(i:j),t} = (\varepsilon_{i,t}, \varepsilon_{i+1,t}, \dots, \varepsilon_{j-1,t}, \varepsilon_{j,t})'$  for  $i < j$ . Assume that  $\mathbf{m}_t$  is a  $k \times 1$  vector of proxies (with  $k < n$ ) that are correlated with the last  $k$  structural shocks, so  $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(n-k+1:n),t}) = \boldsymbol{\Psi}$ , where  $\boldsymbol{\Psi}$  is a full-rank  $k \times k$  matrix. Further, assume that  $\mathbf{m}_t$  is uncorrelated with the first  $n - k$  structural shocks, so  $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:n-k),t}) = \mathbf{0}_{k \times (n-k)}$ . The first condition is commonly referred to as the ‘relevance’ condition and the second as the ‘exogeneity’ condition. We assume that  $\mathbf{m}_t$  is generated by the process

$$\boldsymbol{\Gamma}_0 \mathbf{m}_t = \boldsymbol{\Lambda} \boldsymbol{\varepsilon}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\nu}_t, \quad t = 1, \dots, T, \quad (8)$$

where:  $\boldsymbol{\Gamma}_l, l = 0, \dots, p_m$ , is a  $k \times k$  matrix with  $\boldsymbol{\Gamma}_0$  invertible;  $\boldsymbol{\Lambda}$  is a  $k \times n$  matrix; and the initial conditions  $(\mathbf{m}_{1-p_m}, \dots, \mathbf{m}_0)$  are given. We assume that  $(\boldsymbol{\varepsilon}'_t, \boldsymbol{\nu}'_t)' | \mathcal{F}_{t-1} \sim N(\mathbf{0}_{(n+k) \times 1}, \mathbf{I}_{n+k})$ , where  $\mathcal{F}_{t-1}$  is the information set at time  $t - 1$ , which includes the lags of  $\mathbf{y}_t$  and  $\mathbf{m}_t$ . The assumption about the joint distribution of  $(\boldsymbol{\varepsilon}_t, \boldsymbol{\nu}_t)$  implies that  $\boldsymbol{\nu}_t | \mathcal{F}_{t-1}, \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{k \times 1}, \mathbf{I}_k)$ . This process is an SVAR( $p_m$ ) in  $\mathbf{m}_t$  where the structural shocks  $\boldsymbol{\varepsilon}_t$  are included as exogenous variables. The process implies that the proxies contain information about the structural shocks after allowing for possible serial correlation in the proxies.<sup>7</sup> The information content of each proxy for each structural shock is jointly determined by the matrices  $\boldsymbol{\Gamma}_0$  and  $\boldsymbol{\Lambda}$ . This setup allows for the number of lags of  $\mathbf{m}_t$  in the SVAR for  $\mathbf{m}_t$  to differ to the number of lags

<sup>7</sup>We could also allow for up to  $p$  lags of  $\mathbf{y}_t$  to appear in (8) without altering the reduced form for  $\mathbf{m}_t$  or the restrictions on  $\mathbf{Q}$  implied by proxy exogeneity (derived below).

of  $\mathbf{y}_t$  in the SVAR for  $\mathbf{y}_t$ .<sup>8</sup>

Given the distributional assumption on  $\boldsymbol{\varepsilon}_t$  and  $\boldsymbol{\nu}_t$ , and the exogeneity and relevance assumptions, it follows from (8) that

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} = [\mathbf{0}_{k \times (n-k)}, \boldsymbol{\Psi}]. \quad (9)$$

Left-multiplying (8) by  $\boldsymbol{\Gamma}_0^{-1}$  and substituting out  $\mathbf{y}_t$  using (2) yields

$$\mathbf{m}_t = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_0 \mathbf{y}_t - \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_+ \mathbf{x}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\nu}_t. \quad (10)$$

The reduced-form process for the proxies, which we refer to as the ‘first-stage regression’, is

$$\mathbf{m}_t = \mathbf{D} \mathbf{y}_t + \mathbf{G} \mathbf{x}_t + \sum_{l=1}^{p_m} \mathbf{H}_l \mathbf{m}_{t-l} + \mathbf{v}_t, \quad (11)$$

where:  $\mathbf{D} = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_0$ ;  $\mathbf{G} = -\boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_+$ ;  $\mathbf{H}_l = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_l$  for  $l = 1, \dots, p_m$ ; and  $\mathbf{v}_t = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\nu}_t$  with  $\mathbb{E}(\mathbf{v}_t \mathbf{v}_t') = \boldsymbol{\Upsilon} = \boldsymbol{\Gamma}_0^{-1} (\boldsymbol{\Gamma}_0^{-1})'$ . This is a VAR( $p_m$ ) in  $\mathbf{m}_t$  with exogenous variables  $\mathbf{y}_t$  and  $\mathbf{x}_t$ . The first-stage regression should also include any exogenous variables (e.g., a constant) that are included in the SVAR for  $\mathbf{y}_t$ . Since  $\boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} = \mathbf{D} \boldsymbol{\Lambda}_0^{-1} = \mathbf{D} \boldsymbol{\Sigma}_{tr} \mathbf{Q}$ , we can write (9) as

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}_t') = \mathbf{D} \boldsymbol{\Sigma}_{tr} \mathbf{Q} = [\mathbf{0}_{k \times (n-k)}, \boldsymbol{\Psi}]. \quad (12)$$

The  $(i, j)$ th element of this matrix is  $\mathbf{e}_{i,k}' \mathbf{D} \boldsymbol{\Sigma}_{tr} \mathbf{Q} \mathbf{e}_{j,n} = \mathbf{d}_i' \mathbf{q}_j$ , where  $\mathbf{d}_i' \equiv \mathbf{e}_{i,k}' \mathbf{D} \boldsymbol{\Sigma}_{tr}$  is the  $i$ th row of  $\mathbf{D} \boldsymbol{\Sigma}_{tr}$ . The exogeneity assumption therefore generates linear restrictions on the first  $n - k$  columns of  $\mathbf{Q}$  given the reduced-form parameters  $\mathbf{D}$  and  $\boldsymbol{\Sigma}_{tr}$ . The proxies satisfy the relevance assumption  $\text{rank}(\boldsymbol{\Psi}) = k$  if and only if  $\text{rank}(\mathbf{D}) = k$ .

Let  $f_i$  be the number of equality restrictions on the  $i$ th column of  $\mathbf{Q}$ . Rubio-Ramírez, Waggoner and Zha (2010) show that a necessary and sufficient condition for point identification of the structural parameters in an SVAR is that  $f_i = n - i$  for  $i = 1, \dots, n$ . We focus on cases where  $f_i \leq n - i$  for all  $i = 1, \dots, n$ , with strict inequality for at least one  $i$ , and where interest is in a particular set-identified object. Equations (6) and (7) imply that the impulse response and FEVD corresponding to the  $j$ th structural shock are point-identified if and only if the  $j$ th column of  $\mathbf{Q}$  is point-identified. Assume for now that the only zero restrictions are those corresponding to the exogeneity assumption and that  $n \geq 3$ . Assume also that

---

<sup>8</sup>ARW19 specify a joint SVAR for  $(\mathbf{y}_t', \mathbf{m}_t')$  where zero restrictions rule out feedback from  $\mathbf{m}_t$  to  $\mathbf{y}_t$ . This process also implies that the proxies contain information about the structural shocks and, under the exogeneity assumption, yields the same set of identifying zero restrictions that we derive below.

$\text{rank}(\mathbf{D}) = k$ , so the relevance condition holds. If  $k = 1$ , then  $f_i = 1$  for  $i = 1, \dots, n - 1$  and  $f_n = 0$ . In this case, the first  $n - 1$  columns of  $\mathbf{Q}$  are set-identified and  $\mathbf{q}_n$  is point-identified.<sup>9</sup> If  $k = n - 1$ , then  $f_1 = n - 1$  and  $f_i = 0$  for  $i = 2, \dots, n$ . In this case,  $\mathbf{q}_1$  is point-identified and  $\mathbf{q}_i$ ,  $i = 2, \dots, n$ , is set-identified.<sup>10</sup> For  $1 < k < n - 1$ , all columns of  $\mathbf{Q}$  are set-identified.<sup>11</sup>

As in ARW19, we allow for additional equality and sign restrictions on elements of  $\Psi$ . An example of an equality restriction is that the first proxy variable ( $m_{1t}$ ) is not only uncorrelated with the first  $n - k$  structural shocks, but also uncorrelated with one of the last  $k$  structural shocks (e.g.,  $\mathbb{E}(m_{1t}\varepsilon_{(n-k+1),t}) = 0$ ). This type of restriction is a linear equality restriction on a single column of  $\mathbf{Q}$ . An example of a sign restriction is that the covariance between the first proxy and one of the last  $k$  structural shocks is nonnegative (e.g.,  $\mathbb{E}(m_{1t}\varepsilon_{nt}) \geq 0$ ), which is a linear inequality restriction on a single column of  $\mathbf{Q}$ . Another example is that the covariance between a particular proxy and a particular structural shock is greater than or equal to the covariance between that proxy and another structural shock, which is a linear inequality restriction on two columns of  $\mathbf{Q}$ ; for example,  $\mathbb{E}(m_{1t}\varepsilon_{nt}) \geq \mathbb{E}(m_{1t}\varepsilon_{n-1,t})$  implies that  $\mathbf{d}'_1(\mathbf{q}_n - \mathbf{q}_{n-1}) \geq 0$ .

Our approach also allows for other restrictions commonly used in SVARs, such as zero restrictions on  $\mathbf{A}_0 = \mathbf{Q}'\Sigma_{tr}^{-1}$ ,  $\mathbf{A}_0^{-1} = \mathbf{Q}\Sigma_{tr}$  or the long-run cumulative impulse response  $CIR^\infty = (\mathbf{I}_n - \sum_{l=1}^p \mathbf{B}_l)^{-1}\Sigma_{tr}\mathbf{Q}$ , and sign restrictions on the impulse responses or  $\mathbf{A}_0$ .

## 2.3 Robust Bayesian Inference

We assume for now that the object of interest is the impulse response  $\eta_{i,j,h}$ , although the discussion in this section also applies to the FEVD or any other scalar-valued function of the structural parameters. Given the formulation of the exogeneity restrictions and any additional zero or sign restrictions as restrictions on the columns of  $\mathbf{Q}$ , robust Bayesian inference about the identified set for  $\eta_{i,j,h}$  proceeds as in GK18. We summarise the salient features of this approach here.

Collect the coefficients on  $\mathbf{x}_t$  and  $\mathbf{m}_t$  in (11) as  $\mathbf{J} = [\text{vec}(\mathbf{G})', \text{vec}(\mathbf{H}_1)', \dots, \text{vec}(\mathbf{H}_{p_m})']'$ .

<sup>9</sup>The exogeneity restrictions imply that  $\mathbf{d}'_i \mathbf{q}_i = 0$  for  $i = 1, \dots, n - 1$ . Since the columns of an orthonormal matrix are orthogonal and have unit length,  $\mathbf{q}_n = \pm \mathbf{d}_1 / \|\mathbf{d}_1\|$ . The sign normalization pins down the sign of  $\mathbf{q}_n$ .

<sup>10</sup>The exogeneity restrictions imply that  $\mathbf{D}\Sigma_{tr}\mathbf{q}_1 = 0$ , where  $\mathbf{D}$  is a  $(n - 1) \times n$  matrix. Under the relevance assumption,  $\text{rank}(\mathbf{D}\Sigma_{tr}) = n - 1$  and the nullspace of  $\mathbf{D}\Sigma_{tr}$  is of dimension one by the rank-nullity theorem.  $\mathbf{q}_1$  is therefore a unit-length vector in the (one-dimensional) nullspace of  $\mathbf{D}\Sigma_{tr}$ , which is uniquely determined given the sign normalization.

<sup>11</sup>The result for  $k = 1$  corresponds to Corollary 2 in ARW19. The results for  $k = n - 1$  and  $1 < k < n - 1$  follow from their Proposition 2.

We denote the proxy-SVAR reduced-form parameters as

$$\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})', \text{vec}(\mathbf{D})', \mathbf{J}', \text{vech}(\boldsymbol{\Upsilon})')' \in \boldsymbol{\Phi}. \quad (13)$$

Since the zero restrictions are linear equality restrictions on single columns of  $\mathbf{Q}$  and are otherwise functions only of the reduced-form parameters, we can represent them in the general form

$$\mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \begin{bmatrix} \mathbf{F}_1(\boldsymbol{\phi})\mathbf{q}_1 \\ \vdots \\ \mathbf{F}_n(\boldsymbol{\phi})\mathbf{q}_n \end{bmatrix} = \mathbf{0}_{(\sum_{i=1}^n f_i) \times 1}, \quad (14)$$

where  $\mathbf{F}_i(\boldsymbol{\phi})$  is an  $f_i \times n$  matrix that stacks the coefficient vectors of the zero restrictions constraining  $\mathbf{q}_i$ . If the zero restrictions do not constrain  $\mathbf{q}_i$ ,  $\mathbf{F}_i(\boldsymbol{\phi})$  does not exist and  $f_i = 0$ . We represent the sign restrictions as  $S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}$ , where  $s$  is the number of sign restrictions (excluding the sign normalization). If there are no sign restrictions, then  $S(\boldsymbol{\phi}, \mathbf{Q})$  does not exist and  $s = 0$ .

To simplify implementation of the robust Bayesian inferential approach, we order the variables in  $\mathbf{y}_t$  to satisfy Definition 1.

**Definition 1 (Ordering of Variables):** Given an ordering of the proxies in  $\mathbf{m}_t$ , order the variables in  $\mathbf{y}_t$  so that  $f_i$  satisfies  $f_1 \geq f_2 \geq \dots \geq f_n \geq 0$ . In case of ties, if the impulse response of interest is to the  $j^*$ th structural shock, order the  $j^*$ th variable first. That is, set  $j^* = 1$  when no other column of  $\mathbf{Q}$  has a larger number of restrictions than  $\mathbf{q}_{j^*}$ . If  $j^* \geq 2$ , order the variables so that  $f_{j^*-1} > f_{j^*}$ .

This ordering convention is used when iteratively constructing columns of  $\mathbf{Q}$  satisfying the zero restrictions. It is an extension of the ordering convention used by Rubio-Ramírez, Waggoner and Zha (2010) in the point-identified setting to allow for set identification due to sign restrictions and underidentifying zero restrictions, and mirrors Definition 3 in GK18. The ordering convention uniquely determines  $j^*$ , but the ordering of the remaining variables will not be unique when  $f_i = f_k$  for some  $i, k \neq j^*$ . However, re-ordering the remaining variables will have no effect on the results of our algorithms, as long as the ordering convention is satisfied. The following example illustrates how to order the variables to satisfy Definition 1 and, given the ordering, how the matrices of restrictions are constructed.

**Example 2.1.** Consider a proxy SVAR for  $(c_t, i_t, y_t, \pi_t)$ , where  $c_t$  is consumption growth,  $i_t$  is investment growth,  $y_t$  is output growth and  $\pi_t$  is inflation. Assume that there exist two proxy variables,  $\mathbf{m}_t = (m_{c,t}, m_{i,t})'$ , which are correlated with the structural shocks  $\varepsilon_{c,t}$  and

$\varepsilon_{i,t}$ , and are uncorrelated with  $\varepsilon_{y,t}$  and  $\varepsilon_{\pi,t}$ . In the absence of additional zero restrictions, all impulse responses are set-identified. If the impulse response of interest is to  $\varepsilon_{i,t}$ , an ordering of the variables that satisfies Definition 1 is  $(y_t, \pi_t, i_t, c_t)$ , with  $(f_1, f_2, f_3, f_4) = (2, 2, 0, 0)$  and  $j^* = 3$ . If, instead, the impulse response of interest is to  $\varepsilon_{\pi,t}$ , an ordering of the variables that satisfies Definition 1 is  $(\pi_t, y_t, i_t, c_t)$ , with  $(f_1, f_2, f_3, f_4) = (2, 2, 0, 0)$  and  $j^* = 1$ . In both cases,  $\mathbf{F}_1(\boldsymbol{\phi}) = \mathbf{F}_2(\boldsymbol{\phi}) = \mathbf{D}\boldsymbol{\Sigma}_{tr}$  is a  $2 \times 4$  matrix.

In the case where  $j^* = 3$ , consider the additional sign restrictions  $\mathbb{E}(m_{c,t}\varepsilon_{c,t}) \geq 0$ ,  $\mathbb{E}(m_{i,t}\varepsilon_{i,t}) \geq 0$ ,  $\mathbb{E}(m_{c,t}\varepsilon_{c,t}) \geq \mathbb{E}(m_{c,t}\varepsilon_{i,t})$  and  $\mathbb{E}(m_{i,t}\varepsilon_{i,t}) \geq \mathbb{E}(m_{i,t}\varepsilon_{c,t})$ . The matrix of sign restrictions can be represented as

$$\mathbf{S}(\boldsymbol{\phi}, \mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{1 \times 4} & \mathbf{0}_{1 \times 4} & \mathbf{0}_{1 \times 4} & \mathbf{d}'_1 \\ \mathbf{0}_{1 \times 4} & \mathbf{0}_{1 \times 4} & \mathbf{d}'_2 & \mathbf{0}_{1 \times 4} \\ \mathbf{0}_{1 \times 4} & \mathbf{0}_{1 \times 4} & -\mathbf{d}'_1 & \mathbf{d}'_1 \\ \mathbf{0}_{1 \times 4} & \mathbf{0}_{1 \times 4} & \mathbf{d}'_2 & -\mathbf{d}'_2 \end{bmatrix} \text{vec}(\mathbf{Q}) \geq \mathbf{0}_{4 \times 1}. \quad (15)$$

The identified set for the impulse response  $\eta_{i,j,h}$  given the zero and sign restrictions is

$$IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F, S) = \{\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi}|F, S)\}, \quad (16)$$

where  $\mathcal{Q}(\boldsymbol{\phi}|F, S)$  is the set of orthonormal matrices that satisfy the zero and sign restrictions and the sign normalization:

$$\mathcal{Q}(\boldsymbol{\phi}|F, S) = \{\mathbf{Q} \in \mathcal{O}(n) : \mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{(\sum_i^n f_i) \times 1}, \mathbf{S}(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}, \text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}\}. \quad (17)$$

Let  $\pi_\phi$  be a prior over the reduced-form parameter  $\boldsymbol{\phi}$ . A joint prior for  $\boldsymbol{\theta} = (\boldsymbol{\phi}', \text{vec}(\mathbf{Q})')' \in \boldsymbol{\Phi} \times \text{vec}(\mathcal{O}(n))$  can be written as  $\pi_\theta = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_\phi$ , where  $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$  is supported only on  $\mathcal{Q}(\boldsymbol{\phi}|F, S)$ . Under point identification, the identifying restrictions pin down a unique value of  $\mathbf{Q}$  given  $\boldsymbol{\phi}$ . Consequently, specifying a prior for  $\boldsymbol{\phi}$  is sufficient to induce a single prior – and thus a single posterior – for  $\boldsymbol{\theta}$ . In the set-identified case, the identifying restrictions do not uniquely determine  $\mathbf{Q}$  given  $\boldsymbol{\phi}$ , so specifying a prior for the reduced-form parameters does not induce a single prior for  $\boldsymbol{\theta}$  and thus does not yield a single posterior. Following Uhlig (2005), the vast majority of the empirical literature using Bayesian methods in set-identified SVARs imposes a single prior for  $\mathbf{Q}|\boldsymbol{\phi}$ , including ARW19 in their set-identified proxy SVARs. However, while the prior for  $\boldsymbol{\phi}$  is updated by the data, the conditional prior for  $\mathbf{Q}|\boldsymbol{\phi}$  is not updated, even asymptotically, because the likelihood does not depend on  $\mathbf{Q}$  (Poirier 1998; Moon and Schorfheide 2012). This is problematic, because posterior inference may be driven by an arbitrary prior for  $\mathbf{Q}$ , which has no direct economic interpretation, and even a uniform

prior over  $\mathcal{O}(n)$  may be informative about the objects of interest, such as impulse responses (Baumeister and Hamilton 2015).

Rather than specifying a single prior, the robust Bayesian approach of GK18 considers the class of all priors for  $\mathbf{Q}|\phi$  that are consistent with the identifying restrictions:

$$\Pi_{\mathbf{Q}|\phi} = \{ \pi_{\mathbf{Q}|\phi} : \pi_{\mathbf{Q}|\phi}(\mathcal{Q}(\phi|F, S)) = 1 \}. \quad (18)$$

Combining the class of priors with the posterior for  $\phi$  generates a class of posteriors for  $\theta$ :

$$\Pi_{\theta|\mathbf{Y}, \mathbf{M}} = \{ \pi_{\theta|\mathbf{Y}, \mathbf{M}} = \pi_{\mathbf{Q}|\phi} \pi_{\phi|\mathbf{Y}, \mathbf{M}} : \pi_{\mathbf{Q}|\phi} \in \Pi_{\mathbf{Q}|\phi} \}, \quad (19)$$

where  $\mathbf{Y} = (\mathbf{y}'_{1-p}, \dots, \mathbf{y}'_T)'$  and  $\mathbf{M} = (\mathbf{m}'_{1-p}, \dots, \mathbf{m}'_T)'$ . In turn, the class of posteriors for  $\theta$  induces a class of posteriors for  $\eta_{i,j,h}$ . GK18 suggest summarising this class of posteriors by reporting the ‘set of posterior means’:

$$\left[ \int_{\Phi} l(\phi) d\pi_{\phi|\mathbf{Y}, \mathbf{M}}, \int_{\Phi} u(\phi) d\pi_{\phi|\mathbf{Y}, \mathbf{M}} \right], \quad (20)$$

where  $l(\phi) = \inf\{\eta_{i,j,h}(\phi, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\phi|F, S)\}$  and  $u(\phi) = \sup\{\eta_{i,j,h}(\phi, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\phi|F, S)\}$ . They also suggest reporting a robust credible region with credibility level  $\alpha$  (see Proposition 1 of GK18). This region is interpreted as the shortest interval estimate for  $\eta_{i,j,h}$  such that the posterior probability put on the interval is greater than or equal to  $\alpha$  uniformly over the posteriors in the class. One can also report posterior probability bounds, which are the lowest and highest posterior probabilities of an event over all priors in the class.

When there are zero restrictions only, the identified set is never empty and so the data are not informative about the plausibility of the identifying restrictions. When there are sign restrictions, the identified set may be empty at particular values of  $\phi$ . The posterior probability that the identified set is non-empty,  $\pi_{\phi|\mathbf{Y}, \mathbf{M}}(\{\phi : IS_{\eta_{i,j,h}}(\phi|F, S) \neq \emptyset\})$ , can thus be used to quantify the plausibility of the identifying restrictions.

### 3 Frequentist Validity

In this section we provide conditions under which the robust Bayesian inferential approach provides valid frequentist inference about impulse responses in the proxy SVAR. This may be of interest to frequentists who use Bayesian approaches to inference purely for computational convenience. Bayesians may also be interested in the asymptotic frequentist properties of Bayesian procedures if this facilitates the communication of their results to frequentist



audiences.

The set of posterior means can be interpreted as a consistent estimator of the true identified set if  $IS_{\eta_{i,j,h}}(\phi|F, S)$  is convex and is a continuous correspondence of  $\phi$  at the true value  $\phi_0$  (see Theorem 3 in GK18). If, in addition,  $[l(\phi), u(\phi)]$  is differentiable in  $\phi$  at  $\phi_0$  with nonzero derivatives, and the posterior for  $\phi$  satisfies the Bernstein-von Mises property, the robust credible region is an asymptotically valid confidence set for the true identified set (see Proposition 2 in GK18). In the context of an SVAR, Propositions 3 and 4 of GK18 provide conditions under which the impulse-response identified set is guaranteed to be convex and continuous in  $\phi$ , respectively, while Proposition 5 provides conditions under which it is guaranteed to be differentiable in  $\phi$ . In the proxy SVAR, we can show that having the relevance condition satisfied at  $\phi_0$  is a necessary condition for continuity of the identified-set correspondence at  $\phi_0$ . In Section 3.1 we proceed under the assumption that the relevance condition is satisfied and provide conditions under which the identified-set correspondence is convex and is differentiable in  $\phi$ . We explore issues associated with ‘weak’ proxies – where the relevance condition is ‘close’ to being violated – in Section 3.2.

### 3.1 Strong Proxies

Assume that there are  $k$  proxies correlated with the last  $k$  structural shocks and uncorrelated with the remaining  $n - k$  structural shocks. Assume also that  $n > 3$  and  $1 < k < n - 1$ , so there are multiple proxies for multiple shocks and the impulse responses to all shocks are set-identified. This is the setting in MR13 and Mertens and Montiel-Olea (2018) (before the imposition of additional point-identifying zero restrictions), and so is of empirical relevance. The propositions below clarify conditions for  $IS_{\eta_{i,j,h}}(\phi|F, S)$  to be convex and differentiable in  $\phi$ , in which case the robust Bayesian approach provides asymptotically valid frequentist inference about the impulse-response identified set. We relegate proofs to Appendix 7.

**Proposition 3.1.** *Let the object of interest be  $\eta_{i,j^*,h} = \mathbf{c}_{i,h}(\phi)\mathbf{q}_{j^*}$ , the impulse response of the  $i$ th variable at the  $h$ th horizon to the  $j^*$ th structural shock, where the variables are ordered according to Definition 1.*

(I) *Suppose there are only zero restrictions arising from the exogeneity assumption and that the relevance condition holds, so  $\text{rank}(\mathbf{D}\Sigma_{tr}) = k$ . Then, for every  $i$  and  $h$  and almost every  $\phi \in \Phi$ , the identified set of  $\eta_{i,j^*,h}$ ,  $IS_{\eta_{i,j^*,h}}(\phi|F, S)$ , is convex.*

(II) *Consider the case with both zero and sign restrictions. Suppose the only zero restrictions are those arising from the exogeneity restrictions and that the relevance condition holds, so  $\text{rank}(\mathbf{D}\Sigma_{tr}) = k$ . Also assume that any sign restrictions constrain the  $j^*$ th column of  $\mathbf{Q}$  only and let  $\mathbf{S}_{j^*}(\phi)\mathbf{q}_{j^*} \geq \mathbf{0}_{s \times 1}$  represent the sign restrictions.*

(i) If interest is in the impulse responses to one of the first  $n - k$  structural shocks, then  $j^* = 1$  by Definition 1, and  $IS_{\eta_{i,j^*,h}}(\phi|F, S)$  is convex for every  $i$  and  $h$  if there exists a unit-length vector  $\mathbf{q} \in \mathbb{R}^n$  satisfying

$$\mathbf{D}\Sigma_{tr}\mathbf{q} = \mathbf{0}_{k \times 1} \quad \text{and} \quad \mathbf{S}_1(\phi)\mathbf{q} > \mathbf{0}_{s \times 1}. \quad (21)$$

(ii) Let  $\mathbf{N}(\mathbf{D}\Sigma_{tr})$  be an orthonormal basis for the nullspace of  $\mathbf{D}\Sigma_{tr}$  (so  $\mathbf{N}(\mathbf{D}\Sigma_{tr})$  is an  $n \times (n - k)$  matrix). If interest is in the impulse responses to one of the last  $k$  structural shocks, then  $j^* = n - k + 1$  by Definition 1, and  $IS_{\eta_{i,j^*,h}}(\phi|F, S)$  is nonempty and convex for every  $i$  and  $h$  if there exists a unit-length vector  $\mathbf{q} \in \mathbb{R}^n$  satisfying

$$\mathbf{N}(\mathbf{D}\Sigma_{tr})'\mathbf{q} = \mathbf{0}_{k \times 1} \quad \text{and} \quad \mathbf{S}_{n-k+1}(\phi)\mathbf{q} > \mathbf{0}_{s \times 1}. \quad (22)$$

Proposition 3.1 states that, when there are exogeneity restrictions only, the identified set for the impulse response is convex for almost every  $\phi \in \Phi$ . When there are also sign restrictions constraining  $\mathbf{q}_{j^*}$  only, the identified set is convex conditional on it being nonempty. Note that convexity of the identified set in the empirically relevant case, when interest is in the responses to one of the last  $k$  shocks, does not follow from Proposition 3 of GK18. The key difference from the general setting of GK18 is that, in our case,  $\mathbf{F}_i(\phi) = \mathbf{D}\Sigma_{tr}$  has full row rank and is common for  $i = 1, \dots, j^* - 1$ . These special features of the matrix of zero restrictions makes it possible to characterise the set of feasible values for  $\mathbf{q}_{j^*}$ .

For the same cases in which we can guarantee convexity of the impulse-response identified set, we provide sufficient conditions for the differentiability of  $u(\phi)$  and  $l(\phi)$ . To do this, we follow GK18 by building on results from Gafarov *et al.* (2018), who show the directional differentiability of the upper and lower bound of the impulse-response identified set when there are zero and sign restrictions on  $\mathbf{q}_{j^*}$  only.

**Proposition 3.2.** *Let the object of interest be  $\eta_{i,j^*,h} = \mathbf{c}_{i,h}(\phi)\mathbf{q}_{j^*}$ , where the variables are ordered according to Definition 1. Suppose the only zero restrictions are those arising from the exogeneity assumption and that the relevance condition holds, so  $\text{rank}(\mathbf{D}\Sigma_{tr}) = k$ . Also assume that any sign restrictions constrain the  $j^*$ th column of  $\mathbf{Q}$  only and let  $\mathbf{S}_{j^*}(\phi)\mathbf{q}_{j^*} \geq \mathbf{0}_{s \times 1}$  represent the sign restrictions.*

(i) *Suppose the impulse responses of interest are those to one of the first  $n - k$  structural shocks, so  $j^* = 1$  by Definition 1, and that the column vectors of  $\left[ (\mathbf{D}\Sigma_{tr})', \mathbf{S}_1(\phi)', \Sigma_{tr}^{-1}\mathbf{e}_{1,n} \right]$  are linearly independent at  $\phi = \phi_0$ . If, at  $\phi = \phi_0$ , the set of solutions of the optimization*

problem

$$\max_{\mathbf{q} \in \mathcal{S}^{n-1}} \left( \min_{\mathbf{q} \in \mathcal{S}^{n-1}} \right) \mathbf{c}'_{i,h}(\boldsymbol{\phi}) \mathbf{q} \quad s.t. \quad \mathbf{D}\boldsymbol{\Sigma}_{tr} \mathbf{q} = \mathbf{0}_{k \times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1) \times 1} \quad (23)$$

is singleton, the optimized value  $u(\boldsymbol{\phi})$  ( $l(\boldsymbol{\phi})$ ) is nonzero, and the number of binding sign restrictions at the optimum is less than  $n - k - 1$ , then  $u(\boldsymbol{\phi})$  ( $l(\boldsymbol{\phi})$ ) is differentiable at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ .

(ii) Suppose the impulse responses of interest are those to one of the last  $k$  structural shocks, so  $j^* = n - k + 1$  by Definition 1, and that the column vectors of  $\left[ \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr}), \mathbf{S}_1(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{n-k+1,n} \right]$  are linearly independent at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ . If, at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ , the set of solutions of the optimization problem

$$\max_{\mathbf{q} \in \mathcal{S}^{n-1}} \left( \min_{\mathbf{q} \in \mathcal{S}^{n-1}} \right) \mathbf{c}'_{i,h}(\boldsymbol{\phi}) \mathbf{q} \quad s.t. \quad \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})' \mathbf{q} = \mathbf{0}_{k \times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_{n-k+1}(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{n-k+1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1) \times 1} \quad (24)$$

is singleton, the optimized value  $u(\boldsymbol{\phi})$  ( $l(\boldsymbol{\phi})$ ) is nonzero, and the number of binding sign restrictions at the optimum is less than  $k - 1$ , then  $u(\boldsymbol{\phi})$  ( $l(\boldsymbol{\phi})$ ) is differentiable at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ .

When  $n \geq 3$ , it is also straightforward to show that the identified set is convex when  $k = 1$  and interest is in the impulse responses to one of the first  $n - 1$  structural shocks (the impulse responses to the last structural shock are point-identified), or when  $k = n - 1$  and interest is in the impulse responses to one of the last  $n - 1$  shocks (the impulse responses to the first structural shock are point-identified). Differentiability in these cases is also obtained under similar conditions to those in Proposition 3.2.

When there are sign restrictions that constrain multiple columns of  $\mathbf{Q}$ , we cannot guarantee convexity of the identified set (see Example 4 in GK18), nor differentiability. Nevertheless, the set of posterior means and robust credible region can be interpreted as providing inference about the *convex hull* of the identified set.

Since  $\text{FEVD}_{i,j^*,h}$  is a continuous function of  $\mathbf{q}_{j^*}$ ,  $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F, S)$  is continuous at  $\boldsymbol{\phi}_0$  and it is convex whenever  $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F, S)$  is convex. If  $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F, S)$  is also differentiable in  $\boldsymbol{\phi}$ , we can guarantee frequentist validity of the robust Bayesian inferential procedure when applied to the FEVD in the same cases as for the impulse response. However, we are unaware of results on the differentiability of  $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F, S)$ . We leave exploration of this to further work.

### 3.2 Weak Proxies

In this section we investigate how weak proxies affect robust Bayesian posterior inference about set-identified impulse responses in the proxy SVAR. Our focus is on the asymptotic frequentist properties of our procedure. We consider the case where  $n = 3$ ,  $k = 1$  and the objects of interest are the impulse responses to  $\varepsilon_{1t}$ . We choose this case because it is straightforward to analytically characterise the identified set and hence to discuss the effects of weak proxies.

At a given value of  $\phi \in \Phi$  (and ignoring the sign normalization), the upper bound of  $IS_{\eta_{i,1,h}}(\phi|F, S)$  is the value function associated with the following optimization problem:<sup>12</sup>

$$u(\phi) = \max_{\mathbf{q} \in \mathcal{S}^{n-1}} \mathbf{c}'\mathbf{q} \quad \text{subject to} \quad \mathbf{d}'\mathbf{q} = 0,$$

where  $\mathbf{c} \equiv \mathbf{c}_{i,h}(\phi)$  and  $\mathbf{d}' \equiv \mathbf{D}\Sigma_{tr}$ . Applying the change of variables  $\mathbf{x} = \Sigma_{tr}\mathbf{q}$  yields the problem in Equation (2.5) of Gafarov *et al.* (2018). Using their results, the value function satisfies

$$u(\phi)^2 = \mathbf{c}' \left[ \mathbf{I}_3 - \mathbf{d}(\mathbf{d}'\mathbf{d})^{-1}\mathbf{d}' \right] \mathbf{c}. \quad (25)$$

Equation (12) implies that  $\mathbb{E}(m_t \varepsilon_{3t}) = \mathbf{d}'\mathbf{q}_3 = \Psi$ . The exogeneity restrictions require that  $\mathbf{q}_1$  and  $\mathbf{q}_2$  are orthogonal to  $\mathbf{d}$ . Since the columns of  $\mathbf{Q}$  are orthogonal and have unit length,  $\mathbf{q}_3 = \pm\mathbf{d}/\|\mathbf{d}\|$ , which implies that  $\mathbf{d}'\mathbf{d}/\|\mathbf{d}\| = \|\mathbf{d}\| = |\Psi|$ .

A ‘weak’ proxy correlates with one of the structural shocks only weakly, so  $|\Psi|$  is close to zero. This is equivalent to  $\|\mathbf{d}\|$  being small. Note that  $u(\phi)$  as the square root of (25) is continuous and smooth in  $\mathbf{c}$ , while it is discontinuous in  $\mathbf{d}$  at  $\mathbf{d} = \mathbf{0}_{3 \times 1}$ . Hence, if the posterior distribution of  $\mathbf{d}$  concentrates near a point of singularity of  $u(\phi)$  due to the weak proxy, the posterior of  $u(\phi)$  can exhibit a nonstandard distribution even when the posterior of  $(\mathbf{c}, \mathbf{d})$  is consistent and can be well approximated by a normal distribution centered at the maximum likelihood estimator (MLE).

To investigate the posterior for  $u(\phi)$  in the weak-proxy case, we consider the local asymptotic approximation of the posterior for  $u(\phi)$  with a drifting sequence of the true values of  $\phi$  converging to a point of singularity. We here present the heuristic exposition of the results and defer the regularity conditions and formal proofs to the Appendix.

We consider a drifting sequence of data-generating processes  $\{\phi_T : T = 1, 2, \dots\}$  that induces a drifting sequence of parameter values  $\{(\mathbf{c}_T, \mathbf{d}_T) : T = 1, 2, \dots\}$  converging to a point of singularity. Following the weak-instrument asymptotics of Staiger and Stock (1997),

---

<sup>12</sup>In the absence of sign normalization restrictions, the lower bound of the identified set  $l(\phi)$  is given by  $-u(\phi)$ . This section hence focuses only on the posterior for  $u(\phi)$ .

we consider the drifting sequence of  $(\mathbf{c}, \mathbf{d})$  with  $T^{-1/2}$ -convergence rate,

$$\mathbf{c}_T = \mathbf{c}_0 + \frac{\boldsymbol{\gamma}}{\sqrt{T}}, \quad \mathbf{d}_T = \frac{\boldsymbol{\delta}}{\sqrt{T}}, \quad (26)$$

where  $\mathbf{c}_0 \neq \mathbf{0}_{3 \times 1}$ ,  $\boldsymbol{\delta} \neq \mathbf{0}_{3 \times 1}$  and  $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^3 \times \mathbb{R}^3$  are the localisation parameters. The magnitude of  $\boldsymbol{\delta}$  characterises the relevance of the proxy; that is, a smaller value of  $\|\boldsymbol{\delta}\|$  implies a weaker proxy.

Let  $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$  be the MLE for  $(\mathbf{c}, \mathbf{d})$  (which is a constant once we have conditioned on the sample). We assume that the sampling distribution of the MLE is  $\sqrt{T}$ -asymptotically normal:

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_T - \mathbf{d}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}_{6 \times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right). \quad (27)$$

We also assume that the posterior for  $(\mathbf{c}, \mathbf{d})$  converges to a normal distribution with data-independent variance. That is, conditional on the sampling sequence,

$$\sqrt{T} \begin{pmatrix} \mathbf{c} - \hat{\mathbf{c}}_T \\ \mathbf{d} - \hat{\mathbf{d}}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_c \\ \mathbf{Z}_d \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}_{6 \times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right), \quad (28)$$

as  $T \rightarrow \infty$  for almost every sampling sequence, where  $\boldsymbol{\Omega} \equiv \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix}$  is the posterior asymptotic variance, which does not depend on the sampling sequence. The asymptotic equivalence of the probability laws in (27) and (28) implies that the reduced-form parameters  $(\mathbf{c}, \mathbf{d})$  are regular in the sense that the well-known Bernstein-von Mises Theorem holds. See, for instance, Schervish (1995) and DasGupta (2008) for a set of sufficient conditions for posterior asymptotic normality with the Bernstein-von Mises property.

Under this setting, Proposition 8.1 in the Appendix derives the following asymptotic approximation of the posterior for  $u(\boldsymbol{\phi})$ . Conditional on the sampling sequence,

$$u(\boldsymbol{\phi}) \xrightarrow{d} \sqrt{\mathbf{c}'_0 \left( \mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d\|^2} \right) \mathbf{c}_0}, \quad (29)$$

as  $T \rightarrow \infty$  for almost every sampling sequence, where  $\hat{\mathbf{Z}}_d$  is a constant that depends on the sample, and  $\mathbf{Z}_d \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Omega}_d)$ .

This representation of the asymptotic posterior provides the following insights about the influence of the weak proxy on posterior inference. First, the posterior of  $u(\boldsymbol{\phi})$  is not consistent and remains a non-degenerate distribution in large samples. Second, the asymptotic posterior for  $u(\boldsymbol{\phi})$  depends not only on the localisation parameter  $\boldsymbol{\delta}$ , but also on the statistic

$\hat{\mathbf{Z}}_d$  realized in the data. Hence, unlike in the well-identified case, the influence of the data on the shape of the posterior does not disappear in large samples. Also, the asymptotic posterior mean almost always (in terms of the sampling probability) misses the upper bound of the true identified set defined by the limit along the drifting data-generating processes  $\{\phi_T : T = 1, 2, \dots\}$  yielding (26):

$$\lim_{T \rightarrow \infty} u(\phi_T) = \sqrt{\mathbf{c}'_0 \left( \mathbf{I}_3 - \frac{\boldsymbol{\delta} \boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2} \right) \mathbf{c}_0}.$$

This implies that, under the current weak-proxy asymptotics, the set of posterior means for the impulse response is not a consistent estimator for the identified set.

Under the same drifting sequence inducing (26), Proposition 8.2 in the Appendix derives the asymptotic sampling distribution of the MLE for the upper bound of the identified set:

$$u(\hat{\phi}) \xrightarrow{d} \sqrt{\mathbf{c}'_0 \left( \mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0}, \quad (30)$$

where  $\hat{\mathbf{Z}}_d \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Omega}_d)$ . Like the posterior for  $u(\phi)$ , the sampling distribution of  $u(\hat{\phi})$  is not consistent and remains non-degenerate. A comparison of (29) and (30) shows that the posterior and the sampling distribution of the MLE for the bound of the identified set do not asymptotically coincide for almost every sampling sequence. The Bernstein-von Mises property therefore fails for the estimation of the upper and lower bounds of the identified set. This implies that the asymptotic frequentist coverage of the robust credible region of GK18 can also fail, because a condition analogous to Assumption 4 in GK 18 does not hold in the current setting of weak proxy asymptotics.

When the proxy is strong in the sense that  $|\Psi| = \|\mathbf{d}\|$  is far from zero, the point-wise asymptotic approximation of the posterior of  $u(\phi)$  approximates well the finite-sample posterior. Noting that  $u(\phi)$  is smooth at  $\mathbf{d} \neq \mathbf{0}_{3 \times 1}$  and assuming that the posterior of  $(\mathbf{c}, \mathbf{d})$  centered at the MLE is  $\sqrt{T}$ -asymptotically normal, the delta method implies that  $\sqrt{T}(u(\phi) - u(\hat{\phi}))$  is asymptotically normal with a data-independent variance. This asymptotic posterior coincides with the sampling distribution of the MLE, so correct frequentist coverage of the robust credible region can be attained in addition to posterior consistency. This stark contrast in the asymptotic behavior of the posteriors suggests that, in the current simple setting, whether the posterior of  $u(\phi)$  is non-normal could be useful for diagnosing whether the proxy is weak. We leave a formal analysis of this for future research.

## 4 Numerical Implementation

In this section, we provide numerical algorithms to conduct robust Bayesian inference about set-identified objects of interest in proxy SVARs. The algorithms numerically approximate the set of posterior means and associated robust credible interval. When there are sign restrictions, the algorithms also give estimates of the plausibility of the identifying restrictions. Throughout, we assume that the order of the variables satisfies Definition 1. Since the identifying restrictions are linear restrictions on columns of  $\mathbf{Q}$ , the algorithms are similar to the algorithms in GK18. We repeat them here for completeness and discuss further details, and issues specific to the proxy SVAR case, below. Algorithm 1 assumes that the object of interest is the impulse response; the subsequent remarks discuss how to conduct robust Bayesian inference about other objects of interest. Matlab code implementing the algorithms is available on the authors' personal websites.

**Algorithm 1.** *Let  $\mathbf{F}(\phi, \mathbf{Q}) = \mathbf{0}_{(\sum_{i=1}^n f_i) \times 1}$  and  $S(\phi, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}$  be the set of identifying restrictions and let  $\eta_{i,j^*,h} = \mathbf{c}'_{i,h} \mathbf{q}_{j^*}$  be the impulse response of interest.*

- **Step 1:** *Specify a prior for  $\phi$ ,  $\pi_\phi$ , and obtain the posterior  $\pi_{\phi|\mathbf{Y},\mathbf{M}}$ .<sup>13</sup>*
- **Step 2:** *Draw  $\phi$  from  $\pi_{\phi|\mathbf{Y},\mathbf{M}}$  and check whether  $\mathcal{Q}(\phi|F, S)$  is empty using the subroutine below.*
  - **Step 2.1:** *Draw  $\mathbf{z}_1 \sim N(\mathbf{0}_{n \times 1}, \mathbf{I}_n)$  and let  $\tilde{\mathbf{q}}_1 = [\mathbf{I}_n - \mathbf{F}'_1(\mathbf{F}_1\mathbf{F}'_1)^{-1}\mathbf{F}_1] \mathbf{z}_1$ . For  $i = 2, \dots, n$ , run the following procedure sequentially: draw  $\mathbf{z}_i \sim N(\mathbf{0}_{n \times 1}, \mathbf{I}_n)$  and compute  $\tilde{\mathbf{q}}_i = [\mathbf{I}_n - \tilde{\mathbf{F}}'_i(\tilde{\mathbf{F}}_i\tilde{\mathbf{F}}'_i)^{-1}\tilde{\mathbf{F}}_i] \mathbf{z}_i$ , where  $\tilde{\mathbf{F}}'_i = [\mathbf{F}'_i, \tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_{i-1}]$ .*
  - **Step 2.2:** *Given  $\tilde{\mathbf{q}}_i$ ,  $i = 1, \dots, n$ , define*

$$\mathbf{Q}_0 = \left[ \text{sign}((\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{1,n})' \tilde{\mathbf{q}}_1) \frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|}, \dots, \text{sign}((\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{n,n})' \tilde{\mathbf{q}}_n) \frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|} \right].^{14}$$

- **Step 2.3:** *Check whether  $\mathbf{Q}_0$  satisfies  $\mathbf{S}(\phi, \mathbf{Q}_0) \geq \mathbf{0}_{s \times 1}$ . If so, retain  $\mathbf{Q}_0$  and proceed to Step 3. Otherwise, repeat Steps 2.1 and 2.2 (up to a maximum of  $L$  times) until  $\mathbf{Q}_0$  is obtained satisfying  $\mathbf{S}(\phi, \mathbf{Q}_0) \geq \mathbf{0}_{s \times 1}$ . If no draws of  $\mathbf{Q}_0$  satisfy  $\mathbf{S}(\phi, \mathbf{Q}_0) \geq \mathbf{0}_{s \times 1}$ , approximate  $\mathcal{Q}(\phi|F, S)$  as being empty and return to Step 2.*

<sup>13</sup> $\pi_\phi$  does not have to be proper or to satisfy the condition  $\pi_\phi(\{\phi : \mathcal{Q}(\phi|F, S) \neq \emptyset\}) = 1$  for all  $\phi \in \Phi$ ; that is,  $\pi_\phi$  may assign positive probability to regions of  $\Phi$  that yield an empty set of orthonormal matrices satisfying the identifying restrictions.

<sup>14</sup>If  $(\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{i,n})' \tilde{\mathbf{q}}_i = 0$  for some  $i$ , set  $\text{sign}((\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{i,n})' \tilde{\mathbf{q}}_i)$  equal to 1 or  $-1$  with equal probability.

- **Step 3:** Compute the lower bound of  $IS_{\eta_{i,j^*,h}}(\phi|F, S)$  by solving the following constrained optimization problem with initial value  $\mathbf{Q}_0$ :

$$l(\phi) = \min_{\mathbf{Q}} \mathbf{c}'_{i,h} \mathbf{q}_{j^*}$$

subject to

$$\begin{aligned} \mathbf{F}(\phi, \mathbf{Q}) &= \mathbf{0}_{(\sum_i^n f_i) \times 1} \\ \mathbf{S}(\phi, \mathbf{Q}) &\geq \mathbf{0}_{s \times 1} \\ \text{diag}(\mathbf{Q}' \boldsymbol{\Sigma}_{tr}^{-1}) &\geq \mathbf{0}_{n \times 1} \\ \mathbf{Q}' \mathbf{Q} &= \mathbf{I}_n. \end{aligned}$$

Similarly, obtain  $u(\phi) = \max_{\mathbf{Q}} \mathbf{c}'_{i,h} \mathbf{q}_{j^*}$  under the same set of constraints.

- **Step 4:** Repeat Steps 2–3  $M$  times to obtain  $[l(\phi_m), u(\phi_m)]$  for  $m = 1, \dots, M$ . Approximate the set of posterior means by the sample averages of  $l(\phi_m)$  and  $u(\phi_m)$ .
- **Step 5:** To obtain an approximation of the smallest robust credible region with credibility  $\alpha \in (0, 1)$ , define  $d(\eta, \phi) = \max\{|\eta - l(\phi)|, |\eta - u(\phi)|\}$  and let  $\hat{z}_\alpha(\eta)$  be the sample  $\alpha$ -th quantile of  $\{d(\eta, \phi_m), m = 1, \dots, M\}$ . An approximated smallest robust credible interval for  $\eta_{i,j^*,h}$  is an interval centered at  $\arg \min_\eta \hat{z}_\alpha(\eta)$  with radius  $\min_\eta \hat{z}_\alpha(\eta)$ .
- **Step 6:** Approximate  $\pi_{\phi|\mathbf{Y}, \mathbf{M}}(\{\phi : \mathcal{Q}(\phi|F, S) \neq \emptyset\})$  by the proportion of draws of  $\phi$  passing Step 2.3.

## 4.1 Remarks

### 4.1.1 Further details about Step 2

Given a draw of  $\phi$  from its posterior, Step 2 attempts to draw  $\mathbf{Q}$  satisfying the zero and sign restrictions. The vectors  $\tilde{\mathbf{q}}_i$ ,  $i = 1, \dots, n$ , are residual vectors from the linear projection of multivariate standard normally distributed random variables on vectors representing the zero restrictions and previously constructed columns of  $\mathbf{Q}$ . These residual vectors therefore satisfy the zero restrictions represented in  $\mathbf{F}(\phi, \mathbf{Q})$  and are orthogonal.<sup>15</sup> Step 2.2 rescales the residual vectors to have unit length and imposes the sign normalization that the diagonal elements of  $\mathbf{A}_0$  are nonnegative.

<sup>15</sup>If the relevance condition fails,  $\mathbf{F}_i(\phi)$  is of reduced row rank for  $i = 1, \dots, n - k$  and the coefficients in the linear projection are not identified. This is a measure zero event so long as  $\pi_\phi$  does not place positive probability mass on the event  $\text{rank}(\mathbf{D}) < k$ .



Step 2 can be interpreted as implementing a particular implicit prior over  $\mathcal{Q}(\phi|F, S)$ . This implicit prior is irrelevant for the class of posteriors generated by the robust Bayesian procedure, since the draw of  $\mathbf{Q}$  is used only as an initial value in the numerical optimization step. However, one could use these draws to construct the posterior for  $\eta_{i,j,h}$  induced by this prior; in the empirical application below, we do this to illustrate how posterior inference may be sensitive to the choice of prior for  $\mathbf{Q}|\phi$ . The algorithm used to draw  $\mathbf{Q}$  possesses similarities to the algorithms described in Arias, Rubio-Ramírez and Waggoner (2018), who conduct single-prior Bayesian inference in set-identified SVARs. However, differences in the algorithms mean that the implicit priors are different. Algorithm 3 in Arias *et al.* (2018) draws a value of  $\mathbf{Q}$  satisfying the zero restrictions only once at each draw of  $\phi$  and discards the joint draw of  $\phi$  and  $\mathbf{Q}$  if the sign restrictions are not satisfied. In contrast, we draw  $\mathbf{Q}$  satisfying the zero restrictions until we obtain a value satisfying the sign restrictions. Relative to the prior implicit in the second approach, the prior implicit in the first approach places more weight on values of  $\phi$  with a larger identified set for  $\mathbf{Q}$  (see Uhlig (2017) for a discussion of this point).

#### 4.1.2 Choice of prior

The method used to draw  $\phi|\mathbf{Y}, \mathbf{M}$  depends on the posterior, and thus on the prior. In the empirical application below we use independent (improper) Jeffreys' priors over the blocks of reduced-form parameters in the VAR for  $\mathbf{y}_t$  and the first-stage regression; that is,  $\pi_\phi = \pi_{\mathbf{B}, \Sigma} \pi_{\mathbf{D}, \mathbf{J}, \mathbf{r}}$ , where  $\pi_{\mathbf{B}, \Sigma} \propto |\Sigma|^{-\frac{n+1}{2}}$  and  $\pi_{\mathbf{D}, \mathbf{J}, \mathbf{r}} \propto |\mathbf{Y}|^{-\frac{k+1}{2}}$ .<sup>16</sup> This makes it simple to draw from the posterior of  $\phi|\mathbf{Y}, \mathbf{M}$ , since it is the product of independent normal-inverse-Wishart posteriors.<sup>17</sup> We emphasise that our algorithm does not rely on using independent priors over the reduced-form parameters; all that matters is that one can sample from the posterior of  $\phi$ . For example, if the prior is over the model's structural – rather than reduced-form – parameters, one could draw from the posterior of the structural parameters and transform these draws into draws of the reduced-form parameters.

#### 4.1.3 Convergence issues and alternative algorithms

The optimization problem in Step 3 is nonconvex. Consequently, the convergence of gradient-based optimization methods in this problem is not guaranteed. Accordingly, we suggest

<sup>16</sup> $\pi_{\mathbf{B}, \Sigma}$  is nonzero only for values of  $\mathbf{B}$  such that the VAR is invertible into a VMA( $\infty$ ).

<sup>17</sup>This follows from the fact that the joint likelihood of  $(\mathbf{M}, \mathbf{Y})$  is multiplicatively separable across the two blocks of parameters:  $\pi_{\mathbf{Y}, \mathbf{M}|\phi} = \pi_{\mathbf{M}|\mathbf{Y}, \mathbf{D}, \mathbf{J}, \mathbf{r}} \pi_{\mathbf{Y}|\mathbf{B}, \Sigma}$ . For an algorithm that draws from the normal-inverse-Wishart posterior distribution, see Del Negro and Schorfheide (2011). Imposing independent normal-inverse-Wishart priors would also yield a posterior that is the product of independent normal-inverse-Wishart posteriors.

drawing multiple values of  $\mathbf{Q}_0$  in Steps 2.1–2.3 to use as initial values in the optimization step, and computing optima over the set of solutions obtained from the different initial values. GK18 also provide an algorithm that can be used to check for convergence of, or as an alternative to, the numerical optimization step.

**Algorithm 2.** *In Algorithm 1, replace Step 3 with the following:*

- **Step 3’:** *Iterate Steps 2.1–2.3  $K$  times and let  $\{\mathbf{Q}_l, l = 1, \dots, \tilde{K}\}$  be the  $\tilde{K}$  draws of  $\mathbf{Q}$  that satisfy the identifying restrictions. Let  $\mathbf{q}_{j^*,l}$  be the  $j^*$ th column of  $\mathbf{Q}_l$ . Approximate  $[l(\phi), u(\phi)]$  by  $[\min_l \mathbf{c}'_{i,h} \mathbf{q}_{j^*,l}, \max_l \mathbf{c}'_{i,h} \mathbf{q}_{j^*,l}]$ .*

Algorithm 2 yields an approximated identified set that is smaller than the true identified set at every draw of  $\phi$ . However, the approximated identified set will converge to the true identified set as  $\tilde{K}$  goes to infinity. In our implementation of this algorithm, we fix  $\tilde{K}$  and let  $K$  vary. We suggest determining an appropriate value of  $\tilde{K}$  by fixing the value of  $\phi$  (e.g., at the MLE) and comparing the bounds obtained given different values of  $\tilde{K}$ . In some cases, Algorithm 2 may be computationally less demanding than Algorithm 1. For example, when the dimension of the VAR is large or if interest is in the impulse responses of many variables at many horizons, the computational cost of generating a sufficiently large number of draws of  $\mathbf{Q}$  to accurately approximate the bounds of the identified sets may be smaller than the cost of carrying out the optimization step for each variable of interest at each horizon (particularly when using multiple initial values). Conversely, Algorithm 2 may be computationally more demanding when there are sign restrictions that substantially truncate the support of  $\mathbf{Q}$ , because many draws of  $\mathbf{Q}$  will be rejected. In practice, Step 3 and Step 3’ are parallelizable, so large reductions in computing time are possible in both algorithms by distributing computation across multiple processors.

Under constraints on  $\mathbf{q}_{j^*}$  only, Gafarov *et al.* (2018) develop an algorithm to compute the bounds of the identified set using an analytical expression for the bounds given a set of active zero and/or sign restrictions. This approach will typically be computationally more efficient than approximating the bounds via gradient-based numerical optimization or simulation. However, it is not generally applicable in proxy SVARs that are likely to be of interest empirically, because the types of sign restrictions on  $\Psi$  that naturally arise in this

setting will usually constrain multiple columns of  $\mathbf{Q}$ .<sup>18</sup>

#### 4.1.4 Point identification

If  $f_{j^*} = j^* - 1$ , the equality restrictions on  $\mathbf{q}_{j^*}$  are sufficient to point identify the object of interest. This means that the prior for  $\phi$  induces a single posterior for the object of interest. In this case, Steps 1 and 2.1–2.2 of Algorithm 1 can still be used to draw from this posterior. Because  $\mathbf{q}_{j^*}$  is exactly identified, any draw of  $\mathbf{Q}$  satisfying the zero restrictions will contain the same  $\mathbf{q}_{j^*}$  and thus will yield the same object of interest. We make use of this in the empirical application below when estimating a proxy SVAR under point-identifying restrictions.

#### 4.1.5 Other objects of interest

When interest is in the FEVD rather than the impulse response, Algorithms 1 and 2 can be modified by replacing  $\mathbf{c}_{i,h}(\phi)' \mathbf{q}_{j^*}$  with  $\text{FEVD}_{i,j^*,h}$ . If one is interested in both impulse responses and FEVDs, Algorithm 2 may deliver large gains in computation time over Algorithm 1, because the same draws of  $\mathbf{Q}$  can be used to compute bounds for all objects of interest rather than having to carry out the numerical optimization step for each object separately. Note also that when interest is in the *cumulative* impulse response,  $\mathbf{c}_{i,h}(\phi)' \mathbf{q}_{j^*}$  is replaced with  $\left(\sum_{k=1}^h \mathbf{c}_{i,k}(\phi)'\right) \mathbf{q}_{j^*}$ .

#### 4.1.6 Impulse responses to a unit shock

The algorithms above impose the normalization  $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{I}_n$ , which is typical in set-identified SVARs (e.g., Uhlig 2005). This means that the impulse responses are to a standard-deviation shock. Algorithm 3 shows how to obtain the set of posterior means and the robust credible interval for impulse responses to a unit shock, which may be of more interest in particular applications (see Stock and Watson (2016, 2018) for a discussion of this point).

**Algorithm 3.** *In Algorithm 1, replace Step 3 with the following:*

---

<sup>18</sup>There are special cases where the results in Gafarov *et al.* (2018) could be extended to compute the bounds of the identified set. For example, if  $n > 3$ ,  $1 < k < n - 1$  and  $j^* = n - k + 1$ , one could compute an orthonormal basis for the nullspace of  $\mathbf{D}\boldsymbol{\Sigma}_{tr}$  and include the vectors representing this basis in the set of active restrictions. If  $k = n - 1$  and  $j^* = 2$ , one could include the restriction  $\mathbf{q}'_1 \mathbf{q}_2 = 0$  in the set of active restrictions, since in this case  $\mathbf{q}_1$  is point-identified. As in the example in Section 3.2, the analytical results could also be applied if there are exogeneity restrictions only and interest is in the (set-identified) impulse response to the first structural shock. These examples would all still require that any sign restrictions constrain only  $\mathbf{q}_{j^*}$ .

- **Step 3**: Iterate Steps 2.1–2.3  $K$  times and let  $\{\mathbf{Q}_l, l = 1, \dots, \tilde{K}\}$  be the  $\tilde{K}$  draws of  $\mathbf{Q}$  that satisfy the identifying restrictions. Let  $\mathbf{A}_{0,l}^{-1} = \Sigma_{tr} \mathbf{Q}_l$  and compute  $\mathbf{a}_{j^*,l} = (\mathbf{A}_{0,l}^{-1} \mathbf{e}_{j^*,n}) / (\mathbf{e}'_{j^*,n} \mathbf{A}_{0,l}^{-1} \mathbf{e}_{j^*,n})$ . Approximate  $[l(\phi), u(\phi)]$  by  $[\min_l \mathbf{e}'_{i,n} \mathbf{C}_h \mathbf{a}_{j^*,l}, \max_l \mathbf{e}'_{i,n} \mathbf{C}_h \mathbf{a}_{j^*,l}]$ .

The algorithm generates impulse responses to a standard-deviation shock that are consistent with the identifying restrictions, rescales the impulse responses so that they are with respect to a unit shock (the  $i$ th element of  $\mathbf{a}_{j^*,l}$  is equal to one), and computes the bounds of the identified set using the extreme values of the rescaled impulse responses. One potential issue is that the set of posterior means and robust credible interval may be unbounded when the relevant diagonal elements of  $\mathbf{A}_0^{-1} = \Sigma_{tr} \mathbf{Q}$  are not bounded away from zero for all  $\phi \in \Phi$  and  $\mathbf{Q} \in \mathcal{Q}(\phi|F, S)$ .

## 5 Empirical Application: The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States

We illustrate our methodology using the proxy SVAR considered in MR13, who estimate the macroeconomic effects of shocks to average personal and corporate income tax rates in the United States. The variables included in their benchmark specification are the average personal income tax rate (APITR), the average corporate income tax rate (ACITR), the personal income tax base, the corporate income tax base, government purchases of final goods, gross domestic product and federal government debt. The last five variables are in real per capita terms and are included in logs. MR13 decompose the sequence of plausibly exogenous changes in tax liabilities constructed by Romer and Romer (2010) into those related to personal income taxes and those related to corporate income taxes, and they exclude changes in tax liabilities with a lag between announcement and implementation of more than one quarter. These changes in tax liabilities are divided by the relevant tax base in the previous quarter and the resulting variables are used as proxies for structural shocks to the APITR and ACITR. The data are quarterly and run from 1950Q1 to 2006Q4. The VAR includes a constant and four lags of the endogenous variables. See MR13 for further details about the construction of the variables used in the VAR and the proxies.<sup>19</sup>

When the objects of interest are impulse responses to  $\varepsilon_{APITR,t}$ , any ordering of the variables such that  $\mathbf{y}_t = [\mathbf{x}_t, APITR_t, ACITR_t]$ , where  $\mathbf{x}_t$  contains all variables other than  $APITR_t$  and  $ACITR_t$ , will satisfy Definition 1. When interest is in the impulse responses

<sup>19</sup>We obtained the data from Karel Mertens' website: <https://karelmertens.com/research/>.

to  $\varepsilon_{ACITR,t}$ , any ordering such that  $\mathbf{y}_t = [\mathbf{x}_t, ACITR_t, APITR_t]$  will satisfy Definition 1. In both cases,  $f_i = 2$  for  $i = 1, \dots, 5$ ,  $f_6 = f_7 = 0$  and  $j^* = 6$ . Let  $\mathbf{m}_t = (m_{APITR,t}, m_{ACITR,t})'$ , where  $m_{APITR,t}$  and  $m_{ACITR,t}$  are the rescaled changes in personal and corporate income tax liabilities, respectively. MR13 impose the identifying restrictions that  $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:5),t}) = \mathbf{0}_{2 \times 5}$  and  $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{6:7}) = \boldsymbol{\Psi}$ , where  $\boldsymbol{\Psi}$  is an (unknown) full-rank  $2 \times 2$  matrix. These identifying restrictions are insufficient to point identify any structural shock. As discussed in MR13, if one were willing to assume that  $m_{APITR,t}$  is uncorrelated with  $\varepsilon_{ACITR,t}$ , or vice versa for  $m_{ACITR,t}$  and  $\varepsilon_{APITR,t}$ , the additional zero restriction would be sufficient to point identify both structural shocks of interest.<sup>20</sup> However, positive correlation between the proxies suggests that these assumptions may be inappropriate.

To achieve point identification, MR13 consider additional zero restrictions on the direct contemporaneous response of one tax rate to the other (i.e., causal orderings). For example, when interest is in the impulse responses to  $\varepsilon_{APITR,t}$ , they assume that the ACITR does not respond directly to a structural shock in the APITR on impact. In our setting, this restriction is  $\mathbf{e}'_{7,7} \mathbf{A}_0 \mathbf{e}_{6,7} = (\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{e}_{7,7})' \mathbf{q}_6 = 0$ . This restriction also point-identifies the ACITR shock. To assess robustness of their results, they consider the alternative causal ordering that the APITR does not respond directly to a structural shock in the ACITR on impact. Either of these zero restrictions could be violated if, for instance, there are constraints that impinge on the ability of the government to change personal and corporate income tax rates independently of one another. Accordingly, we extend their robustness analysis by providing an estimator of the set of impulse responses compatible with relaxing the additional zero restriction and replacing it with a set of – arguably weaker – sign restrictions.

We assume that each proxy is positively correlated with its associated structural shock (i.e.,  $\mathbb{E}(m_{APITR,t} \varepsilon_{APITR,t}) \geq 0$  and  $\mathbb{E}(m_{ACITR,t} \varepsilon_{ACITR,t}) \geq 0$ ) and that each proxy is more highly correlated with its associated structural shock than with the structural shock to the other average tax rate (i.e.,  $\mathbb{E}(m_{APITR,t} \varepsilon_{APITR,t}) \geq \mathbb{E}(m_{APITR,t} \varepsilon_{ACITR,t})$  and  $\mathbb{E}(m_{ACITR,t} \varepsilon_{ACITR,t}) \geq \mathbb{E}(m_{ACITR,t} \varepsilon_{APITR,t})$ ). We also assume that the response of each average tax rate to its own structural shock is nonnegative on impact, which is a sign restriction on impulse responses (as in Uhlig (2005)). Importantly, our approach allows us to relax the additional point-identifying zero restriction while avoiding the need to impose an unrevisable prior over the

<sup>20</sup>  $\mathbf{q}_i$ ,  $i = 1, \dots, 5$ , is restricted to the 5-dimensional subspace of  $\mathbb{R}^7$  in the nullspace of  $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ .  $\mathbf{q}_6$  and  $\mathbf{q}_7$  therefore lie in the 2-dimensional subspace spanned by the rows of  $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ . If interest is in impulse responses to  $\varepsilon_{APITR,t}$  and  $\mathbb{E}(m_{ACITR,t} \varepsilon_{APITR,t}) = \mathbf{d}'_2 \mathbf{q}_6 = 0$ ,  $\mathbf{q}_6$  is additionally constrained to be orthogonal to  $\mathbf{d}_2$  and so lies in a 1-dimensional subspace.  $\mathbf{q}_7$  is orthogonal to  $\mathbf{q}_6$ , and so lies in the 1-dimensional subspace spanned by  $\mathbf{d}_2$ . Assuming that  $\mathbb{E}(m_{APITR,t} \varepsilon_{ACITR,t}) = 0$  yields point identification through similar reasoning (given a re-ordering of the variables to satisfy Definition 1). Assuming that  $\mathbb{E}(m_{APITR,t} \varepsilon_{ACITR,t}) = 0$  and  $\mathbb{E}(m_{ACITR,t} \varepsilon_{APITR,t}) = 0$  would yield one overidentifying restriction, but our algorithms do not allow for this.

model’s set-identified parameters.

First, we obtain impulse responses under point-identifying restrictions. When interest is in responses to the APITR, the additional point-identifying restriction is that the ACITR does not react contemporaneously to a shock in the APITR, and vice versa when interest is in responses to the ACITR. We compare estimates under these restrictions against those obtained under the set-identifying restrictions and using the single prior for  $\mathbf{Q}|\phi$  implied by Steps 2.1–2.3 of Algorithm 1 (see the discussion in Section 4). The purpose of this exercise is to explore the effect of the additional zero restriction on posterior inference. We then compare the impulse responses under the set-identifying restrictions and the single prior against those obtained using our robust Bayesian approach. This isolates the effect of the single prior on posterior inference. To quantify the sensitivity of posterior inference in this model to the choice of prior for  $\mathbf{Q}$ , we report the ‘prior informativeness’ statistic proposed in GK18, which measures the extent to which the Bayesian credible region is tightened by choosing a particular prior:

Prior informativeness =

$$1 - \frac{\text{Width of Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}{\text{Width of robust Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}. \quad (31)$$

As discussed in Section 4, we assume independent Jeffreys’ priors over the reduced-form parameters such that the VAR for  $\mathbf{y}_t$  is invertible into a VMA( $\infty$ ). The posterior is the product of independent normal-inverse-Wishart distributions, from which it is straightforward to obtain independent draws. We obtain 10,000 draws from the posterior of  $\phi$  with non-empty identified set. In the first-stage regression, we include a constant and exclude lags of the proxies. In this application, the optimization step of Algorithm 1 is slow due to the dimension of the VAR and the number of horizons considered. Consequently, we use Algorithm 2 with  $\tilde{K} = 10,000$  to approximate the bounds of the identified set at each draw of  $\phi$  via simulation.<sup>21</sup> If we cannot obtain a single draw of  $\mathbf{Q}$  satisfying the sign restrictions after 100,000 draws satisfying the zero restrictions, we approximate the identified set as being empty at that draw of  $\phi$ . Under the additional zero restriction in MR13, we obtain the point-identified object of interest at each draw of  $\phi$  by drawing a single value of  $\mathbf{Q}$  using Steps 2.1–2.2 of Algorithm 1.

Figure 1 plots impulse responses to a positive standard-deviation shock in the APITR under the point-identifying restrictions, under the set-identifying restrictions with a single

---

<sup>21</sup>We have verified that using 10,000 draws of  $\mathbf{Q}$  is sufficient to accurately approximate the bounds of the identified set for the impulse response and FEVD at the MLE of  $\phi$ .

prior for  $\mathbf{Q}|\phi$ , and under the set-identifying restrictions with the robust Bayesian approach.<sup>22</sup> The posterior distribution of the response of the APITR is similar under the point- and set-identifying restrictions when the single prior is used. Focusing on the output response, the 90 per cent highest posterior density (HPD) credible intervals include zero at all horizons under both sets of restrictions. Considering the class of all priors consistent with the set-identifying restrictions widens the credible intervals further; the prior informativeness statistic indicates that the choice of the single prior shrinks the width of the 90 credible interval for the output response by about 25 per cent on average over the horizons considered.

Figure 2 repeats Figure 1 for a shock to the ACITR. The response of the ACITR to its own shock is qualitatively similar under the two sets of identifying restrictions when a single prior is used. Under the point-identifying restrictions, the 90 per cent HPD intervals for the output response include zero at all horizons, which suggests that shocks to the ACITR have no effect on output. In contrast, under the set-identifying restrictions, the HPD intervals exclude zero at short horizons. However, inferences about the response of output are sensitive to the choice of single prior; the 90 per cent robust credible intervals for the output response include zero at all horizons and the prior informativeness statistic is about 20 per cent on average over the horizons considered.

Figure 3 plots the FEVD of output with respect to the two income tax shocks. Focusing on the posterior mean of the FEVD, the APITR shock accounts for about 20 per cent of the forecast error variance at the one-year horizon under the point-identifying restrictions. This figure falls to 10 per cent under the set-identifying restrictions and the single prior, but the result is sensitive to the choice of prior for  $\mathbf{Q}|\phi$ ; the set of posterior means includes values from about 5 per cent to about 25 per cent. Under the point-identifying restrictions, the ACITR shock accounts for around 20 per cent of the forecast error variance of output at the one-year horizon, which is similar to the contribution of the APITR under the same identifying restrictions. This contribution rises to around 26 per cent under the set-identifying restrictions and the single prior. The set of posterior means ranges from 15 to 35 per cent, which suggests that ACITR shocks explain a nontrivial share of the unexpected variation in output at short horizons regardless of the choice of prior.

Since our set-identifying restrictions include both zero and sign restrictions, the identified set may be empty at particular draws of  $\phi$ . The posterior probability that the identified set

---

<sup>22</sup>The impulse responses of government debt are omitted for brevity. Note that the posterior mean and credible intervals under the point-identifying restrictions do not necessarily lie within the set of posterior means and robust credible intervals, respectively. One reason for this is because parameter values satisfying the point-identifying restrictions do not necessarily satisfy the set-identifying restrictions; for example, under the point-identifying restrictions used to identify  $\varepsilon_{ACITR,t}$ , the posterior probability that our sign restrictions are satisfied is around 45 per cent. Another reason is that the robust credible interval is not a union of the highest posterior density intervals over the class of posteriors.

is non-empty,  $\pi_{\phi|\mathbf{Y},\mathbf{M}}(\{\phi : IS_{\eta_{i,j,h}}(\phi|F,S) \neq \emptyset\})$ , is over 90 per cent, which suggests that the identifying restrictions are consistent with the data.

## 6 Conclusion

This paper develops algorithms for robust Bayesian inference in proxy SVARs where the impulse responses or FEVDs of interest are set-identified. This approach allows researchers to relax potentially controversial point-identifying restrictions without having to specify a single, unrevisable prior over the model's set-identified parameters. This is likely to be of particular value in proxy SVARs where more than one proxy is used to identify more than one structural shock.



## References

- Angelini, G. and L. Fanelli (2019). “Exogenous Uncertainty and the Identification of Structural Vector Autoregressions with External Instruments”. *Journal of Applied Econometrics* 34 (6), pp. 951–971.
- Arias, J.E., D. Caldara and J.F. Rubio-Ramírez (2019). “The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure”. *Journal of Monetary Economics* 101, pp. 1–13.
- 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), pp. 685–720.
- (2019). “Inference in Bayesian Proxy-SVARs”. Working paper.
- Bahaj, S. (forthcoming). “Sovereign Spreads in the Euro Area: Cross Border Transmission and Macroeconomic Implications”. *Journal of Monetary Economics*.
- Baumeister, C. and J.D. Hamilton (2015). “Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information”. *Econometrica* 83 (5), pp. 1963–1999.
- (2018). “Inference in Structural Vector Autoregressions when the Identifying Assumptions are not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations”. *Journal of Monetary Economics* 100, pp. 48–65.
- (2019). “Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks”. *American Economic Review* 109 (5), pp. 1873–1910.
- Blanchard, O.J. and D. Quah (1989). “The Dynamic Effects of Aggregate Demand and Supply Disturbances”. *The American Economic Review* 79 (4), pp. 655–673.
- Braun, R. and R. Brüggemann (2017). “Identification of SVAR Models by Combining Sign Restrictions With External Instruments”. University of Konstanz Department of Economics Working Paper Series 2017-07.
- Caldara, D. and E. Herbst (2019). “Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs”. *American Economic Journal: Macroeconomics* 11 (1), pp. 157–92.
- Caldara, D. and C. Kamps (2017). “The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers”. *The Review of Economic Studies* 84 (3), pp. 1015–1040.
- Christiano, L.J., M. Eichenbaum and C.L. Evans (1999). “Chapter 2 – Monetary Policy Shocks: What Have we Learned and to What End?” In: *Handbook of Macroeconomics*. Ed. by M. Woodford and J.D. Taylor. Vol. 1. Elsevier, pp. 65–148.

- DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer Texts in Statistics. Springer-Verlag New York.
- Del Negro, M. and F. Schorfheide (2011). “Bayesian Macroeconometrics”. In: *Oxford Handbook of Bayesian Econometrics*. Ed. by J. Geweke, G. Koop and H. Van Dijk. Oxford University Press, pp. 293–389.
- Drautzburg, T. (forthcoming). “A Narrative Approach to a Fiscal DSGE Model”. *Quantitative Economics*.
- Gafarov, B., M. Meier and J.L. Montiel-Olea (2018). “Delta-Method Inference for a Class of Set-Identified SVARs”. *Journal of Econometrics* 203 (2), pp. 316–327.
- Gertler, M. and P. Karadi (2015). “Monetary Policy Surprises, Credit Costs, and Economic Activity”. *American Economic Journal: Macroeconomics* 7 (1), pp. 44–76.
- Giacomini, R. and T. Kitagawa (2018). “Robust Bayesian Inference for Set-identified Models”. cemmap Working Paper CWP61/18.
- Giacomini, R., T. Kitagawa and H. Uhlig (2019). “Estimation Under Ambiguity”. cemmap Working Paper CWP24/19.
- Granziera, E., H.R. Moon and F. Schorfheide (2018). “Inference for VARs Identified with Sign Restrictions”. *Quantitative Economics* 9 (3), pp. 1087–1121.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Jentsch, C. and K.G. Lunsford (2019). “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment”. *American Economic Review* 109 (7), pp. 2655–2678.
- Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.
- Kline, B. and E. Tamer (2016). “Bayesian Inference in a Class of Partially Identified Models”. *Quantitative Economics* 7 (2), pp. 329–366.
- Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer-Verlag New York.
- Ludvigson, S.C., S. Ma and S. Ng (2018). “Shock Restricted Structural Vector-Autoregressions”. National Bureau of Economic Research Working Paper No. 23225.
- Lunsford, K.G. (2015). “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”. *The Quarterly Journal of Economics* 133 (4), pp. 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), pp. 1212–47.

- Mertens, K. and M.O. Ravn (2014). “A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers”. *Journal of Monetary Economics* 68 (Supplement), S1–S19.
- (2019). “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply”. *American Economic Review* 109 (7), pp. 2679–91.
- Montiel-Olea, J.L., J.H. Stock and M.W. Watson (2018). “Inference in Structural Vector Autoregressions Identified With an External Instrument”. Working Paper.
- Moon, H.R. and F. Schorfheide (2012). “Bayesian and Frequentist Inference in Partially Identified Models”. *Econometrica* 80 (2), pp. 755–782.
- Piffer, M. and M. Podstawski (2018). “Identifying Uncertainty Shocks Using the Price of Gold”. *The Economic Journal* 128 (616), pp. 3266–3284.
- Plagborg-Møller, M. and C.K. Wolf (2019). “Instrumental Variable Identification of Dynamic Variance Decompositions”. Working paper.
- Poirier, D.J. (1998). “Revising Beliefs in Nonidentified Models”. *Econometric Theory* 14 (4), pp. 483–509.
- Ramey, V.A. (2016). “Chapter 2 – Macroeconomic Shocks and Their Propagation”. In: *Handbook of Macroeconomics*. Ed. by J.B. Taylor and H. Uhlig. Vol. 2. Elsevier, pp. 71–162.
- Romer, C.D. and D.H. Romer (2010). “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks”. *American Economic Review* 100 (3), pp. 763–801.
- Rubio-Ramírez, J.F., D.F. Waggoner and T. Zha (2010). “Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference”. *The Review of Economic Studies* 77 (2), pp. 665–696.
- Schervish, M.J. (1995). *Theory of Statistics*. Springer Series in Statistics. Springer-Verlag New York.
- Sims, C.A. (1980). “Macroeconomics and Reality”. *Econometrica* 48 (1), pp. 1–48.
- Staiger, D. and J.H. Stock (1997). “Instrumental Variables Regression with Weak Instruments”. *Econometrica* 65 (3), pp. 557–586.
- Stock, J.H. (2008). *What’s New in Econometrics—Time Series, Lecture 7: Structural VARs*. Cambridge, MA.: National Institute for Economic Research. URL: [http://www.nber.org/minicourse\\_2008.html](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, pp. 81–156.
- (2016). “Chapter 8 – Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics”. In: *Handbook of Macroeconomics*. Ed. by J.B. Taylor and H. Uhlig. Vol. 2. Elsevier, pp. 415–525.

- Stock, J.H. and W.H. Watson (2018). “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments”. *The Economic Journal* 128 (610), pp. 917–948.
- Stock, J.H., J.H. Wright and M. Yogo (2002). “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”. *Journal of Business & Economic Statistics* 20 (4), pp. 518–529.
- Uhlig, H. (2005). “What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure”. *Journal of Monetary Economics* 52 (2), pp. 381–419.
- (2017). “Shocks, Sign Restrictions, and Identification”. In: *Advances in Economics and Econometrics: Eleventh World Congress*. Ed. by B. Honoré A. Pakes, M. Piazzesi and L. Samuelson. Vol. 2. Cambridge University Press, pp. 95–127.

## 7 Appendix A

This appendix contains the proofs of the propositions in Section 3.1.

**Proof of Proposition 3.1.** Consider the first case (I), where the only restrictions are due to exogeneity of the proxies. If interest is in the impulse responses to the first shock,  $j^* = 1$  by Definition 1; if interest were in the responses to the  $j$ th shock for some  $j \in \{2, \dots, n - k\}$ , Definition 1 would require a re-ordering of the variables such that  $j^* = 1$ . The exogeneity assumption requires  $\mathbf{D}\Sigma_{tr}\mathbf{q}_1 = \mathbf{0}_{k \times 1}$ .  $\mathbf{q}_1$  therefore lies in the nullspace of  $\mathbf{D}\Sigma_{tr}$ , which, by the rank-nullity theorem, is a linear subspace of  $\mathbb{R}^n$  with dimension  $n - k$ . Since  $k < n - 1$  by assumption, this subspace has dimension of at least two. The sign normalization  $(\Sigma_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q}_1 \geq 0$  further constrains  $\mathbf{q}_1$  to lie in a halfspace of  $\mathbb{R}^n$ . The set of feasible  $\mathbf{q}_1$  is the intersection of the  $k$ -dimensional linear subspace satisfying the exogeneity restrictions, the halfspace generated by the sign normalization and the unit sphere, which is a path-connected set. Since the impulse response is a continuous function of  $\mathbf{q}_1$ , the identified set is an interval and is thus convex, because the set of a continuous function with a path-connected domain is always an interval.<sup>23</sup>

If, instead, interest is in responses to one of the last  $k$  shocks,  $j^* = n - k + 1$  by Definition 1; if interest were in the responses to the  $j$ th shock for some  $j \in \{n - k + 2, \dots, n\}$ , Definition 1 would require a re-ordering of the variables such that  $j^* = n - k + 1$ . For  $i = 1, \dots, n - k$ ,  $\mathbf{F}_i(\phi) = \mathbf{D}\Sigma_{tr}$ , which is a  $k \times n$  matrix with rank  $k$  under the relevance assumption.  $\mathbf{q}_i$ ,  $i = 1, \dots, n - k$ , lies in the nullspace of  $\mathbf{D}\Sigma_{tr}$ , which is of dimension  $n - k$  by the rank-nullity theorem. Since the columns of an orthonormal matrix are orthogonal,  $\mathbf{q}_{n-k+1}$  is orthogonal to this nullspace and so lies in the  $k$ -dimensional linear subspace of  $\mathbb{R}^n$  spanned by the rows of  $\mathbf{D}\Sigma_{tr}$ . By assumption,  $k > 1$ , so this subspace has dimension of at least two. The sign normalization  $(\Sigma_{tr}^{-1}\mathbf{e}_{n-k+1,n})'\mathbf{q}_{n-k+1} \geq 0$  further constrains  $\mathbf{q}_{n-k+1}$  to lie in a halfspace of  $\mathbb{R}^n$ . The set of feasible  $\mathbf{q}_{n-k+1}$  is the intersection of the  $k$ -dimensional linear subspace, the halfspace and the unit sphere, which is a path-connected set, and convexity of the identified set follows as above.<sup>24</sup>

Now consider case (II), where there are sign restrictions constraining  $\mathbf{q}_{j^*}$  only in addition to the exogeneity restrictions. Each of the  $s$  sign restrictions described by  $\mathbf{S}_{j^*}(\phi)\mathbf{q}_{j^*} \geq \mathbf{0}_{s \times 1}$  defines a halfspace in which  $\mathbf{q}_{j^*}$  must lie. The intersection of these halfspaces with the

<sup>23</sup>This result also follows directly from Proposition 3(I)(i) of GK18, since  $f_1 = k < n - 1$ .

<sup>24</sup>This result does not follow from Proposition 3 of GK18. The conditions for Proposition 3(I)(ii) are not satisfied because  $f_{j^*-1} = k \not\leq n - (j^* - 1)$ . The conditions for Proposition 3(I)(iii) are not satisfied because there does not exist  $1 \leq i^* \leq j^* - 1$  such that  $f_i < n - i$  for all  $i = i^* + 1, \dots, j^*$  and  $[\mathbf{q}_1, \dots, \mathbf{q}_{i^*}]$  is exactly identified. To see this, note that the necessary condition for exact identification of  $[\mathbf{q}_1, \dots, \mathbf{q}_{i^*}]$  is that  $f_i = n - i$  for all  $i = 1, \dots, i^*$ . But  $f_1 = k < n - 1$ , so this condition fails.

halfspace defined by the sign normalization and the linear subspace with dimension of at least two defined by the exogeneity restrictions will still have dimension of at least two. The intersection of the resulting linear subspace with the unit sphere will therefore be a path-connected set and the impulse-response identified set will be convex.

**Proof of Proposition 3.2.** Consider case (i), where  $j^* = 1$ . The optimization problem to find the upper bound of the identified set can be written as

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q} \in \mathcal{S}^{n-1}} \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q} \quad \text{s.t.} \quad \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{q} = \mathbf{0}_{k \times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1) \times 1}. \quad (32)$$

One-to-one differentiable reparameterization of this problem using  $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$  yields the optimization problem in Equation (2.5) of Gafarov *et al.* (2018). Differentiability of  $u(\boldsymbol{\phi})$  at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$  follows from their Theorem 2 under the assumptions that, at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ , the column vectors of  $\left[ (\mathbf{D}\boldsymbol{\Sigma}_{tr})', \mathbf{S}_1(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n} \right]$  are linearly independent, the set of solutions to the optimization problem is singleton, the optimized value  $u(\boldsymbol{\phi})$  is nonzero, and the number of binding sign restrictions at the optimum is less than  $n - k - 1$ . Differentiability of  $l(\boldsymbol{\phi})$  follows similarly, with  $l(\boldsymbol{\phi})$  defined as the minimizer of  $\mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}$  with respect to  $\mathbf{q} \in \mathcal{S}^{n-1}$  and subject to the same set of constraints.

Consider case (ii), where  $j^* = n - k + 1$ . Let  $\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})$  be an orthonormal basis for the nullspace of  $\mathbf{D}\boldsymbol{\Sigma}_{tr}$  (an  $(n - k) \times n$  matrix). The optimization problem to find the upper bound of the identified set can be written as

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q} \in \mathcal{S}^{n-1}} \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q} \quad \text{s.t.} \quad \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})'\mathbf{q} = \mathbf{0}_{(n-k) \times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_{n-k+1}(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1) \times 1}. \quad (33)$$

One-to-one differentiable reparameterization of this problem using  $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$  yields the optimization problem in Equation (2.5) of Gafarov *et al.* (2018) with the expanded set of equality restrictions including  $\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{x} = \mathbf{0}_{(n-k) \times 1}$ . Differentiability of  $u(\boldsymbol{\phi})$  at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$  follows from their Theorem 2 under the assumptions that, at  $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ , the column vectors of  $\left[ \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr}), \mathbf{S}_{n-k+1}(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n} \right]$  are linearly independent, the set of solutions to the optimization problem is singleton, the optimized value  $u(\boldsymbol{\phi})$  is nonzero, and the number of binding sign restrictions at the optimum is less than  $n - (n - k) - 1 = k - 1$ . Differentiability of  $l(\boldsymbol{\phi})$  follows similarly.

## 8 Appendix B

This appendix sets up the framework for the weak-proxy approximations of the posterior distribution and the sampling distribution of the MLE for the upper bound of the identified set, and derives formally the claims in (29) and (30).

As in the main text, we consider the simple setting of  $n = 3$  and  $k = 1$ , where the upper bound of the identified set  $u(\phi)$  is given by (25). Since  $u(\phi)$  depends on the reduced-form parameters only through  $(\mathbf{c}, \mathbf{d})$ , we express  $u(\phi)$  as  $u(\mathbf{c}, \mathbf{d})$ . The singularity points of  $u(\mathbf{c}, \mathbf{d})$  that we focus on are  $\mathbf{c} \neq \mathbf{0}_{3 \times 1}$  and  $\mathbf{d} = \mathbf{0}_{3 \times 1}$ , where the weak-proxy scenario corresponds to values of  $\mathbf{d}$  close to  $\mathbf{0}_{3 \times 1}$ . We hence consider a sequence of reduced-form parameters  $\{\phi_T : T = 1, 2, \dots\}$  along which the implied parameters  $(\mathbf{c}_T, \mathbf{d}_T)$ ,  $T = 1, 2, \dots$ , converge to  $(\mathbf{c}_0, \mathbf{0}_{3 \times 1})$ ,  $\mathbf{c}_0 \neq \mathbf{0}_{3 \times 1}$ , as  $T \rightarrow \infty$ . As in the main text, we specify a drifting sequence of  $\{\phi_T\}$  that leads to

$$\begin{pmatrix} \mathbf{c}_T \\ \mathbf{d}_T \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 + \boldsymbol{\gamma}/\sqrt{T} \\ \boldsymbol{\delta}/\sqrt{T} \end{pmatrix}, \quad (34)$$

where  $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^3 \times \mathbb{R}^3$  are the localisation parameters.

Let  $\hat{\mathbf{S}}_T \in \mathbb{R}^s$ ,  $s < \infty$ ,  $T = 1, 2, \dots$ , be a finite-dimensional vector of sufficient statistics for  $\phi$  that converges in distribution to a random vector  $\hat{\mathbf{S}} \in \mathbb{R}^s$  as  $T \rightarrow \infty$ . Since we consider a Gaussian proxy SVAR, these sufficient statistics are the first and second sample moments of the observables. By the Skhorohod representation theorem, we can embed this sequence of sufficient statistics  $\{\hat{\mathbf{S}}_T\}$  and the limiting random variables  $\hat{\mathbf{S}}$  into a common probability space on which

$$\hat{\mathbf{S}}_T \rightarrow \hat{\mathbf{S}} \text{ as } T \rightarrow \infty, \text{ almost surely,} \quad (35)$$

holds.

Let  $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$  be the MLE of  $(\mathbf{c}, \mathbf{d})$ . Since the MLE depends only on the sufficient statistics  $\hat{\mathbf{S}}_T$ , we can embed the MLE into the probability space on which  $\{\hat{\mathbf{S}}_T\}$  and  $\hat{\mathbf{S}}$  are commonly defined. Hence, conditioning on the sequence of sufficient statistics  $\{\hat{\mathbf{S}}_T : T = 1, 2, \dots\}$  pins down the constant sequence of MLEs. We assume that the (unconditional) sampling distribution of the MLEs centered at the drifting true values is asymptotically normal:

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_T - \mathbf{d}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}_{6 \times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right). \quad (36)$$

Following the Skhorohod representation for the sufficient statistics (35), we have the almost-

sure convergence of the MLE to the limiting Gaussian random variables

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \rightarrow \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \text{ as } T \rightarrow \infty, \text{ almost surely,} \quad (37)$$

on the common probability space. We also impose a high-level assumption of the strong consistency of the MLE for  $\mathbf{c}$  in the sense of

$$\hat{\mathbf{c}}_T \rightarrow \mathbf{c}_0 \text{ as } T \rightarrow \infty, \text{ almost surely,} \quad (38)$$

on the same probability space.

Since the posterior distribution depends on the data only through the sufficient statistics, it suffices to consider the convergence of the posterior distribution for  $u(\mathbf{c}, \mathbf{d})$  conditional on the sequence of sufficient statistics  $\{\hat{\mathbf{S}}_T\}$ . We assume that the posterior for  $(\mathbf{c}, \mathbf{d})$  centered at their MLEs is asymptotically normal in the following sense. Let

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \mathbf{c} - \hat{\mathbf{c}}_T \\ \mathbf{d} - \hat{\mathbf{d}}_T \end{pmatrix}, \quad (39)$$

and assume

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_c \\ \mathbf{Z}_d \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}_{6 \times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right), \quad (40)$$

for almost every conditioning sequence of  $\{\mathbf{S}_T\}$ . We assume that the asymptotic posterior variance given in (40) is independent of the conditioning variable  $\{\hat{\mathbf{S}}_T : T = 1, 2, \dots\}$  and coincides with the asymptotic variance of the MLE given in (36).

The asymptotic normality of the posterior (centered at the MLE with data-independent variance) holds for a wide class of regular parametric models, and its almost-sure coincidence with the asymptotic (sampling) distribution of the MLE leads to the Bernstein-von Mises Theorem. See, for instance, Schervish (1995) and DasGupta (2008) for a set of sufficient conditions for posterior asymptotic normality.

Under these assumptions, we obtain the following weak-proxy asymptotic approximation of the posterior for  $u(\phi)$ .

**Proposition 8.1.** *Consider a drifting sequence of reduced-form parameters that satisfy (34) with  $\mathbf{c}_0 \neq \mathbf{0}_{3 \times 1}$ , along which we assume that the MLE for  $(\mathbf{c}, \mathbf{d})$  and its posterior satisfies (36), (37), (38) and (40). Then, for almost every conditioning sequence of the sufficient*



statistics  $\{\hat{\mathbf{S}}_T\}$ , the asymptotic posterior of  $u(\mathbf{c}, \mathbf{d})$  is

$$u(\mathbf{c}, \mathbf{d}) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d) = \sqrt{\mathbf{c}'_0 \left( \mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d\|^2} \right) \mathbf{c}_0},$$

where  $\hat{\mathbf{Z}}_d$  is a constant given the sampling sequence, and  $\mathbf{Z}_d \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Omega}_d)$ .

*Proof.* Since  $u(\mathbf{c}, \mathbf{d})$  is homogeneous of degree zero with respect to  $\mathbf{d}$ , we have

$$\begin{aligned} u(\mathbf{c}, \mathbf{d}) &= u(\mathbf{c}, T^{1/2}\mathbf{d}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, T^{1/2}\hat{\mathbf{d}}_T + \mathbf{Z}_{dT}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, T^{1/2}\mathbf{d}_T + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT}), \end{aligned}$$

where the second equality uses (39), the third equality uses (36), and the fourth equality uses (34). Conditional on the sampling sequence of the sufficient statistics  $\{\hat{\mathbf{S}}_T\}$ , the assumptions of almost-sure convergence (37) and (38) and the posterior distributional convergence (40) imply

$$\begin{pmatrix} \hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT} \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d \end{pmatrix}, \quad (41)$$

as  $T \rightarrow \infty$ , where  $(\mathbf{c}_0, \boldsymbol{\delta}, \hat{\mathbf{Z}}_d)$  are constants and  $\mathbf{Z}_d$  is a random vector following  $\mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Omega}_d)$ . Since  $u(\mathbf{c}, \mathbf{d})$  is discontinuous at  $\mathbf{d} = \mathbf{0}_{3 \times 1}$ , and  $\{\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d = \mathbf{0}_{3 \times 1}\}$  is the null event in terms of the probability law of the limiting random variables, an application of the continuous mapping theorem (see, for example, Theorem 10.8 of Kosorok (2008)) yields the conclusion.  $\square$

The next proposition gives the asymptotic sampling distribution of  $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$ .

**Proposition 8.2.** *Consider a drifting sequence of reduced-form parameters that satisfy (34) with  $\mathbf{c}_0 \neq \mathbf{0}_{3 \times 1}$ , along which we assume that the MLE of  $(\mathbf{c}, \mathbf{d})$  satisfies (36). Then, the asymptotic distribution of  $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$  is*

$$u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta} + \hat{\mathbf{Z}}_d) = \sqrt{\mathbf{c}'_0 \left( \mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0},$$

where  $\hat{\mathbf{Z}}_d \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Omega}_d)$ .

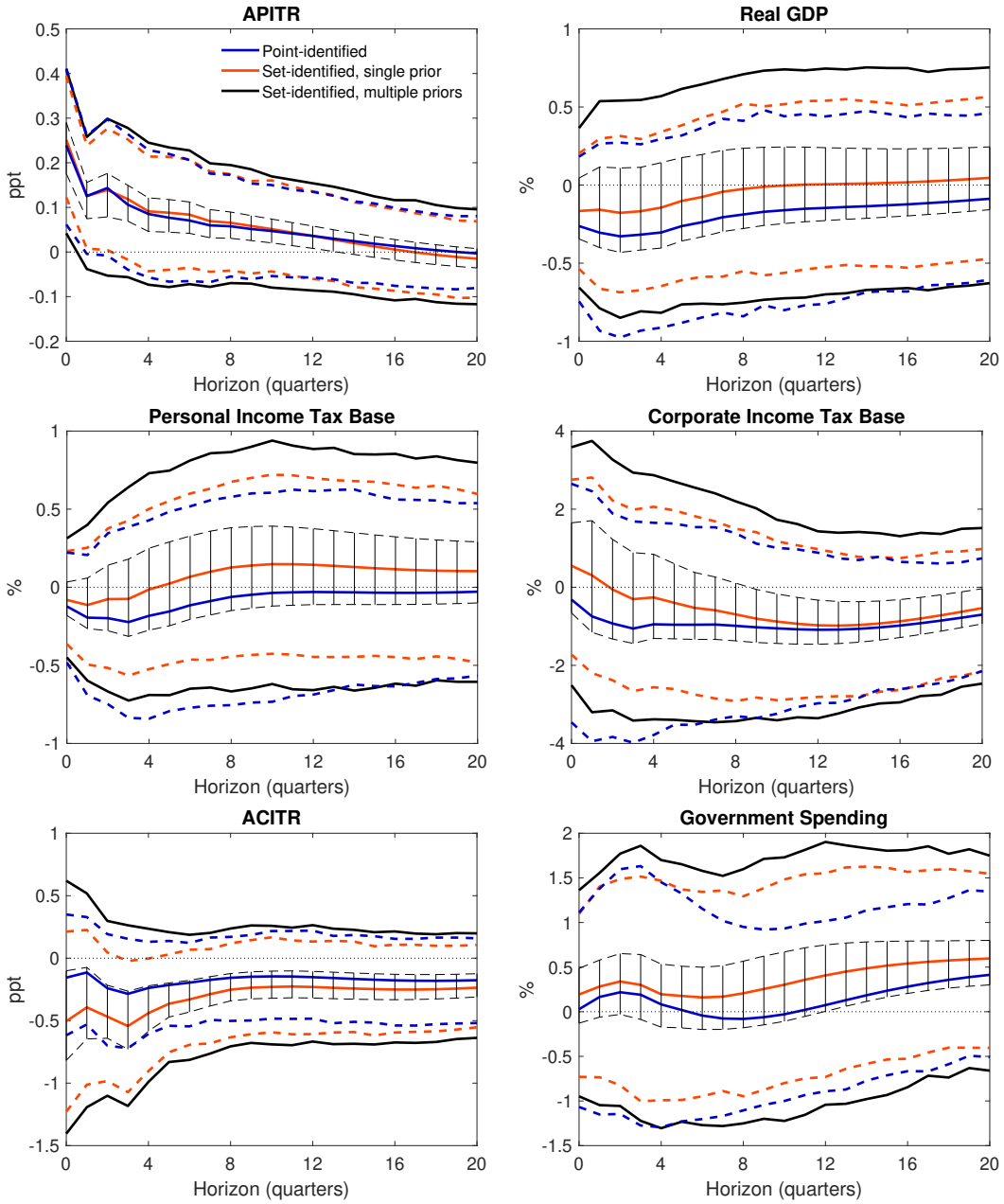
*Proof.* Since  $u(\mathbf{c}, \mathbf{d})$  is homogeneous of degree zero with respect to  $\mathbf{d}$ , it holds that  $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T) =$

$u(\hat{\mathbf{c}}_T, T^{1/2}\hat{\mathbf{d}}_T)$ . Under the drifting sequence (34) and  $\sqrt{T}$ -asymptotic normality of the MLE (36),

$$\begin{pmatrix} \hat{\mathbf{c}}_T \\ T^{1/2}\hat{\mathbf{d}}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_d \end{pmatrix}.$$

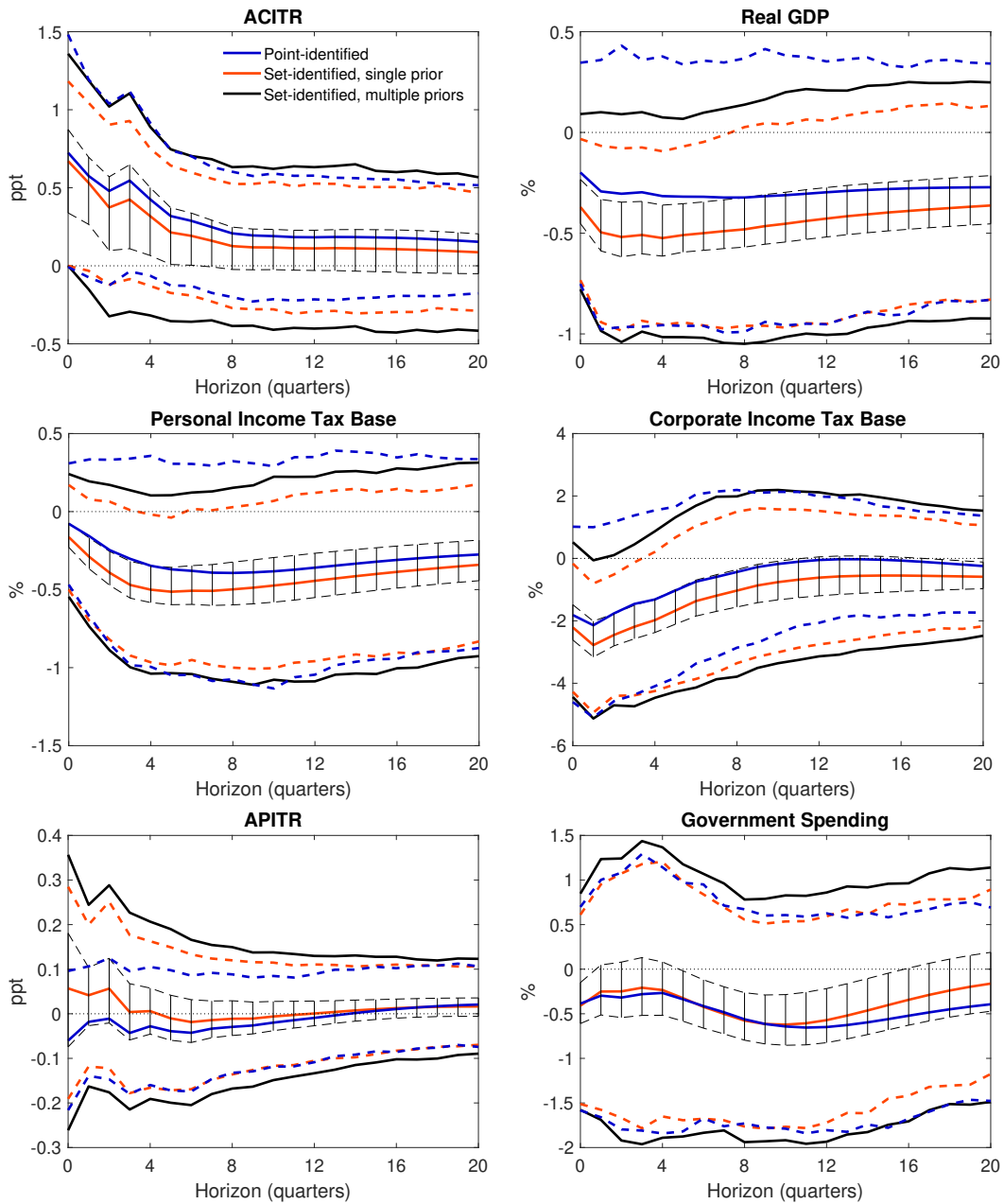
Noting that  $\{\boldsymbol{\delta} + \hat{\mathbf{Z}}_d = \mathbf{0}_{3 \times 1}\}$  is a null event in terms of the limiting probability law, an application of the continuous mapping theorem leads to the conclusion.  $\square$

Figure 1: Impulse Responses to APITR Shock



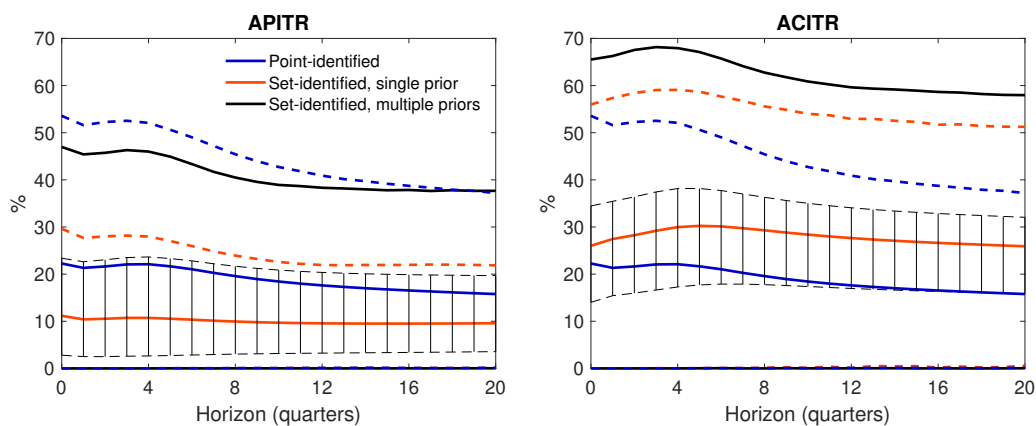
Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.

Figure 2: Impulse Responses to ACITR Shock



Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.

**Figure 3: Contribution of Tax Shocks to Forecast Error Variance of Real GDP**



Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.