Robust Bayesian Inference in Proxy SVARs

Raffaella Giacomini
Toru Kitagawa
Matthew Read

Robust Bayesian Inference in Proxy SVARs[*]

Raffaella Giacomini[†], Toru Kitagawa[‡] and Matthew Read[§]

July 11, 2019

Abstract

We develop methods for robust Bayesian inference in structural vector autoregressions (SVARs) where the impulse responses or forecast error variance decompositions of interest are set-identified using external instruments (or 'proxy SVARs'). Existing Bayesian approaches to inference in proxy SVARs require researchers to specify a single prior over the model's parameters. When parameters are set-identified, a component of the prior is never updated by the data. Giacomini and Kitagawa (2018) propose a method for robust Bayesian inference in set-identifed models that delivers inference about the identified set for the parameter of interest. We extend this approach to proxy SVARs, which allows researchers to relax potentially controversial point-identifying restrictions without having to specify an unrevisable prior. We also explore the effect of instrument strength on posterior inference. We illustrate our approach by revisiting Mertens and Ravn (2013) and relaxing the assumption that they impose to obtain point identification.

1   Introduction

Proxy structural vector autoregressions (SVARs) are an increasingly popular method for estimating the dynamic causal effects of macroeconomic shocks.[1] 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). Examples of papers that estimate impulse responses to a single shock using a single proxy include Gertler and Karadi (2015) in the frequentist setting and Caldara and Herbst (2019) in the Bayesian setting. Mertens and Ravn (2013) (henceforth MR13) develop a proxy SVAR with two instruments for two structural shocks and show that point identification of the impulse responses to these shocks requires a zero restriction on the structural parameters in addition to the zero restrictions implied by exogeneity of the proxies (see also Arias, Rubio-Ramírez and Waggoner (2019) for a discussion of this point). Examples of other papers that use multiple proxies to identify multiple structural shocks 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 does so in the frequentist setting. Notable exceptions are Braun and Brüggemann (2017) and Arias, Rubio-Ramírez and Waggoner (2019) (henceforth ARW19), who propose 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.[2] However, under set identification, posterior inference is sensitive to the choice of prior over the set-identified parameters, even asympotically (Poirier 1998),

---

[1]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), Arias, Rubio-Ramírez and Waggoner (2019), Caldara and Herbst (2019) and Jentsch and Lunsford (2019).

[2]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.

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 objects of interest, such as impulse responses. To address these issues, Giacomini and Kitagawa (2018) (GK18) propose an approach to Bayesian inference in set-identifed 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 the associated 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 identifying restrictions. For other values of n and k, it may be the case that particular impulse responses are point-identified, while other impulse responses are set-identified. We focus on cases where the impulse responses of interest are set-identified. Additionally, we extend the algorithms developed in GK18 to conduct posterior inference 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.[3] We also provide an algorithm for computing unit-effect impulse responses (as opposed to unit-standard-deviation impulse responses), which are often considered in the proxy SVAR literature.

As in Braun and Brüggemann (2017) and 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

---

[3]Plagborg-Moller and Wolf (2019) develop frequentist inferential procedures for 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.

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.[4] Additionally, our algorithms allow for sign restrictions and zero restrictions of the kind considered in GK18, including 'short-run' zero restrictions (as in Sims (1980)), '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 whose elements determine the contemporaneous relationships among the endogenous variables (as in Arias, Caldara and Rubio-Ramírez (2019)).

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 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. Their approach also allows identifying restrictions to be imposed non-dogmatically. 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 structural parameters into a revisable prior over reduced-form parameters and an unrevisable prior over the 'rotation matrix' that maps VAR innovations into structural shocks (see, for example, Uhlig (2005)). We then allow for multiple priors for the rotation matrix, which delivers inference that is robust to the choice of unrevisable priors. 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.[5]

---

[4]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), ARW19 and Piffer and Podstawski (2018). Piffer and Podstawski (2018) allow for the possibility 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.

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

It is well-known that frequentist inference in the linear instrumental-variables (IV) 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, particularly if they use these procedures purely for computational convenience. Accordingly, we investigate the asymptotic properties of our robust Bayesian procedure in the presence of weak instruments.[6] We further provide some discussion about the implications of weak proxies for the numerical implementation of our method.

We illustrate our procedure by considering the analysis in MR13, which is also discussed in Mertens and Ravn (2019) and Jentsch and Lunsford (2019). MR13 use series of plausibly exogenous, unanticipated changes in 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. MR13 impose a zero restriction in addition to those implied by exogeneity of the proxies to achieve point identification, which is equivalent to restricting the direct contemporaneous response of one tax rate to the other. This assumption could be violated if, for instance, there are political economy 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 can be seen as taking a step forward and formalizing the robustness analysis by providing an estimator of the set of impulse responses compatible with relaxing the additional restriction. We also compare the results under our multiple-prior Bayesian approach to those obtained

---

[6]We are not aware of previous work considering the problem of weak instruments in set-identified models, including set-identified versions of the linear IV model such as that considered in Nevo and Rosen (2012).

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 and provides results on the frequentist properties of this approach. Section 3 investigates how weak proxies affect posterior inference. Section 4 details the numerical algorithms used to implement the approach. Section 5 contains the empirical application described above and Section 6 concludes.

2 Framework

2.1 The SVAR

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

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

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 $(\mathbf{y}_{1-p}, ..., \mathbf{y}_0)$ are given. We omit exogenous regressors (such as a constant) for simplicity of exposition, but these are straightforward to include. Letting $\mathbf{x}_t = (\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p})$ and $\mathbf{A}_+ = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, we can rewrite the SVAR(p) as

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

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

$$\mathbf{y}_t = \mathbf{B} \mathbf{x}_t + \mathbf{u}_t, \quad t = 1, ..., T, \tag{3}$$

where $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)$, $\mathbf{B}_l = \mathbf{A}_0^{-1} \mathbf{A}_l$ for $l = 1, \ldots, p$, 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) model.

To facilitate computing the identified set of the impulse response, we reparameterise the model into its 'orthogonal reduced form':

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

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 (or 'rotation') matrix and $\mathcal{O}(n)$ is the set of all such matrices. The parameterisations 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 $\operatorname{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}$.

The VMA representation of the model is

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

where $\mathbf{C}_h$ is the hth 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 ith variable to the jth structural shock at the hth horizon:

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

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

Another object that is often of interest is the relative contribution of a particular shock to the unexpected variation in a particular variable over some horizon, or 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 $\operatorname{var}(y_{i,t+h}|\mathcal{F}_t) = \sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{c}_{i,k}$. The contribution of the jth structural shock to the forecast error variance of the ith variable at the hth horizon is $\operatorname{var}(y_{i,t+h}|\mathcal{F}_t, \boldsymbol{\varepsilon}_{-j,t+1}, \ldots, \boldsymbol{\varepsilon}_{-j,t+h}) = \sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{q}_j\mathbf{q}_j'\mathbf{c}_{i,k}$, where $\boldsymbol{\varepsilon}_{-j,t} = \{\varepsilon_{i,t} : i \neq j \land i = 1, \ldots, n\}$. The contribution of the jth structural shock to the forecast error variance of the ith variable at the hth horizon as a fraction of the total forecast error variance is then

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

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 and 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 for the structural parameters and any associated objects of interest.

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 $\boldsymbol{\varepsilon}_{(i:j),t} = (\varepsilon_{i,t}, \varepsilon_{i+1,t}, ..., \varepsilon_{j-1,t}, \varepsilon_{j,t})$ for $i < j$. Assume that $\mathbf{m}_t$ is a $k \times 1$ vector of proxies 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 $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:n-k),t}) = \mathbf{0}_{k,n-k}$. 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

$$\boldsymbol{\Gamma}_0 \mathbf{m}_t = \boldsymbol{\Lambda} \boldsymbol{\varepsilon}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\nu}_t, \quad t = 1, ..., T, \tag{7}$$

where: $\boldsymbol{\Gamma}_l$, $l = 0, ..., p_m$, is a $k \times k$ matrix with $\boldsymbol{\Gamma}_0$ invertible; $\boldsymbol{\Lambda}$ is a $k \times n$ matrix; and the initial conditions $(\mathbf{m}_{1-p_m}, \ldots, \mathbf{m}_0)$ 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 (noisy) information about the structural shocks after allowing for possible serial correlation in the proxies. 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 of $\mathbf{y}_t$ in the SVAR for $\mathbf{y}_t$.[7]

Given the distributional assumption on $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$, and the exogeneity and relevance assumptions, it follows from (7) that

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_t) = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} = [\mathbf{0}_{k,n-k}, \boldsymbol{\Psi}]. \tag{8}$$

Left-multiplying (7) by $\boldsymbol{\Gamma}_0^{-1}$ and substituting out $\mathbf{y}_t$ using (2) yields

$$\mathbf{m}_t = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_0 \mathbf{y}_t - \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Lambda} \mathbf{A}_+ \mathbf{x}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\nu}_t. \tag{9}$$

---

[7] ARW19 specify a joint SVAR for $(\mathbf{y}'_t, \mathbf{m}'_t)'$ where zero restrictions rule out contemporaneous feedback from $\mathbf{m}_t$ to $\mathbf{y}_t$. This process also implies that the proxies contain (noisy) information about the structural shocks and yields the same set of identifying zero restrictions that we derive below.

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

$$\mathbf{m}_t = \mathbf{D}\mathbf{y}_t + \mathbf{G}\mathbf{x}_t + \sum_{l=1}^{p_m} \mathbf{H}_l \mathbf{m}_{t-l} + \mathbf{v}_t, \tag{10}$$

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 $l = 1, ..., p_m$; 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 $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda} = \mathbf{D}\mathbf{A}_0^{-1} = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, we can write (8) as

$$\mathbb{E}(\mathbf{m}_t\boldsymbol{\varepsilon}_t') = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q} = [\mathbf{0}_{k,n-k}, \boldsymbol{\Psi}]. \tag{11}$$

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 $\mathbf{d}_i'$ is the ith row of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$. The exogeneity condition is therefore equivalent to linear restrictions on the first $n-k$ columns of $\mathbf{Q}$ given the reduced-form parameters $\mathbf{D}$ and $\boldsymbol{\Sigma}_{tr}$. The relevance condition rank($\boldsymbol{\Psi}$) $= k$ holds if and only if rank($\mathbf{D}$) $= k$.

Let f_i be the number of equality restrictions on the ith 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 $i = 1, ..., n$. We focus on cases where, for all $i = 1, ..., n$, $f_i \leq n - i$, with strict inequality for at least one i, and where interest is in a particular set-identified impulse response or FEVD.

Assume for now that the only zero restrictions are those corresponding to the exogeneity assumption and that $n \geq 3$. Assume also that rank($\mathbf{D}$) $= k$, so the relevance condition holds. If $k = 1$, then $f_i = 1$ for $i = 1, ..., n-1$ and $f_n = 0$. In this case, the first $n-1$ columns of $\mathbf{Q}$ (and thus the impulse responses to the first $n-1$ structural shocks) are set-identified, but $\mathbf{q}_n$ is point-identified.[8] If $k = n-1$, then $f_1 = n-1$ and $f_i = 0$ for $i = 2, ..., n$. In this case, $\mathbf{q}_1$ is point-identified, but $\mathbf{q}_i$, $i = 2, ..., n$, is set-identified.[9] For $1 < k < n-1$, in the absence of additional zero restrictions, all columns of $\mathbf{Q}$ – and thus all impulse responses –

---

[8]To see why $\mathbf{q}_n$ is point-identified, note that the exogeneity restrictions imply that $\mathbf{d}_1'\mathbf{q}_i = 0$ for $i = 1, ..., n-1$; that is, $\mathbf{q}_i$ is orthogonal to $\mathbf{d}_1'$ for $i = 1, ..., n-1$. Since the columns of an orthonormal matrix are orthogonal and have unit length, $\mathbf{q}_n = \pm\mathbf{d}_1'/\|\mathbf{d}_1'\|$, where $\|.\|$ is the Euclidean norm. The sign normalization pins down the sign of $\mathbf{q}_1$, so it is point-identified.

[9]To see why $\mathbf{q}_1$ is point-identified, note that the exogeneity restrictions imply that $\mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{q}_1 = 0$, where $\mathbf{D}$ is a $(n-1) \times n$ matrix. Assuming that the relevance condition rank($\mathbf{D}$) $= k$ holds, the nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ is a linear subspace of $\mathbb{R}^n$. Since the columns of an orthonormal matrix are orthogonal and have unit length, $\mathbf{q}_1$ is a unit-length vector in the nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$, which is point-identified given a sign normalization.

are set-identified.[10]

As in Braun and Brüggemann (2017) and ARW19, we allow for additional equality and sign restrictions involving covariances between the proxies and shocks. 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. $\mathbb{E}(m_{1t}\varepsilon_{(n-k+1),t}) = 0$). These types of restrictions are linear restrictions on the columns of $\mathbf{Q}$. An example of a sign restriction is that the covariance between the first proxy and one of the last $n - k$ 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 zero and sign restrictions commonly used in SVARs, such as zero restrictions on $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$, $\mathbf{A}_0^{-1} = \mathbf{Q}\boldsymbol{\Sigma}_{tr}$ and on the long-run cumulative impulse response $CIR^\infty = (\mathbf{I}_n - \sum_{l=1}^{p} \mathbf{B}_l)^{-1}\boldsymbol{\Sigma}_{tr}\mathbf{Q}$. Sign restrictions can be placed on the impulse responses or on elements of $\mathbf{A}_0$.

2.3   Robust Bayesian Inference About the Impulse Responses

We assume for now that the object of interest is an impulse response, although the discussion in this section also applies to the FEVD (or any other scalar object of interest that is a 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 of the impulse responses proceeds similarly to GK18. We summarise the salient features of this approach here.

Collect the coefficients on $\mathbf{x}_t$ and $\mathbf{m}_t$ in (10) as $\mathbf{J} = [\mathbf{G}, \mathbf{H}_1, \ldots, \mathbf{H}_{p_m}]$. The proxy-SVAR reduced-form parameters are

$$\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})', \text{vec}(\mathbf{D})', \text{vec}(\mathbf{J})', \text{vech}(\boldsymbol{\Upsilon})')' \in \boldsymbol{\Phi} \subset \mathbb{R}^{np+n(n+1)/2+(p+1)nk+p_m k^2+k(k+1)/2}.$$

$$(12)$$

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

---

[10]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.

general form

$$F(\boldsymbol{\phi}, \mathbf{Q}) = \begin{bmatrix} F_1(\boldsymbol{\phi})\mathbf{q}_1 \\ \vdots \\ F_n(\boldsymbol{\phi})\mathbf{q}_n \end{bmatrix} = \mathbf{0}_{(\sum_{i=1}^n f_i) \times 1}, \tag{13}$$

where $F_i(\boldsymbol{\phi})$ 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 $\mathbf{q}_i$, $F_i(\boldsymbol{\phi})$ does not exist and $f_i = 0$. When the only zero restrictions are those relating to exogeneity of the proxies, the relevance condition implies that, for almost every $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, $\text{rank}(F_i(\boldsymbol{\phi})) = f_i$ for $i = 1, \dots, n - k$. 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).

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

$$IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F, S) = \{\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi}|F, S)\}, \tag{14}$$

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

$$\mathcal{Q}(\boldsymbol{\phi}|F, S) = \{\mathbf{Q} \in \mathcal{O}(n) : F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{(\sum_i^n f_i) \times 1}, S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}, \text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}\}. \tag{15}$$

Let $\pi_{\boldsymbol{\phi}}$ 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 $\pi_{\boldsymbol{\theta}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}}$, 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 Braun and Brüggemann (2017) and 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 over the rotation matrix, which has no direct economic interpretation, and even 'uniform' priors 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 $\mathbf{Q}|\boldsymbol{\phi}$ that are consistent with the identifying restrictions:

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}} = \left\{ \pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}(\mathcal{Q}(\boldsymbol{\phi}|F, S)) = 1 \right\}. \tag{16}$$

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

$$\Pi_{\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}} = \left\{ \pi_{\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}} \pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}} \in \Pi_{\mathbf{Q}|\boldsymbol{\phi}} \right\}, \tag{17}$$

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

$$\left[\int_{\boldsymbol{\Phi}} l(\boldsymbol{\phi}) d\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}, \int_{\boldsymbol{\Phi}} u(\boldsymbol{\phi}) d\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}} \right], \tag{18}$$

where $l(\boldsymbol{\phi}) = \inf\{\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi}|F, S)\}$ and $u(\boldsymbol{\phi}) = \sup\{\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi}|F, S)\}$. They also suggest reporting a robust credible region with credibility level α (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 α 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 $\boldsymbol{\phi}$. The posterior probability that the identified set is non-empty, $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}(\{\boldsymbol{\phi} : IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F, S) \neq \emptyset\})$, can thus be used to quantify the plausibility of the identifying restrictions.

In order to implement our robust Bayesian inferential approach, it is necessary to order the variables in $\mathbf{y}_t$ to satisfy Definition 1 (which mirrors Definition 3 in GK18).

Definition 1 (Ordering of Variables): Order the variables in $\mathbf{y}_t$ so that f_i satisfies $f_1 \geq f_2 \geq \ldots \geq f_n \geq 0$. 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 vector has a larger number of restrictions than $\mathbf{q}_{j^*}$. If $j^* \geq 2$, order the variables so that $f_{j^*-1} > f_{j^*}$.[11]

---

[11]The ordering is not necessarily unique.

The following example illustrates how to order the variables to satisfy Definition 1.

Example 1: Consider a proxy SVAR for (c_t, i_t, y_t, π_t), where c_t is consumption growth, i_t is investment growth, y_t is output growth and π_t is inflation. Assume that there exist two proxy variables, $m_{c,t}$ and $m_{i,t}$, 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 that to $\varepsilon_{i,t}$, an ordering of the variables that satisfies Definition 1 is (y_t, π_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 that to $\varepsilon_{\pi,t}$, an ordering of the variables that satisfies Definition 1 is (π_t, y_t, i_t, c_t), with $(f_1, f_2, f_3, f_4) = (2, 2, 0, 0)$ and $j^* = 1$.

2.4 Frequentist Validity

In this section we use results from GK18 to ascertain 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 practitioners who use Bayesian approaches to inference purely for computational convenience.

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 a continuous and convex function of ϕ. If, in addition, $[l(\phi), u(\phi)]$ is differentiable in ϕ at the true value ϕ_0, the robust credible region is an asymptotically valid confidence set for the true identified set. Propositions 3 and 4 of GK18 provide conditions under which the impulse-response identified set is guaranteed to be convex and continuous in ϕ, respectively, while Proposition 5 provides conditions under which it is guaranteed to be differentiable in ϕ. We can show that having the relevance condition satisfied at ϕ_0 is a necessary condition for continuity of the identified-set correspondence at ϕ_0. Differentiability of the bounds additionally requires that a unique set of sign restrictions are binding at the bounds of the identified set in a neighborhood around ϕ_0. In what follows, we focus on verifying the conditions under which convexity holds. We discuss issues associated with 'weak' proxies – where the relevance condition is 'close' to being violated – in the next section.

Assume that there are k proxy variables correlated with k structural shocks and uncorrelated with the remaining $n - k$ structural shocks, and that there are no additional zero or sign restrictions. First, consider the case where $n \geq 3$ and there is one proxy, so $k = 1$. Then $f_i = 1$ for $i = 1, \ldots, n - 1$ and $f_n = 0$. If 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), then $j^* = 1$ by Definition 1 and $f_1 < n-1$, which implies that the identified set is convex for almost every $\phi \in \mathbf{\Phi}$ by Proposition 3(I)(i) of GK18. If there are proxies for all-but-one structural shock, so $k = n-1$, then $f_1 = n-1$ and $f_i = 0$ for $i = 2, \ldots, n$. If interest is in the impulse responses to one of the last $n-1$ structural shocks (the impulse responses to the first structural shock are point-identified), then $j^* = 2$ by Definition 1. As long as $\text{rank}(F_1(\phi)) = n-1$ for almost every $\phi \in \mathbf{\Phi}$, then $\mathbf{q}_1$ is exactly identified and the impulse-response identified set is convex for almost every $\phi \in \mathbf{\Phi}$ by Proposition 3(I)(iii). In these two cases, assuming that the relevance condition holds, frequentist validity follows under the regularity conditions in Propositions 4 and 5 of GK18.[12]

Now, consider the case where $n \geq 4$ and $1 < k < n-1$, which implies that the impulse responses to all structural shocks are set-identified. $f_i = k$ for $i = 1, \ldots, n-k$ and $f_i = 0$ for $i = n-k+1, \ldots, n$. If interest is in the impulse responses to one of the first $n-k$ structural shocks, then $j^* = 1$ by Definition 1. Since $f_1 = n-k < n-1$, the identified set is convex for almost every $\phi \in \mathbf{\Phi}$ by Proposition 3(I)(i), and frequentist validity follows under the regularity conditions in Propositions 4 and 5 of GK18. If, instead, interest is in the impulse responses to one of the last k structural shocks, then $j^* = n-k+1$ by Definition 1. In this case, the conditions in Proposition 3(I)(ii) or (iii) are not satisfied, so we cannot guarantee convexity of the identified set.[13] Nevertheless, the set of posterior means and robust credible region can be interpreted as providing valid posterior inference – and valid asymptotic frequentist inference – about the *convex hull* of the identified set.[14]

3   Weak Proxies

In this section, we investigate how weak proxies affect robust Bayesian posterior inference about set-identified impulse responses in the proxy SVAR. For simplicity of analysis and

---

[12]To illustrate how proxy relevance affects continuity, consider the first case. Continuity requires that there exists an open neighbourhood $G \subset \mathbf{\Phi}$ around the true value of the reduced-form parameter ϕ_0 such that $\text{rank}(F_1(\phi)) = 1$ for all $\phi \in G$, and that there exists a unit-length vector satisfying the zero restriction and sign normalization constraining $\mathbf{q}_1$ (see Proposition 4(i)). If the relevance condition were to fail, $F_1(\phi_0) = \mathbf{0}_{1 \times n}$ and $\text{rank}(F_1(\phi_0)) = 0$.

[13]The conditions for Proposition 3(I)(ii) are not satisfied because $f_{j^*-1} = k \not< n-(j^*-1)$. 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 $i = i^*+1, \ldots, j^*$ and $[\mathbf{q}_1, \ldots, \mathbf{q}_{i^*}]$ is exactly identified. To see this, note that the necessary condition for exact identification of $[\mathbf{q}_1, \ldots, \mathbf{q}_{i^*}]$ is that $f_i = n-i$ for all $i = 1, \ldots, i^*$. But $f_1 = k < n-1$, so this condition fails.

[14]When there are also sign restrictions that constrain $\mathbf{q}_{j^*}$ only, convexity of the identified set can be checked at each draw of ϕ using Proposition 3(II)(iv)–(v).

exposition, we focus on the case where $n = 3$, $k = 1$ and the objects of interest are the (set-identified) impulse responses to ε_{1t}. We choose this case because it is straightforward to analytically characterise the identified set and hence discuss the effects of weak proxies.

Since the objects of interest are the impulse responses to ε_{1t}, we set $j^* = 1$ and $(f_1, f_2, f_3) = (1, 1, 0)$. At a given value of $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ (and ignoring the sign normalization), the upper bound of the identified set for the impulse response of the ith variable to the first structural shock at the h-th horizon is the value function associated with the following optimisation problem:

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q} \in \mathcal{S}^{n-1}} \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q} \quad \text{subject to} \quad F_1(\boldsymbol{\phi})\mathbf{q} = 0,$$

where $\mathcal{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$. Applying the change of variables $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$ yields the problem in Equation (2.5) of Gafarov *et al.*. Using their results, the value function satisfies

$$u(\boldsymbol{\phi})^2 = \mathbf{c}'_{i,h}(\boldsymbol{\phi}) \left[\mathbf{I}_n - F_1(\boldsymbol{\phi})' \left(F_1(\boldsymbol{\phi})F_1(\boldsymbol{\phi})' \right)^{-1} F_1(\boldsymbol{\phi}) \right] \mathbf{c}_{i,h}(\boldsymbol{\phi})$$

$$= \mathbf{c}'_{i,h}(\boldsymbol{\phi}) \left[\mathbf{I}_n - \mathbf{d}_1 \left(\mathbf{d}'_1 \mathbf{d}_1 \right)^{-1} \mathbf{d}'_1 \right] \mathbf{c}_{i,h}(\boldsymbol{\phi}) \tag{19}$$

where the second line uses the fact that, in this case, $F_1(\boldsymbol{\phi}) = \mathbf{D}\boldsymbol{\Sigma}_{tr} \equiv \mathbf{d}'_1$.[15] (11) implies that $\mathbf{d}'_1 \mathbf{q}_3 = \Psi$, where $\Psi = \mathbb{E}(m_t \varepsilon_{3t})$. The exogeneity restrictions require that $\mathbf{d}'_1 \mathbf{q}_1 = 0$ and $\mathbf{d}'_1 \mathbf{q}_2 = 0$; that is, $\mathbf{q}_1$ and $\mathbf{q}_2$ are orthogonal to $\mathbf{d}_1$. Since the columns of an orthonormal matrix are orthogonal and have unit length, it must therefore be the case that $\mathbf{q}_3 = \pm \mathbf{d}_1 / \|\mathbf{d}_1\|$, where $\|.\|$ is the Euclidean norm. This implies that $\mathbf{d}'_1 \mathbf{d}_1 / \|\mathbf{d}_1\| = \|\mathbf{d}_1\| = |\Psi|$.

For ease of notation, in this section we henceforth refer to $\mathbf{c}_{i,h}(\boldsymbol{\phi})$ simply as $\mathbf{c}(\boldsymbol{\phi})$. A 'weak' proxy correlates with one of the structural shocks only weakly, so $|\Psi|$ is close to zero. This is equivalent to $\|\mathbf{d}_1\|$ being small. Note that $u(\boldsymbol{\phi})$ as the square-root of (19) is continuous and smooth in $c(\boldsymbol{\phi})$, while it is discontinuous in $\mathbf{d}_1$ at $\mathbf{d}_1 = \mathbf{0}_{3 \times 1}$. Hence, if the posterior distribution of $\mathbf{d}_1$ concentrates near $\mathbf{0}_{3 \times 1}$ due to the weak proxy, the posterior of $u(\boldsymbol{\phi})$ can exhibit a nonstandard distribution due to the singularity of $u(\boldsymbol{\phi})$ at $\mathbf{d}_1 = \mathbf{0}_{3 \times 1}$ even when the posterior of $(\mathbf{c}'(\boldsymbol{\phi}), \mathbf{d}'_1)'$ is consistent and can be well approximated by a normal distribution centered at their maximum likelihood estimates.

To investigate the posterior for $u(\boldsymbol{\phi})$ in the weak proxy case, we consider the local asymptotic approximation of the posterior for $u(\boldsymbol{\phi})$ with a drifting sequence of the true values of $\boldsymbol{\phi}$

---

[15]In the absence of sign normalization restrictions, the lower bound of the identified set $l(\boldsymbol{\phi})$ is given by $-u(\boldsymbol{\phi})$. This section hence focuses only on the posterior for $u(\boldsymbol{\phi})$.

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 $\{(\mathbf{c}_T, \mathbf{d}_{1T}) : T = 1, 2, \dots,\}$ converging to a point of singularity. Following the weak instrument asymptotics of Staiger and Stock (1997), we consider the drifting sequence of $(\mathbf{c}, \mathbf{d}_1)$ with $T^{-1/2}$-convergence rate,

$$\mathbf{c}_T = \mathbf{c}_0 + \frac{\boldsymbol{\gamma}}{\sqrt{T}}, \quad \mathbf{d}_{1T} = \frac{\boldsymbol{\delta}}{\sqrt{T}}, \tag{20}$$

where $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^6$ is the vector of localisation parameters, and the magnitude of $\boldsymbol{\delta}$ characterises the relevance of the proxy; that is, the smaller the $\|\boldsymbol{\delta}\|$, the weaker the proxy.

Let $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T})$ be the maximum likelihood estimator (MLE) for $\mathbf{c}(\phi)$ and $\mathbf{d}_1$ (which are constants once we have conditioned on the sample). We assume that the sampling distribution of the MLE is $\sqrt{T}$-asymptotically normal,

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_{1T} - \mathbf{d}_{1T} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right). \tag{21}$$

We also assume that the posterior for $(\mathbf{c}'(\phi), \mathbf{d}'_1)'$ converges to a normal distribution with data-independent variance. That is, conditional on the sampling sequence,

$$\sqrt{T} \begin{pmatrix} \mathbf{c}(\phi) - \hat{\mathbf{c}}_T \\ \mathbf{d}_1 - \hat{\mathbf{d}}_{1T} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_c \\ \mathbf{Z}_d \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix} \right), \tag{22}$$

as $T \to \infty$ for almost every sampling sequence, and $\boldsymbol{\Omega} \equiv \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}'_{cd} & \boldsymbol{\Omega}_d \end{pmatrix}$ is the posterior asymptotic variance, which does not depend on the sampling sequence. The asymptotic equivalence of the probability laws in (21) and (22) implies that the reduced-form parameters $(\mathbf{c}, \mathbf{d}_1)$ 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 7.1 in the Appendix derives the following asymptotic

16

approximation of the posterior for $u(\boldsymbol{\phi})$: conditional on the sampling sequence,

$$u(\boldsymbol{\phi}) \xrightarrow{d} \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d\|^2} \right) \mathbf{c}_0}, \tag{23}$$

as $T \to \infty$ for almost every sampling sequence, where $\hat{\mathbf{Z}}_d$ is a constant depending 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 parameters $\boldsymbol{\delta}$, 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 $\{\boldsymbol{\phi}_T : T = 1, 2, \dots\}$ yielding (20),

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

This implies that, under the current weak-proxy asymptotics, the set of posterior means for the impulse response considered in GK18 is not a consistent estimator for the identified set.

Under the same drifting sequence inducing (20), the asymptotic sampling distribution of the MLE for the upper bound of the identified set is (see Proposition 7.2 in Appendix):

$$u(\hat{\boldsymbol{\phi}}) \xrightarrow{d} \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0}, \tag{24}$$

where $\hat{\mathbf{Z}}_d \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$. Like the posterior for $u(\boldsymbol{\phi})$, the sampling distribution of $u(\hat{\boldsymbol{\phi}})$ is not consistent and remains non-degenerate. A comparison of (23) and (24) 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, implying that the Bernstein-von Mises property does not hold in the presence of the weak proxy. This violates one of the sufficient conditions of Theorem 3 in GK18 for asymptotically valid frequentist coverage.

In the strong proxy case where $|\Psi| = \|\mathbf{d}_1\|$ is far from zero, the pointwise asymptotic approximation of the posterior of $u(\boldsymbol{\phi})$ approximates well the finite-sample posterior. Noting that $u(\boldsymbol{\phi})$ is smooth at $\mathbf{d}_1 \neq \mathbf{0}_{3 \times 1}$ and assuming that the posterior of $(\mathbf{c}, \mathbf{d}_1)$ centered at the MLE is $\sqrt{T}$-asymptotically normal, the delta method implies that $\sqrt{T}(u(\boldsymbol{\phi}) - u(\hat{\boldsymbol{\phi}}))$ is asymptotically normal with a data-independent variance. This asymptotic posterior coincides with the sampling distribution of the MLE, so in addition to posterior consistency, the correct frequentist coverage of the robust credible region can be attained. This stark contrast in the asymptotic behavior of the posteriors suggests that in the current simple setting, whether the posterior of $u(\boldsymbol{\phi})$ is close to a normal distribution could be useful for diagnosing whether the proxy is strong or 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 impulse responses 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 details specific to the proxy SVAR case below. Matlab code implementing the algorithms is available on the authors' personal websites or on request.

Algorithm 1. Let $\mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{(\sum_{i=1}^{n} f_i) \times 1}$ and $S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}$ 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. The algorithm proceeds as follows:

- **Step 1**: Specify a prior for the reduced-form parameters, $\pi_{\boldsymbol{\phi}}$, and obtain the posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}$.[16]

- **Step 2**: Draw $\boldsymbol{\phi}$ from $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}$. Check whether $\mathcal{Q}(\boldsymbol{\phi}|F, S)$ is empty using the subroutine below.

---

[16]$\pi_{\boldsymbol{\phi}}$ does not have to be proper or to satisfy the condition $\pi_{\boldsymbol{\phi}}(\{\boldsymbol{\phi} : \mathcal{Q}(\boldsymbol{\phi}|F, S) \neq \emptyset\}) = 1$; that is, the prior may assign positive probability to regions of the reduced-form parameter space that yield an empty set of rotation matrices satisfying the identifying restrictions.

- **Step 2.1**: Draw $\mathbf{z}_1 \sim N(\mathbf{0}_{n\times1}, \mathbf{I}_n)$. Let $\tilde{\mathbf{q}}_1 = \mathbf{M}_1\mathbf{z}_1$ be the $n \times 1$ residual vector in the linear projection of $\mathbf{z}_1$ onto an $n \times f_1$ regressor matrix $F_1(\boldsymbol{\phi})'$, where $\mathbf{M}_1 = [\mathbf{I}_n - F_1(\boldsymbol{\phi})'(F_1(\boldsymbol{\phi})F_1(\boldsymbol{\phi})')^{-1}F_1(\boldsymbol{\phi})]$. For $i = 2, \ldots, n$, run the following procedure sequentially: draw $\mathbf{z}_i \sim N(\mathbf{0}_{n\times1}, \mathbf{I}_n)$ and compute $\tilde{\mathbf{q}}_i = \mathbf{M}_i\mathbf{z}_i$, where $\mathbf{M}_i\mathbf{z}_i$ is the residual vector in the linear projection of $\mathbf{z}_i$ onto the $n \times (f_i + i - 1)$ regressor matrix $[F_i(\boldsymbol{\phi})', \tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_{i-1}]$. The vectors $\tilde{\mathbf{q}}_i$, $i = 1, \ldots, n$, are orthogonal and satisfy the zero restrictions represented in $F(\boldsymbol{\phi}, \mathbf{Q})$.[17]

- **Step 2.2**: Given $\tilde{\mathbf{q}}_i$, $i = 1, \ldots, n$, define

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

  where $\|.\|$ is the Euclidean norm in $\mathbb{R}^n$. If $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{i,n})'\tilde{\mathbf{q}}_i = 0$ for some i, set $\mathrm{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_i)$ equal to 1 or -1 with equal probability. This step rescales $\tilde{\mathbf{q}}_i$, $i = 1, \ldots, n$, to have unit length and imposes the sign normalization that the diagonal elements of $\mathbf{A}_0$ are nonnegative.

- **Step 2.3**: Check whether $\mathbf{Q}_0$ satisfies $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times1}$. 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 $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times1}$. If no draws of $\mathbf{Q}_0$ satisfy $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times1}$, approximate $\mathcal{Q}(\boldsymbol{\phi}|F, S)$ as being empty and return to Step 2.

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

$$l(\boldsymbol{\phi}) = \min_{\mathbf{Q}} \mathbf{c}_{i,h}(\boldsymbol{\phi})'\mathbf{q}_j$$

subject to

$$F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{(\sum_i^n f_i)\times1}$$
$$S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times1}$$
$$\mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times1}$$
$$\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n.$$

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

---

[17]If the relevance condition fails, $F_i(\boldsymbol{\phi})$ is of reduced row rank for $i = 1, \ldots, n - k$ and the coefficients in the linear projection of $\mathbf{z}_i$ on $F_i(\boldsymbol{\phi})'$ are not identified. This is a measure zero event so long as the prior for $\boldsymbol{\phi}$ does not place positive probability mass on the event $\mathrm{rank}(\mathbf{D}) < k$.

- **Step 4**: Repeat Steps 2–3 M times to obtain $[l(\boldsymbol{\phi_m}), u(\boldsymbol{\phi_m})]$ for $m = 1, ..., M$. Approximate the set of posterior means by the sample averages of $l(\boldsymbol{\phi_m})$ and $u(\boldsymbol{\phi_m})$.

- **Step 5**: To obtain an approximation of the smallest robust credible region with credibility $\alpha \in (0, 1)$, define $d(\eta, \boldsymbol{\phi}) = \max\{|\eta - l(\boldsymbol{\phi})|, |\eta - u(\boldsymbol{\phi})|\}$ and let $\hat{z}_\alpha(\eta)$ be the sample α-th quantile of $\{d(\eta, \boldsymbol{\phi_m}), m = 1, ..., M\}$. 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**: The proportion of draws of $\boldsymbol{\phi}$ passing Step 2.3 is an approximation of the posterior probability that the identified set is nonempty: $\tilde{\pi}_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}(\{\boldsymbol{\phi} : \mathcal{Q}(\boldsymbol{\phi}|F, S) \neq \emptyset\})$.

4.1 Remarks

4.1.1 Choice of priors

The method used to draw $\boldsymbol{\phi}|\mathbf{Y}, \mathbf{M}$ 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, $\pi_{\boldsymbol{\phi}} = \pi_{\mathbf{B},\boldsymbol{\Sigma}}\pi_{\mathbf{D},\mathbf{J},\boldsymbol{\Upsilon}}$, where $\pi_{\mathbf{B},\boldsymbol{\Sigma}} \propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}}$ and $\pi_{\mathbf{D},\mathbf{J},\boldsymbol{\Upsilon}} \propto |\boldsymbol{\Upsilon}|^{-\frac{k+1}{2}}$. This makes it easy to draw from the posterior of $\boldsymbol{\phi}|\mathbf{Y}, \mathbf{M}$, since it is the product of independent normal-inverse-Wishart posteriors.[18] An advantage of using independent priors is that it allows the sample for $\mathbf{m}_t$ to be shorter than the sample for $\mathbf{y}_t$ without discarding information that could otherwise be used to estimate the VAR for $\mathbf{y}_t$, which can be the case in empirical applications (e.g. Gertler and Karadi 2015). 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 $\boldsymbol{\phi}$. In particular, 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.

4.1.2 Convergence issues

As in GK18, the optimisation step (Step 3) involves minimising (or maximising) a linear function subject to linear and quadratic equality restrictions and linear inequality restrictions. This is a quadratically constrained linear program and hence is nonconvex. Consequently, the convergence of gradient-based optimisation methods in this problem is not guaranteed.

---

[18]This follows from the fact that the joint likelihood of $(\mathbf{M}, \mathbf{Y})$ is multiplicatively separable across the two blocks of parameters. 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.

Accordingly, we suggest drawing multiple values of $\mathbf{Q}_0$ in Steps 2.2 and 2.3 to use as initial values in the optimisation 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 optimisation step.

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

- **Step 3':** Iterate Steps 2.1–2.3 K times and let $\{\mathbf{Q}_l, l = 1, ..., \tilde{K}\}$ be the draws of $\mathbf{Q}$ that satisfy the sign restrictions (either fix K and let $\tilde{K}$ vary or fix $\tilde{K}$ and let K vary). Let $\mathbf{q}_{j,l}$ be the jth column of $\mathbf{Q}_l$. Approximate $[l(\boldsymbol{\phi}), u(\boldsymbol{\phi})]$ 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 $\boldsymbol{\phi}$. However, the approximated identified set will converge to the true identified set as $\tilde{K}$ goes to infinity. 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 numerical optimisation 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 $\mathbf{Q}$, in which case it may take many draws of $\mathbf{Q}$ given the equality restrictions to generate a number of draws satisfying the sign restrictions that is sufficient to accurately approximate the bounds of the identified set. In practice, Step 3 and Step 3' are parallelisable, so large reductions in computating time are possible in both algorithms by distributing computation across multiple processors.

4.1.3 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 $\boldsymbol{\phi}$ induces a single posterior for $\eta_{i,j^*,h}$ and/or $FEVD_{i,j^*,h}$. In this case, Steps 1 and 2.1–2.2 of Algorithm 2 can still be used to draw from the posterior of the object of interest. 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 $\eta_{i,j^*,h}$ or $FEVD_{i,j^*,h}$. We make use of this in the empirical application below when estimating a proxy SVAR under point-identifying restrictions.

4.1.4 Other objects of interest

When interest is in the FEVD rather than the impulse response, Algorithms 1 and 2 can be modified by replacing $\mathbf{c}_{i,h}(\boldsymbol{\phi})'\mathbf{q}_j$ with $\sum_{k=0}^{h-1}\mathbf{c}_{i,k}(\boldsymbol{\phi})'\mathbf{q}_j\mathbf{q}_j'\mathbf{c}_{i,k}(\boldsymbol{\phi})/\sum_{k=0}^{h-1}\mathbf{c}_{i,k}(\boldsymbol{\phi})'\mathbf{c}_{i,k}(\boldsymbol{\phi})$. In practice, the optimisation step of Algorithm 1 is likely to be computationally demanding when the objective is the FEVD, so we recommend using Algorithm 2 to numerically approximate the bounds of the identified set via simulation. If one is interested in both impulse responses and the FEVD, 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 both the impulse response and the FEVD rather than having to carry out numerical optimisation for each object and for each variable and horizon. It would be straightforward to extend the algorithms to other objects that are a function of the structural parameters, such as the contribution of a particular shock to the observed unexpected change in a particular variable between two dates (i.e. the historical decomposition). Note also that when interest is in the *cumulative* impulse response, $\mathbf{c}_{i,h}(\boldsymbol{\phi})'\mathbf{q}_j$ is simply replaced with $\left(\sum_{k=1}^{h}\mathbf{c}_{i,k}(\boldsymbol{\phi})'\right)\mathbf{q}_j$.

4.1.5 Impulse responses to a unit shock

The algorithms above impose the unit standard deviation 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 shock that is one standard deviation in magnitude (see, for example, Stock and Watson (2016, 2018) for a discussion of this point). Algorithm 3 shows how to obtain the set of posterior means and the robust credible interval for impulse responses to a unit shock.

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

- **Step 3''**: Iterate Steps 2.1–2.3 K times and let $\{\mathbf{Q}_l, l = 1, ..., \tilde{K}\}$ be the draws of $\mathbf{Q}$ that satisfy the sign restrictions (either fix K and let $\tilde{K}$ vary or fix $\tilde{K}$ and let K vary). Let $\mathbf{A}_{0,l}^{-1} = \boldsymbol{\Sigma}_{tr}\mathbf{Q}_l$. Rescale the ith column of $\mathbf{A}_{0,l}^{-1}$ such that the ith element is equal to one; that is, set $\mathbf{a}_{i,l} = (\mathbf{A}_{0,l}^{-1}\mathbf{e}_{i,n})/(\mathbf{e}_{i,n}'\mathbf{A}_{0,l}^{-1}\mathbf{e}_{i,n})$. Approximate $[l(\boldsymbol{\phi}), u(\boldsymbol{\phi})]$ by $[\min_l \mathbf{e}_{i,n}'\mathbf{C}_h\mathbf{a}_i, \max_l \mathbf{e}_{i,n}\mathbf{C}_h\mathbf{a}_i]$.

The algorithm generates impulse responses to a one standard deviation shock that are consistent with the identifying restrictions, rescales the impulse responses so that they are with respect to a unit shock, and computes the bounds of the identified set using the extreme values of the rescaled impulse responses. We note that one potential issue arising under the unit-effect normalization is that the posterior mean bounds and robust credible interval may be unbounded when the relevant diagonal elements of $\mathbf{A}_0^{-1} = \boldsymbol{\Sigma}_{tr}\mathbf{Q}$ are not bounded away

from zero for all $\phi \in \Phi$ and $\mathbf{Q} \in \mathcal{Q}(\phi | F, S)$.

4.1.6 Weak proxies

When the proxy is 'weak', in the sense that $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ is 'almost' rank deficient at some or all draws of ϕ, our algorithms still deliver a numerical approximation of the exact finite-sample posterior of the identified set. However, there may be problems with the quality of the numerical approximation. First, the optimisation step in Algorithm 1 may be poorly behaved when the matrix of linear equality restrictions is close to rank deficient. Second, Algorithm 2 requires repeating Step 2.1 many times, which involves projecting a vector onto a matrix representing the zero and orthogonality restrictions. If this matrix is close to rank deficient, the matrix inversion step involved in the linear projection will be inaccurate. This could then affect the accuracy of the bounds of the identified set. However, these issues should be readily identifiable in practice. In the Matlab code implementing our approach, the constrained nonlinear optimiser ('fmincon') prints a warning when the matrix of constraints is close to rank deficient. Similarly, Matlab prints warnings when attempting to invert an ill-conditioned matrix. In both cases, proximity to rank deficiency is assessed using a condition number (e.g., the 2-norm condition number is the ratio of the maximum and minimum singular values), with an infinitely large condition number representing singularity. Practitioners could also compute the condition number of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ at each draw of ϕ and plot its posterior or compute the posterior probability that it is larger than some threshold.

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 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.[19]

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

To achieve point identification, MR13 impose different zero restrictions. When interest is in the impulse responses to a structural shock to the APITR, they assume that the ACITR does not respond directly to a structural shock in the APITR on impact, and vice versa when interest is in the impulse responses to a structural shock to the ACITR. Note, however, that these restrictions do not necessarily mean that the tax rates do not respond at all to the shock in the other tax rate, since the tax rates can still respond to the other variables, which themselves are able to respond directly to the shock. It can be shown that these additional zero restrictions are equivalent to zero restrictions on $\mathbf{A}_0$.[21] These zero

---

[19] We thank Morten Ravn and Karel Mertens for making their data available. We obtained the data from Karel Mertens' website: https://karelmertens.com/research/.

[20] Assuming that $\mathbb{E}(m_{APITR,t}\varepsilon_{ACITR,t}) = 0$ and $\mathbb{E}(m_{APITR,t}\varepsilon_{ACITR,t}) = 0$ would yield one overidentifying restriction. Our algorithms rule out this possibility.

[21] ARW19 show that the additional identifying restrictions used in the proxy SVARs of MR13 and Mertens and Montiel-Olea (2018) are equivalent to restrictions on $\mathbf{A}_0$. Mertens and Montiel-Olea (2018) estimate the effect of shocks to different personal income tax rates, but their identifying assumptions are analogous to those in MR13. In our setting, the additional zero restriction is $\mathbf{e}'_{7,7}\mathbf{A}_0\mathbf{e}_{6,7} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{7,7})'\mathbf{q}_6 = 0$.

restrictions could be violated if, for instance, there are political economy constraints that impinge on the ability of the government to change personal and corporate income tax rates independently of one another. Consequently, we investigate the sensitivity of the results to replacing the additional point-identifying zero restriction with set-identifying sign restrictions. In particular, 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 $\mathbb{E}(m_{ACITR,t}\varepsilon_{ACITR,t}) \geq \mathbb{E}(m_{ACITR,t}\varepsilon_{APITR,t})$). One advantage of this approach is that the identifying restrictions are the same regardless of which shock is of interest, whereas MR13 impose a different point-identifying restriction on $\mathbf{A}_0$ depending on whether the shock of interest is the APITR shock or the ACITR shock. Importantly, our approach allows us to relax the additional point-identifying zero restriction while avoiding the need to impose a single, unrevisable prior over the model's set-identified parameters.

First, we compare impulse responses obtained under the point-identifying restrictions used in MR13 against those obtained under the set-identifying restrictions and using a single prior for $\mathbf{Q}|\boldsymbol{\phi}$. 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 over the rotation matrix, 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 =
$$1 - \frac{\text{Width of Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}{\text{Width of robust Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}. \quad (25)$$

As discussed in Section 4, we assume independent Jeffreys' priors over the reduced-form parameters, $\pi_{\boldsymbol{\phi}} = \pi_{\mathbf{B},\boldsymbol{\Sigma}}\pi_{\mathbf{D},\mathbf{G},\boldsymbol{\Upsilon}}$, where $\pi_{\mathbf{B},\boldsymbol{\Sigma}} \propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}}$ and $\pi_{\mathbf{D},\mathbf{G},\boldsymbol{\Upsilon}} \propto |\boldsymbol{\Upsilon}|^{-\frac{k+1}{2}}$. The posterior is thus the product of independent normal-inverse-Wishart distributions, from which it is straightforward to obtain independent draws. In practice, we obtain 1,000 draws from the posterior of $\boldsymbol{\phi}$ (with non-empty identified set) by estimating separate Bayesian reduced-form VARs for $\mathbf{y}_t$ and $\mathbf{m}_t$ (with $(\mathbf{y}_t, \ldots, \mathbf{y}_{t-p})$ included as exogenous variables in the latter). In the first-stage regression, we include a constant and exclude lags of the proxies. When the impulse response of interest is point-identified, we obtain the impulse responses at each draw

of ϕ by drawing a single value of $\mathbf{Q}$ using Steps 2.1–2.2 of Algorithm 1. When the impulse responses are set-identified, the single prior for $\mathbf{Q}|\phi$ is that implied by Steps 2.1–2.3 of Algorithm 1. We present impulse responses to one-standard-deviation increases in the structural shocks of interest. In this application, the optimisation step of Algorithm 1 is slow due to the large dimension of the VAR. Consequently, we use Algorithm 2 with $\tilde{K} = 10,000$ to approximate the bounds of the identified set at each draw of ϕ via simulation. If we cannot obtain a draw of $\mathbf{Q}$ satisfying the identifying restrictions after 10,000 draws of $\mathbf{Q}$, we approximate the identified set as being empty at that draw of ϕ.

Figure 1 plots impulse responses to a positive one standard deviation shock in the APITR under the point-identifying restrictions used in MR13 and under our set-identifying restrictions but with a single prior for $\mathbf{Q}|\phi$.[22] The two sets of restrictions yield very similar posterior distributions for the response of the APITR to its own shock. In contrast, the posterior distribution of the responses of the other variables differs somewhat across the two sets of identifying restrictions. For example, the posterior probability that the response of output is negative after one year is 80 per cent under the point-identifying restrictions, whereas this probability is only 68 per cent under the set-identifying restrictions. A notable difference between the two sets of results is the response of the ACITR. Under the point-identifying restrictions, the 90 per cent highest posterior density (HPD) intervals for the response of the ACITR include zero at all horizons and the posterior mean response is about 0.2 percentage points on average in the first year. Under the set-identifying restrictions, the HPD intervals exclude zero at some horizons within the first year after the shock and the posterior mean response is about 0.5 percentage points on average in the first year.

Figure 2 plots the impulse responses to an APITR shock under the set-identifying restrictions and using our robust Bayesian approach. The impulse responses under the single prior are also reported for ease of comparison. The set of posterior means (the estimator of the identified set) for the impulse response of output to the APITR shock includes zero at all horizons and at the one-year horizon spans from -0.40 to 0.16. The posterior probability that the response of output is negative after one year is between 0.27 and 0.94 across the priors consistent with the identifying restrictions. 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 40 per cent on average over the horizons considered. These results indicate that posterior inferences about the effect of output depend heavily on the choice of single prior for $\mathbf{Q}|\phi$. Notably, the 90 per cent robust credible intervals for the impulse

---

[22]The impulse responses of government debt are omitted for brevity.

response of the ACITR include zero at all horizons, whereas the HPD intervals do not. In other words, using the single prior suggests that the ACITR falls following an APITR shock with nontrivial posterior probability, but this result is not robust to the choice of prior for $\mathbf{Q}|\phi$.

Figures 3 and 4 repeat Figures 1 and 2, but for a shock to the ACITR. The response of the ACITR 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 AICTR 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 30 per cent on average over the horizons considered.

Figures 5 and 6 plot the FEVDs of output with respect to the two income tax shocks under the different sets of identifying assumptions and the two approaches to posterior inference. Focusing on the posterior mean of the FEVD, the APITR shock accounts for about 20 per cent of the forecast error variance in output 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 over the rotation matrix: the set of posterior means ranges from about 5 per cent to about 25 per cent. Under the point-identifying restrictions, the ACITR shock accounts for close to 25 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 30 per cent under the set-identifying restrictions and the single prior. The set of posterior means ranges from 15 to 40 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 ϕ. The posterior probability that the identified set is non-empty, $\pi_{\phi|\mathbf{Y},\mathbf{M}}(\{\phi : IS_{\eta_{i,j,h}}(\phi|F,S) \neq \emptyset\})$, is about 96 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

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 (2019). "Inference in Bayesian Proxy-SVARs". Working paper.

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). "Identication 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.

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.

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.

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.

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.

— (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.

Nevo, A. and A.M. Rosen (2012). "Identification With Imperfect Instruments". *The Review of Economics and Statistics* 93 (3), pp. 659–671.

Piffer, M. and M. Podstawski (2018). "Identifying Uncertainty Shocks Using the Price of Gold". *The Economic Journal* 128 (616), pp. 3266–3284.

Plagborg-Moller, 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`.

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.

7 Appendix

This appendix sets up the framework for the weak-proxy approximations for the posterior distribution and the sampling distribution (of the MLE) for the upper bound of the identified set, and derives formally the claims (23) and (24) in the main text.

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 (19). Since $u(\phi)$ depends on the reduced-form parameters only through $(\mathbf{c}, \mathbf{d}_1)$, we express $u(\mathbf{c}, \mathbf{d}_1) \equiv u(\phi)$. The singularity points of $u(\mathbf{c}, \mathbf{d}_1)$ that we focus on are $\mathbf{c} \neq \mathbf{0}_{3 \times 1}$ and $\mathbf{d}_1 = \mathbf{0}_{3 \times 1}$, where the weak proxy scenario corresponds to the value of $\mathbf{d}_1$ 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 $(\mathbf{c}_T, \mathbf{d}_{1T})$, $T = 1, 2, \dots$, converge to $(\mathbf{c}_0, \mathbf{0}_{3 \times 1})$, $\mathbf{c}_0 \neq \mathbf{0}_{3 \times 1}$, as $T \to \infty$. As in the main text, we specify a drifting sequence of $\{\phi_T\}$ that leads to

$$
\begin{pmatrix} \mathbf{c}_T \\ \mathbf{d}_{1T} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 + \gamma/\sqrt{T} \\ \delta/\sqrt{T} \end{pmatrix},
\tag{26}
$$

where $(\gamma, \delta) \in \mathbb{R}^3 \times \mathbb{R}^3$ 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 ϕ that converges in distribution to a random vector $\mathbf{S} \in \mathbb{R}^s$ as $T \to \infty$. 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

$$
\hat{\mathbf{S}}_T \to \hat{\mathbf{S}} \text{ as } T \to \infty, \text{ almost surely},
\tag{27}
$$

holds.

Let $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T})$ be the MLE for $\mathbf{c}$ and $\mathbf{d}_1$. 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

32

normal:

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_{1T} - \mathbf{d}_{1T} \end{pmatrix} \overset{d}{\to} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}_{6\times1}, \begin{pmatrix} \mathbf{\Omega}_c & \mathbf{\Omega}_{cd} \\ \mathbf{\Omega}'_{cd} & \mathbf{\Omega}_d \end{pmatrix} \right). \tag{28}$$

Following the Skhorohod representation for the sufficient statistics (27), we have the almost sure convergence of the MLE to the limiting Gaussian random variables

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \to \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \text{ as } T \to \infty, \text{ almost surely,} \tag{29}$$

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

$$\hat{\mathbf{c}}_T \to \mathbf{c}_0 \text{ as } T \to \infty, \text{ almost surely,} \tag{30}$$

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 $u(\mathbf{c}, \mathbf{d}_1)$ conditional on the sequence of sufficient statistics $\{\hat{\mathbf{S}}_T\}$. We assume that the posterior for $(\mathbf{c}, \mathbf{d}_1)$ centered at their MLEs is asymptotically normal in the following sense. Let

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \mathbf{c} - \hat{\mathbf{c}}_T \\ \mathbf{d}_1 - \hat{\mathbf{d}}_{1T} \end{pmatrix}, \tag{31}$$

and we assume

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \overset{d}{\to} \begin{pmatrix} \mathbf{Z}_c \\ \mathbf{Z}_d \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}_{6\times1}, \begin{pmatrix} \mathbf{\Omega}_c & \mathbf{\Omega}_{cd} \\ \mathbf{\Omega}'_{cd} & \mathbf{\Omega}_d \end{pmatrix} \right), \tag{32}$$

for almost every conditioning sequence of $\{\mathbf{S}_T\}$. We assume that the asymptotic posterior variance given in (32) 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 (28).

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 $u(\boldsymbol{\phi})$.

Proposition 7.1. *Consider a drifting sequence of reduced-form parameters that satisfy (26) with $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, along which we assume that the MLE for $(\mathbf{c}, \mathbf{d}_1)$ and its posterior satisfies (28), (29), (30) and (32). Then, for almost every conditioning sequence of the sufficient statistics $\{\hat{\mathbf{S}}_T\}$, the asymptotic posterior of $u(\mathbf{c}, \mathbf{d}_1)$ is*

$$u(\mathbf{c}, \mathbf{d}_1) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d) = \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d\|^2} \right) \mathbf{c}_0},$$

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 $u(\mathbf{c}, \mathbf{d}_1)$ is homogeneous of degree zero with respect to $\mathbf{d}_1$, we have

$$u(\mathbf{c}, \mathbf{d}_1) = u(\mathbf{c}, T^{1/2}\mathbf{d}_1) = u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, T^{1/2}\hat{\mathbf{d}}_{1T} + \mathbf{Z}_{dT})$$
$$= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT}),$$

where the second equality uses (28), and the third equality uses (31). Conditional on the sampling sequence of the sufficient statistics $\{\hat{\mathbf{S}}_T\}$, the assumptions of almost-sure convergence (29) and (30) and the posterior distributional convergence (32) imply

$$\begin{pmatrix} \hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT} \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d \end{pmatrix}, \tag{33}$$

as $T \to \infty$, 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 $u(\mathbf{c}, \mathbf{d}_1)$ is discontinuous at $\mathbf{d}_1 = \mathbf{0}_{3\times 1}$, 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, e.g., Theorem 10.8 of Kosorok (2008)) yields the conclusion. $\square$

The next proposition gives the asymptotic sampling distribution of $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T})$.

Proposition 7.2. *Consider a drifting sequence of reduced-form parameters that satisfy (26) with $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, along which we assume that the MLE of $(\mathbf{c}, \mathbf{d}_1)$ satisfies (28). Then, the asymptotic distribution of $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T})$ is*

$$u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T}) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta} + \hat{\mathbf{Z}}_d) = \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0},$$

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

Proof. Since $u(\mathbf{c}, \mathbf{d}_1)$ is homogeneous of degree zero with respect to $\mathbf{d}_1$, it holds that $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_{1T}) = u(\hat{\mathbf{c}}_T, T^{1/2}\hat{\mathbf{d}}_{1T})$. Under the drifting sequence (26) and $\sqrt{T}$-asymptotic normality of the MLE (28),

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

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. $\square$

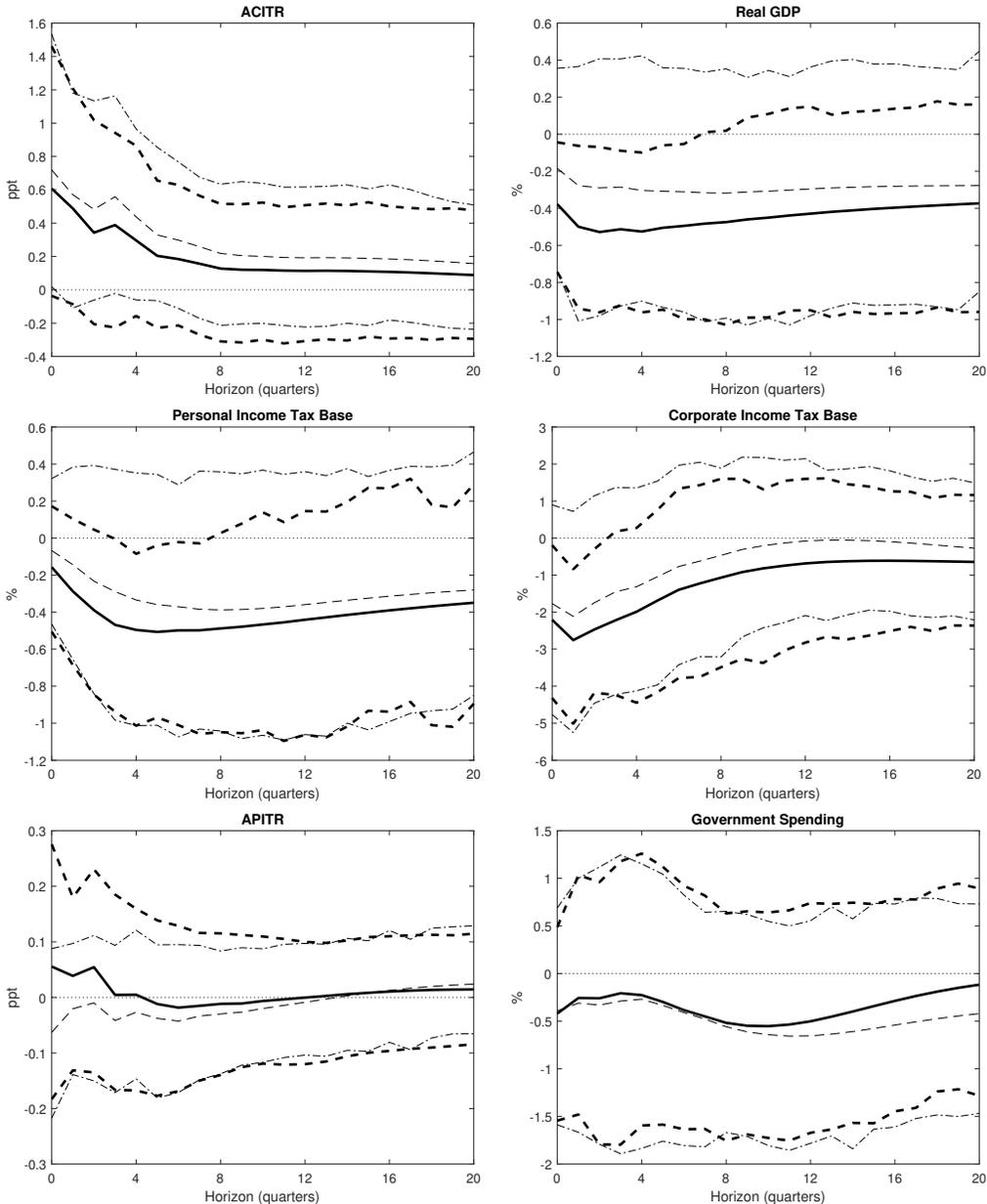# Figure 1: Impulse Responses to $\varepsilon_{APITR,t}$ Under Different Identifying Restrictions (Single Prior)
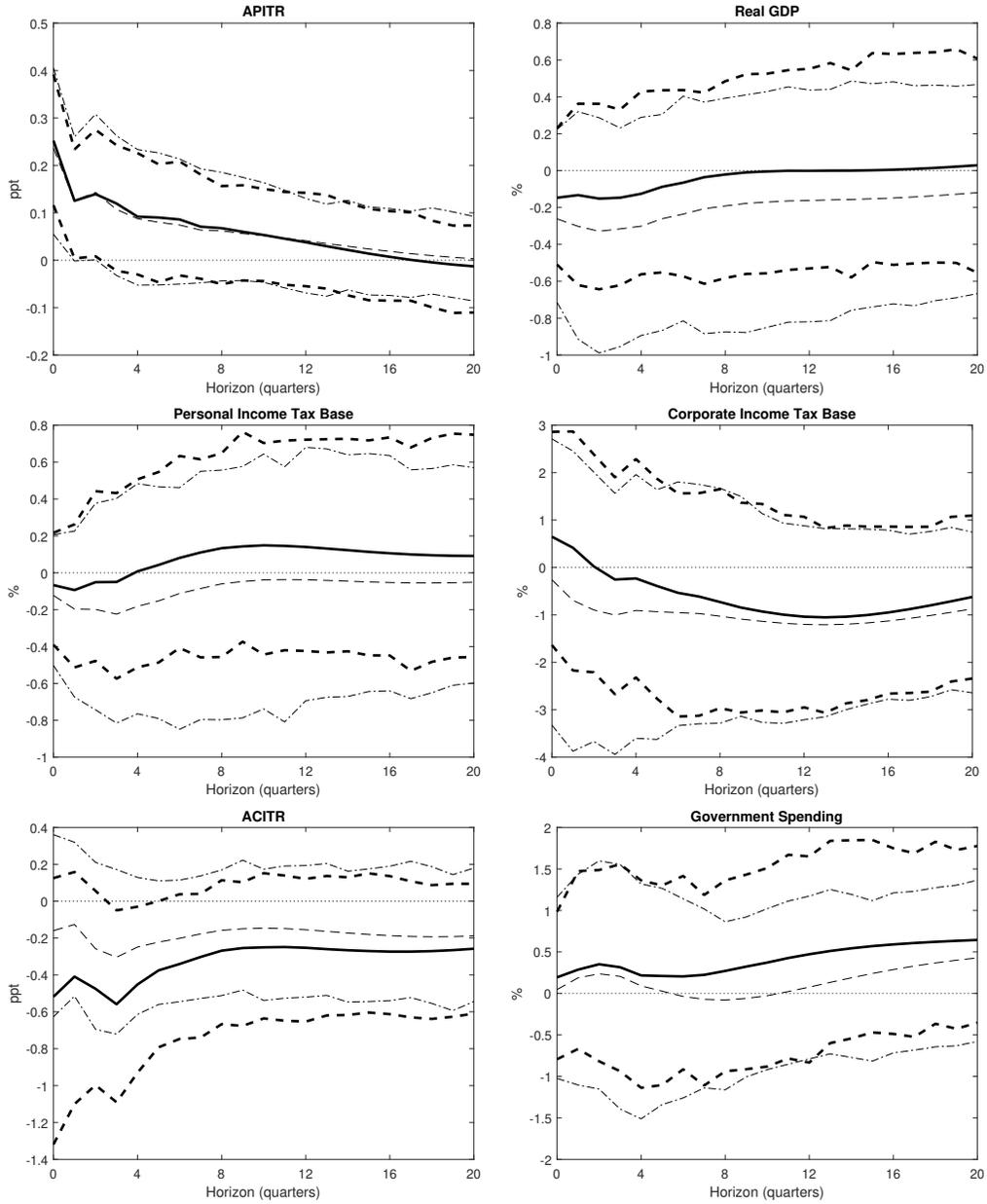
Lighter lines represent the posterior mean and 90 per cent highest posterior density intervals under the point-identifying assumptions in MR13; darker lines represent the posterior mean and 90 per cent highest posterior density intervals under our set-identifying restrictions and given the single prior for $\mathbf{Q}|\phi$; impulse responses are to a one standard deviation shock.

Figure 2: Impulse Responses to $\varepsilon_{APITR,t}$ Under Set-Identifying Restrictions (Single and Multiple Priors)
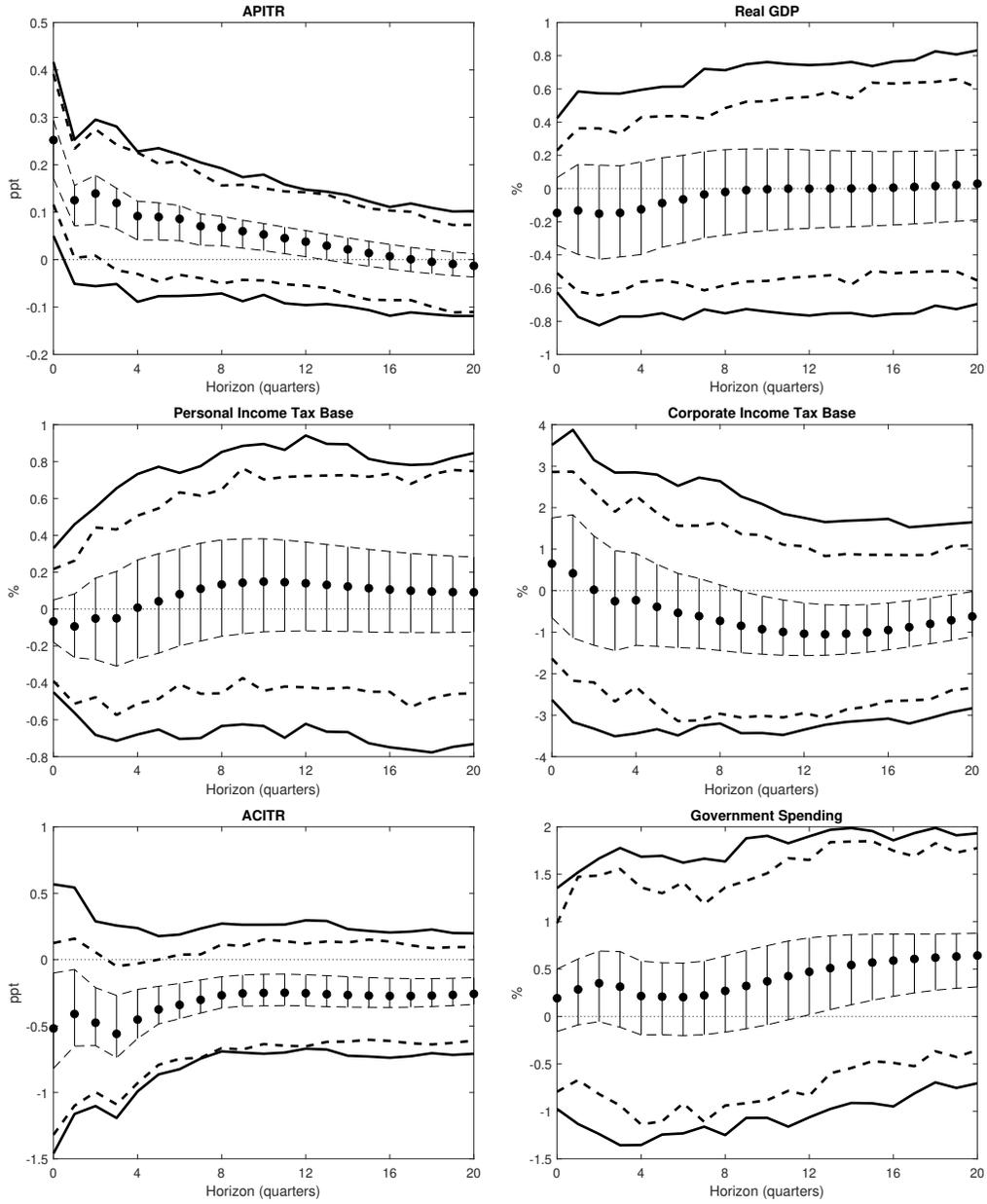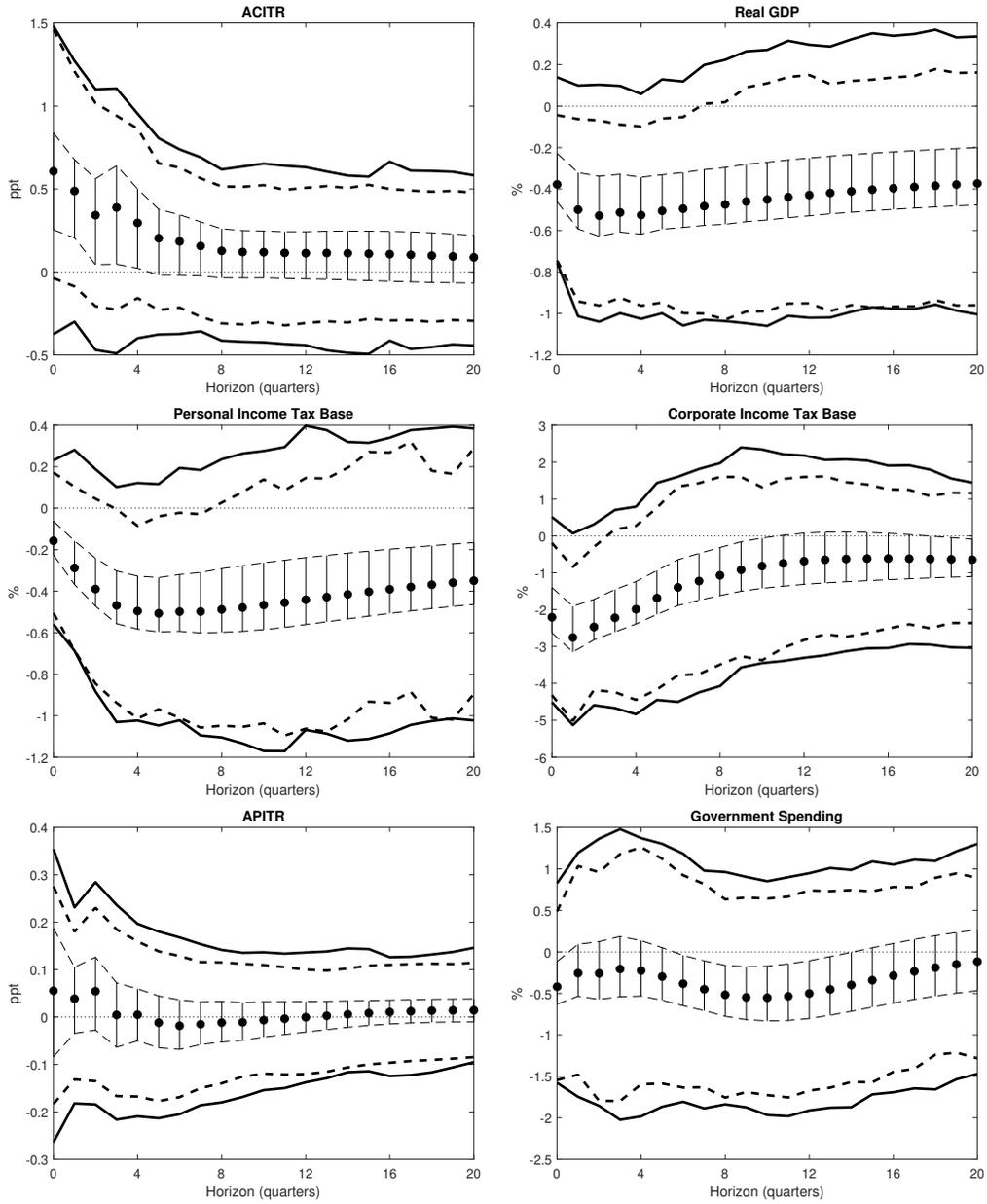

Notes: Circles and dashed lines are, respectively, posterior means and 90 per cent highest posterior density intervals under the single prior for $\mathbf{Q}|\boldsymbol{\phi}$; vertical bars are posterior mean bounds and solid lines are 90 per cent robust credible regions obtained using Algorithm 2 with $\tilde{K} = 10,000$ and 1,000 draws from the posterior of ϕ with non-empty identified set; impulse responses are to a one standard deviation shock.

Figure 3: Impulse Responses to $\varepsilon_{ACITR,t}$ Under Different Identifying Restrictions (Single Prior)
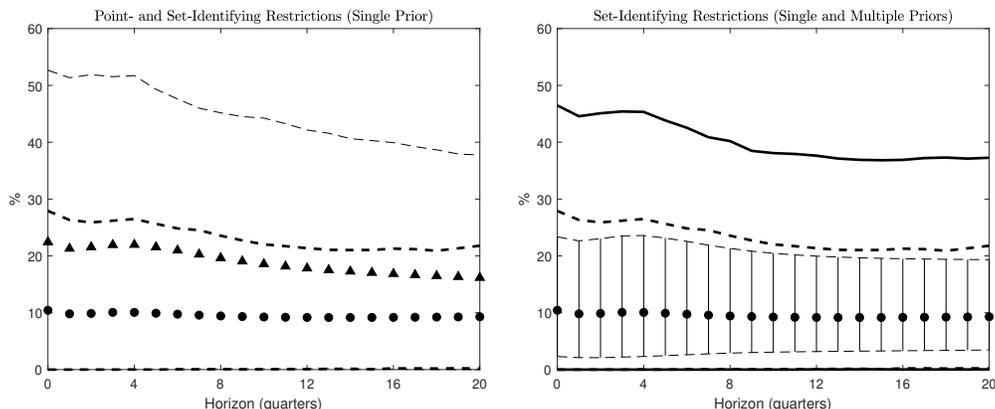

Notes: See notes to Figure 1.

Figure 4: Impulse Responses to $\varepsilon_{ACITR,t}$ Under Set-Identifying Restrictions (Single and Multiple Priors)
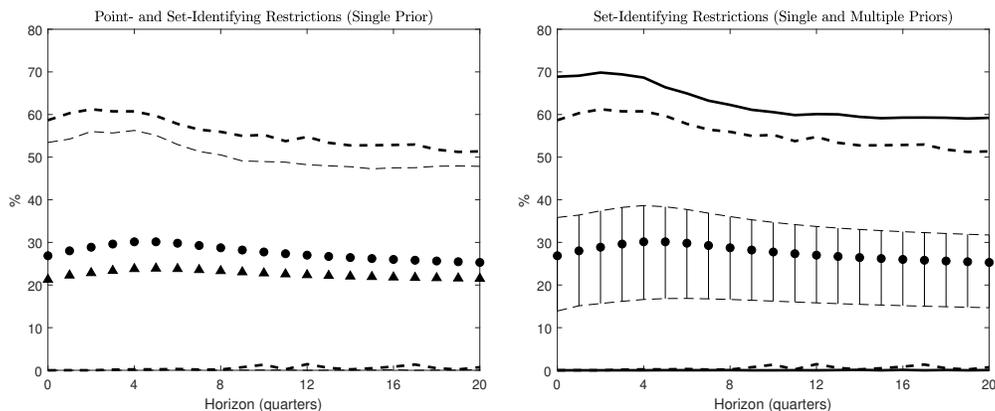

Notes: See notes to Figure 2.

Figure 5: Contribution of $\varepsilon_{APITR,t}$ to Forecast Error Variance of Real GDP

Notes: Circles and darker dashed lines are, respectively, posterior means and 90 per cent highest posterior density intervals under the set-identifying restrictions and the single prior for $\mathbf{Q}|\boldsymbol{\phi}$; triangles and lighter dashed lines are, respectively, posterior means and 90 per cent highest posterior density intervals under the point-identifying restrictions; vertical bars are posterior mean bounds and solid lines are 90 per cent robust credible regions obtained using Algorithm 2.

Figure 6: Contribution of $\varepsilon_{ACITR,t}$ to Forecast Error Variance of Real GDP

Notes: See notes to Figure 5.