Narrative Restrictions and Proxies*

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

5 January 2022

Abstract

We compare two approaches to inference in structural vector autoregressions given information about the signs of structural shocks at specific dates: a likelihood-based approach that imposes 'narrative restrictions' (NR) on the shock signs and a 'narrative-proxy' (NP) approach that casts the information about the shock signs as proxies for the shocks. When the number of NR is fixed, the robust Bayesian approach to inference under NR described in Giacomini, Kitagawa and Read (2021) delivers robust credible intervals with valid asymptotic frequentist coverage of the true impulse responses. In contrast, under the NP approach, the assumptions for validity of the weak-proxy robust confidence intervals in Montiel Olea, Stock and Watson (2021) are violated. A Monte Carlo exercise suggests that the weak-proxy robust confidence intervals have incorrect coverage unless the sign of the shock is known in a large number of periods, whereas the robust credible intervals under the NR approach always display coverage exceeding the nominal level.

*We thank Mikkel Plagborg-Möller, James Stock, Christian Wolf and seminar participants at various venues for discussions that motivated this work. We gratefully acknowledge financial support from ERC grants (numbers 536284 and 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001). The views in this paper are the authors' and do not reflect the views of the Federal Reserve Bank of Chicago, the Federal Reserve system or the Reserve Bank of Australia.

†University College London, Department of Economics and Federal Reserve Bank of Chicago. Email: r.giacomini@ucl.ac.uk

‡Brown University, Department of Economics and University College London, Department of Economics. Email: toru_kitagawa@brown.edu

§University College London, Department of Economics and Reserve Bank of Australia. Email: readm@rba.gov.au

1 Introduction

Structural vector autoregressions (SVARs) are used in macroeconomics to estimate the dynamic causal effects of structural shocks. A common approach to identifying the effects of these shocks is to impose a set of sign and/or zero restrictions on functions of the SVAR's structural parameters that together result in set-identification of the parameters of interest (e.g. sign restrictions on impulse responses, as in Uhlig (2005)). A growing number of papers augment these 'traditional' set-identifying restrictions with restrictions that involve the values of the structural shocks in specific periods. For instance, Antolín-Díaz and Rubio-Ramírez (2018) propose restricting the signs of structural shocks based on the historical narrative about the nature of the structural shocks hitting the economy in particular episodes. An example is the restriction that there was a positive monetary policy shock in the United States in October 1979, which is the year in which the Federal Reserve dramatically raised the federal funds rate following Paul Volcker becoming chairman.[1]

In this paper, we compare two alternative approaches to using information about the signs of structural shocks in specific periods in an SVAR framework. The first approach is to impose the information as restrictions on the signs of the structural shocks in a likelihood-based framework, which we refer to as 'shock-sign narrative restrictions' (NR). This is the approach proposed in Antolín-Díaz and Rubio-Ramírez (2018). The second approach follows a suggestion in Plagborg-Møller and Wolf (2021b) that the information about the shock signs could be recast as an external instrument or 'proxy' for use in a proxy SVAR (e.g. Mertens and Ravn 2013; Stock and Watson 2018). We refer to this as the 'narrative-proxy' (NP) approach. We compare the two approaches through the lens of a simple bivariate SVAR with no dynamics, which makes the trade-offs between the two approaches clear and allows us to leverage analytical results.

Following the approach proposed in Antolín-Díaz and Rubio-Ramírez (2018), the literature that makes use of NR typically imposes these restrictions within a Bayesian framework.[2] This involves specifying a uniform-normal-inverse-Wishart prior over

[1] Other papers that impose NR include Ben Zeev (2018), Ludvigson, Ma and Ng (2018, 2021), Furlanetto and Robstad (2019), Cheng and Yang (2020), Kilian and Zhou (2020a, 2020b), Laumer (2020), Redl (2020), Zhou (2020) and Inoue and Kilian (in press).

[2] An exception is Ludvigson et al. (2018), who use a bootstrap to conduct inference. However, the frequentist validity of this bootstrap is unknown.

the orthogonal reduced-form parameterisation of the SVAR, which is the standard approach in set-identified SVARs (e.g. Arias et al. 2018). However, Giacomini, Kitagawa and Read (2021) – henceforth, GKR – point out some undesirable features of this approach. Under shock-sign restrictions, the likelihood function possesses flat regions, which implies that a component of the prior is never updated by the data. Furthermore, as discussed by Baumeister and Hamilton (2015) in the context of standard sign restrictions, the prior is also informative about parameters of interest, such as impulse responses. Given that the prior is primarily chosen for computational convenience and does not reflect credible prior information about the parameters, these features of standard Bayesian inference under NR raise the concern that posterior inference may be sensitive to the choice of prior.[3]

To address the issue of posterior sensitivity to the choice of prior under NR, GKR propose applying a variant of the 'robust' (multiple-prior) Bayesian approach to inference for set-identified models developed in Giacomini and Kitagawa (2021). This involves replacing the unrevisable component of the prior with a class of priors that are consistent with the identifying restrictions. The class of priors generates a class of posteriors, which can be summarised in various ways. For example, rather than having a single posterior mean, there is a *set* of posterior means, which is an interval containing every possible posterior mean that could be obtained under the class of priors. One can also report a 'robust credible interval', which is an interval that receives at least a given posterior probability under every posterior in the class of posteriors. This approach eliminates the source of posterior sensitivity arising due to the unrevisable component of the prior. GKR also provide conditions under which the robust credible intervals have valid frequentist coverage of the true impulse response asymptotically.

As noted in Plagborg-Møller and Wolf (2021b), an alternative way to use information about the sign of a particular structural shock in specific periods is to recast this information as a discrete-valued proxy for the structural shock. Specifically, one

[3]An additional problem arises under more-general classes of NR, such as restrictions on the historical decomposition. In constructing the posterior, Antolín-Díaz and Rubio-Ramírez (2018) use the *conditional* likelihood, which is the likelihood function conditional on the NR holding. GKR show that the use of this likelihood distorts the posterior distribution towards parameter values that result in a lower *ex ante* probability that the NR are satisfied whenever this probability depends on the structural parameters. To address this problem, they propose using the *unconditional* likelihood to construct the posterior. This problem does not arise under shock-sign restrictions, because the conditional and unconditional likelihoods are identical (up to a multiplicative constant).

can construct a variable that is equal to one in periods where the shock is known to be positive, minus one in periods where it is known to be negative, and zero in all other periods. This variable will clearly be positively correlated with the underlying structural shock (i.e. 'relevant') and uncorrelated with all other structural shocks (i.e. 'exogenous'). The variable is therefore a valid proxy for the structural shock and can be used to point-identify impulse responses to that shock in a proxy SVAR (Mertens and Ravn 2013; Stock and Watson 2018).[4]

Montiel Olea, Stock and Watson (2021) – henceforth, MSW – show that frequentist inference about impulse responses is nonstandard when the proxy variable is only weakly correlated with the target structural shock (see also Lunsford (2015)). They propose a weak-proxy robust approach to inference, which – under some conditions – has asymptotically valid coverage of the true impulse response regardless of the strength of the correlation between the proxy and shock. We show that the narrative proxy is likely to be weakly correlated with the target structural shock when there are only a small number of periods in which the sign of the shock is known. Accordingly, the weak-proxy robust approach to inference is the natural approach to frequentist inference when using narrative proxies.

Using an analytically convenient bivariate example, we compare the two approaches to using information about shock signs. First, we assume that the reduced-form parameters are known and contrast the 'conditional identified sets' for the impulse responses that are obtained under the two approaches when there is a single shock-sign NR or a NP that takes value one in a single period. The conditional identified set is the set of values of the impulse response that are consistent with the reduced-form parameters and the imposed restrictions (i.e. the shock-sign NR or the restriction that the target structural shock is uncorrelated with the NP).[5]

Under the NR approach, the conditional identified set is guaranteed to contain the true impulse response when the reduced-form parameters are equal to their true values. Under the NP approach, the conditional identified set is a singleton that depends on the realisations of the shocks in the single period in which the NP is

[4]The variable is also uncorrelated with leads and lags of the structural shocks, so it could alternatively be used to point-identify the impulse responses to that shock in an instrumental-variables local projection, which does not require assuming that the structural shock is invertible (Stock and Watson 2018).

[5]The concept of a conditional identified set differs to that of a standard identified set. The latter is defined by a (set-valued) mapping from reduced-form to structural parameters, while the former additionally depends on the realisation of the data in particular periods. See GKR for further details.

non-zero. In particular, the conditional identified set for an impulse response to the target shock is contaminated by the impulse response to the non-target shock unless the non-target shock is exactly equal to zero. In fact, we show that the conditional identified set under the NP coincides with the conditional identified set that would be obtained under the 'narrative zero restriction' that the realisation of the second shock is equal to zero in the period in which the sign of the first shock is known. When the realisation of the non-target shock is large relative to the target shock, the conditional identified set for an impulse response to the target shock obtained using the NP is approximately equal to the true impulse response to the non-target shock. This suggests that NP-based estimates may be sensitive to the realisation of non-target shocks in the periods in which the proxy is non-zero.

Second, we discuss the theoretical properties of the inferential procedures described above when there is a single shock-sign NR or a single NP. Using results from GKR, we argue that the robust credible region for an impulse response will have asymptotically valid frequentist coverage of the true impulse response under a single shock-sign NR. We then discuss whether the assumptions required for the asymptotic validity of the weak-proxy robust confidence intervals in MSW are satisfied when there are NP. One of these assumptions is that the covariance between the proxy and the reduced-form VAR innovations is $\sqrt{T}$-asymptotically normal. We describe conditions under which this assumption will be violated under NP. Consequently, the frequentist validity of the MSW confidence intervals is not guaranteed in this setting.

Third, we provide Monte Carlo evidence in support of our theoretical results.

To the best of our knowledge, this is the first paper to discuss whether information about structural shocks should be used to impose NR in an otherwise set-identified SVAR or to construct a proxy for use in a proxy SVAR. Boer and Lütkepohl (in press) compare the efficiency of proxy SVAR estimators that use only the signs of structural shocks on particular dates against proxies that also use information about the size of the shock (i.e. standard proxies). Based on Monte Carlo simulations, they conclude that the proxy SVAR estimator that uses information only about the sign of the shock is nearly as (or, in some cases, more) efficient than the estimator based on the quantitative information. Budnik and Rünstler (2020) consider identification of impulse responses in Bayesian proxy SVARs when the proxy represents the sign of a certain structural shock in particular periods (i.e. what we call narrative proxies). Their approach to identification departs from the standard proxy SVAR setting and is implemented using linear discriminant analysis or a non-parametric sign-concordance criterion.

2 Stylised SVAR

This section introduces a stylised bivariate SVAR that we use as a laboratory for considering the two approaches to using information about the sign of a structural shock. We employ this model because it allows us to derive convenient analytical expressions.

Consider the SVAR(0) for the 2×1 vector of endogenous variables $\mathbf{y}_t = (y_{1t}, y_{2t})'$:

$$\mathbf{A}_0 \mathbf{y}_t = \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ with $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$. We abstract from dynamics for ease of exposition, but this is without loss of generality. The orthogonal reduced form of the model reparameterizes $\mathbf{A}_0^{-1}$ as $\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor (with positive diagonal elements) of $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{y}_t\mathbf{y}_t') = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$. We parameterize $\boldsymbol{\Sigma}_{tr}$ directly as

$$\boldsymbol{\Sigma}_{tr} = \begin{bmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \quad (\sigma_{11}, \sigma_{22} > 0), \tag{2}$$

and denote the vector of reduced-form parameters as $\boldsymbol{\phi} = \text{vech}(\boldsymbol{\Sigma}_{tr})$. $\mathbf{Q}$ is an orthonormal matrix in the space of 2×2 orthonormal matrices, $\mathcal{O}(2)$:

$$\mathbf{Q} \in \mathcal{O}(2) = \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\} \cup \left\{ \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}, \tag{3}$$

where the first set is the set of 'rotation' matrices and the second set is the set of 'reflection' matrices.

In the absence of any identifying restrictions, the identified set for $\mathbf{A}_0^{-1}$ (the matrix of contemporaneous impulse responses) is

$$\mathbf{A}_0^{-1} \in \left\{ \begin{bmatrix} \sigma_{11}\cos\theta & -\sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{22}\cos\theta - \sigma_{21}\sin\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}$$

$$\cup \left\{ \begin{bmatrix} \sigma_{11}\cos\theta & \sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{21}\sin\theta - \sigma_{22}\cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}. \quad (4)$$

The identified set for $\mathbf{A}_0$ is then

$$\mathbf{A}_0 \in \left\{ \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ -\sigma_{21}\cos\theta - \sigma_{22}\sin\theta & \sigma_{11}\cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}$$
$$\cup \left\{ \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ \sigma_{22}\sin\theta + \sigma_{21}\cos\theta & -\sigma_{11}\cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}. \quad (5)$$

Henceforth, we leave implicit that $\theta \in [-\pi, \pi]$ and we impose the sign normalisation $\text{diag}(\mathbf{A}_0) \geq \mathbf{0}_{2\times 1}$, which is a normalisation on the signs of the structural shocks.

The impulse response of the second variable to a shock to the first variable that raises the first variable by one unit is

$$\tilde{\eta}_{21} = \frac{\sigma_{21}\cos\theta + \sigma_{22}\sin\theta}{\sigma_{11}\cos\theta} = \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}}\tan\theta. \quad (6)$$

In what follows, we assume that this is the parameter of interest.

3 Conditional Identified Sets

This section derives the 'conditional identified set' for $\tilde{\eta}_{21}$ under the two approaches to identification.

3.1 Shock-sign narrative restriction

Consider the shock-sign restriction that ε_{1k} is nonnegative for some $k \in \{1, \ldots, T\}$:

$$\varepsilon_{1k} = \mathbf{e}_{1,2}'\mathbf{A}_0\mathbf{y}_k = (\sigma_{11}\sigma_{22})^{-1}\left(\sigma_{22}y_{1k}\cos\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\sin\theta\right) \geq 0. \quad (7)$$

Under the sign normalization and the shock-sign restriction, θ is restricted to the set

$$\theta \in \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \geq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}$$
$$\cup \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \leq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}. \quad (8)$$

Since y_{1k} and y_{2k} enter the inequalities characterising this set, the shock-sign restriction induces a set-valued mapping from $\boldsymbol{\phi}$ to θ that depends on the realization of $\mathbf{y}_k$; GKR refer to this mapping as the conditional identified set. GKR derive an analytical characterisation of the conditional identified set for θ under this restriction. The exact expression for this set depends on the signs of σ_{21} and the realisations of the data. We draw on their results to derive the corresponding conditional identified set for $\tilde{\eta}_{21}$.

In what follows, let $h(\boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{21}y_{1k} - \sigma_{11}y_{2k}$ and $C(\boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{22}y_{1k}/h(\boldsymbol{\phi}, \mathbf{y}_k)$. There are four main cases to consider depending on the signs of $h(\boldsymbol{\phi}, \mathbf{y}_k)$ and $C(\boldsymbol{\phi}, \mathbf{y}_k)$.

First, consider the case where $\sigma_{21} < 0$ and $h(\boldsymbol{\phi}, \mathbf{y}_k) < 0$. In this case, the conditional identified set for θ is

$$\theta \in \left[\arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, C(\boldsymbol{\phi}, \mathbf{y}_k)\right\}\right), \pi + \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, C(\boldsymbol{\phi}, \mathbf{y}_k)\right\}\right)\right]. \quad (9)$$

Since $-\pi/2 < \arctan(x) < \pi/2$, the conditional identified set for θ is a closed interval that contains the point $\pi/2$. For $\theta \nearrow \pi/2$, $\tan\theta \to \infty$, so $\tilde{\eta}_{21} \to \infty$. For $\theta \searrow \pi/2$, $\tan\theta \to -\infty$, so $\tilde{\eta}_{21} \to -\infty$. Hence,

$$\tilde{\eta}_{21} \in (-\infty, \infty). \quad (10)$$

Next, consider the case where $\sigma_{21} < 0$ and $h(\boldsymbol{\phi}, \mathbf{y}_k) > 0$. If $y_{1k} > 0$ or if $y_{1k} < 0$ and $\sigma_{22}/\sigma_{21} < C(\boldsymbol{\phi}, \mathbf{y}_k)$,

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan(C(\boldsymbol{\phi}, \mathbf{y}_k))\right]. \quad (11)$$

Since $\tan\theta$ is a strictly increasing function over this interval, it immediately follows that

$$\tilde{\eta}_{21} \in \left[\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}^2}{\sigma_{11}\sigma_{21}}, \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}}C(\boldsymbol{\phi}, \mathbf{y}_k)\right]. \quad (12)$$

If $y_{1k} < 0$ and $\sigma_{22}/\sigma_{21} > C(\boldsymbol{\phi}, \mathbf{y}_k)$,

$$\theta \in \left[\pi + \arctan(C(\boldsymbol{\phi}, \mathbf{y}_k)), \pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]. \quad (13)$$

Since $\tan\theta$ is a strictly increasing function over this interval, it immediately follows that

$$\tilde{\eta}_{21} \in \left[\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}} C(\boldsymbol{\phi}, \mathbf{y}_k), \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}^2}{\sigma_{11}\sigma_{21}}\right]. \tag{14}$$

The next case is where $\sigma_{21} > 0$ and $h(\boldsymbol{\phi}, \mathbf{y}_k) < 0$. If $y_{1k} > 0$ or if $y_{1k} < 0$ and $\sigma_{22}/\sigma_{21} > C(\boldsymbol{\phi}, \mathbf{y}_k)$,

$$\theta \in \left[\arctan(C(\boldsymbol{\phi}, \mathbf{y}_k)), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]. \tag{15}$$

Since $\tan\theta$ is a strictly increasing function over this interval, it immediately follows that

$$\tilde{\eta}_{21} \in \left[\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}} C(\boldsymbol{\phi}, \mathbf{y}_k), \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}^2}{\sigma_{11}\sigma_{21}}\right]. \tag{16}$$

If $y_{1k} < 0$ and $\sigma_{22}/\sigma_{21} < C(\boldsymbol{\phi}, \mathbf{y}_k)$,

$$\theta \in \left[-\pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), -\pi + \arctan(C(\boldsymbol{\phi}, \mathbf{y}_k))\right]. \tag{17}$$

Since $\tan\theta$ is a strictly increasing function over this interval, it immediately follows that

$$\tilde{\eta}_{21} \in \left[\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}^2}{\sigma_{11}\sigma_{21}}, \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}} C(\boldsymbol{\phi}, \mathbf{y}_k)\right]. \tag{18}$$

Finally, if $\sigma_{21} > 0$ and $h(\boldsymbol{\phi}, \mathbf{y}_k) > 0$,

$$\theta \in \left[-\pi + \arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, C(\boldsymbol{\phi}, \mathbf{y}_k)\right\}\right), \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, C(\boldsymbol{\phi}, \mathbf{y}_k)\right\}\right)\right]. \tag{19}$$

Since $-\pi/2 < \arctan(x) < \pi/2$, the conditional identified set for θ is a closed interval that contains the point $-\pi/2$. For $\theta \nearrow -\pi/2$, $\tan\theta \to \infty$, so $\tilde{\eta}_{21} \to \infty$. For $\theta \searrow -\pi/2$, $\tan\theta \to -\infty$, so $\tilde{\eta}_{21} \to -\infty$. Hence,

$$\tilde{\eta}_{21} \in (-\infty, \infty). \tag{20}$$

Remarks. Consider setting the parameters $(\boldsymbol{\phi}, \theta)$ equal to their true values, $(\boldsymbol{\phi}_0, \theta_0)$. In this case, the conditional identified set for $\tilde{\eta}_{21}$ always includes the true value of the impulse response when the shock-sign restriction is correct. This may be considered an attractive feature of imposing shock-sign restrictions; in the limiting case where the reduced-form parameters are known with certainty, the set of impulse responses obtained is guaranteed to include the true value of the impulse response.

Another feature of these conditional identified sets is that they may be unbounded depending on the value of the reduced-form parameters and the realisation of $\mathbf{y}_k$. As a consequence, the set of posterior means and robust credible intervals under the robust Bayesian approach to inference may also be unbounded. Rather than being a drawback of this approach, one could argue that it is a strength, since it transparently reflects the information contained in the data (through the reduced-form parameters) and the identifying restrictions about the parameter of interest.

3.2 Narrative proxy

Assume there is a variable Z_t satisfying $\mathbb{E}(Z_t\varepsilon_{1t}) \neq 0$ and $\mathbb{E}(Z_t\varepsilon_{2t}) = 0$. After expressing ε_{2t} in terms of $\mathbf{y}_t$ and the parameters, the exogeneity condition implies that

$$\sigma_{22}\mathbb{E}(Z_t y_{1t})\sin\theta = \mathbb{E}(Z_t(\sigma_{11}y_{2t} - \sigma_{21}y_{1t}))\cos\theta. \tag{21}$$

If the instrument is not relevant, so that $\mathbb{E}(Z_t\varepsilon_{1t}) = 0$, the restriction $\mathbb{E}(Z_t\varepsilon_{2t}) = 0$ carries no information about θ, since $\mathbb{E}(Z_t y_{1t}) = 0$ and $\mathbb{E}(Z_t(\sigma_{11}y_{2t} - \sigma_{21}y_{1t})) = 0$. Otherwise,

$$\tan\theta = \frac{\mathbb{E}(Z_t(\sigma_{11}y_{2t} - \sigma_{21}y_{1t}))}{\sigma_{22}\mathbb{E}(Z_t y_{1t})}. \tag{22}$$

This equation has two solutions in $[-\pi, \pi]$, one of which will be ruled out by the sign normalization restrictions. For example, if $\sigma_{21} < 0$ and the term on the right-hand side of Equation (22) (henceforth denoted by C) is positive, then θ is either equal to $\arctan(C) - \pi$ or $\arctan(C)$. The sign normalization implies that $\theta \in [\arctan(\sigma_{22}/\sigma_{21}), \arctan(\sigma_{22}/\sigma_{21}) + \pi]$, which rules out the first solution, so $\theta = \arctan(C)$. If C is negative, then θ is either equal to $\arctan(C)$ or $\arctan(C) + \pi$. If $C > \sigma_{22}/\sigma_{21}$, then the sign normalization selects the first solution, otherwise it selects the second solution. Similar arguments apply when $\sigma_{21} > 0$.[6]

Consider the case where information about the sign of the first structural shock is recast as a binary variable. Specifically, as in the shock-sign example, assume the econometrician knows that $\varepsilon_{1k} \geq 0$ for some $k \in \{1, \ldots, T\}$, and let $Z_k = \text{sgn}(\varepsilon_{1k})$

[6] When $\sigma_{21} > 0$, the sign normalization restricts θ to lie in $\left[-\pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]$. If $C > \frac{\sigma_{22}}{\sigma_{21}}$, the sign normalization implies that $\theta = \arctan(C) - \pi$. Otherwise, the sign normalization implies that $\theta = \arctan(C)$.

with $Z_t = 0$ for $t \neq k$. What happens if the econometrician imposes the identifying restriction that $\mathbb{E}(Z_t \varepsilon_{2t}) = 0$?

Maintaining the assumption that $\boldsymbol{\phi}$ is known with $\sigma_{21} < 0$, in the case where $(\sigma_{11} y_{2k} - \sigma_{21} y_{1k}) > 0$ and $y_{1k} > 0$, an analogue estimator of θ is

$$\begin{aligned}\hat{\theta} &= \arctan\left(\frac{\frac{1}{T}\sum_{t=1}^{T} Z_t(\sigma_{11} y_{2t} - \sigma_{21} y_{1t})}{\sigma_{22}\frac{1}{T}\sum_{t=1}^{T} Z_t y_{1t}}\right) \\ &= \arctan\left(\frac{\sigma_{11} y_{2k} - \sigma_{21} y_{1k}}{\sigma_{22} y_{1k}}\right). \end{aligned} \tag{23}$$

Note that this is equal to the estimator that would be obtained if one were to impose the 'narrative zero restriction' $\varepsilon_{2k} = 0$.[7]

How does this estimator relate to the true value of θ? Assume that the data are generated by a process with parameter $\theta_0 \in (0, \frac{\pi}{2})$ (with $\mathbf{Q}$ equal to the rotation matrix). Replacing y_{1k} and y_{2k} in (23) using $\mathbf{y}_k = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_k$ yields an expression for $\hat{\theta}$ in terms of the true parameters and the underlying structural shocks:

$$\begin{aligned}\hat{\theta} = \arctan\Bigg(&\left(\sigma_{22}\left(\sigma_{11}\cos\theta_0\varepsilon_{1k} - \sigma_{11}\sin\theta_0\varepsilon_{2k}\right)\right)^{-1}\Big[\sigma_{11}\Big[\left(\sigma_{21}\cos\theta_0 + \sigma_{22}\sin\theta_0\right)\varepsilon_{1k} + \\ &\left(\sigma_{22}\cos\theta_0 - \sigma_{21}\sin\theta_0\right)\varepsilon_{2k}\Big] - \sigma_{21}\left(\sigma_{11}\cos\theta_0\varepsilon_{1k} - \sigma_{11}\sin\theta_0\varepsilon_{2k}\right)\Big]\Bigg). \end{aligned} \tag{25}$$

If $\varepsilon_{2k} = 0$, we have that $\hat{\theta} = \theta_0$. Otherwise, $\hat{\theta}$ will not in general coincide with θ_0. For example, for $\varepsilon_{2k} \neq 0$ and $\varepsilon_{1k} \approx 0$, $\hat{\theta} \approx \arctan(-\cot\theta_0)$.[8] In this case, the corresponding value of $\tilde{\eta}_{21}$ is

$$\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}}\tan\left(\arctan\left(-\cot(\theta_0)\right)\right) = \frac{\sigma_{21}}{\sigma_{11}} - \frac{\sigma_{22}}{\sigma_{11}}\cot\theta_0, \tag{26}$$

which is equal to the true impulse response of the second variable to the *second* shock such that the first variable increases by one unit. More generally, these results

[7] Under the narrative zero restriction $\varepsilon_{2k} = 0$, θ is restricted to lie in the set

$$\begin{aligned}\theta \in \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \geq 0, (\sigma_{21}\cos\theta + \sigma_{22}\sin\theta)y_{1k} = \sigma_{11}\cos\theta y_{2k}\} \\ \cup \{\sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \leq 0, (\sigma_{22}\sin\theta + \sigma_{21}\cos\theta)y_{1k} = \sigma_{11}\cos\theta y_{2k}\}. \end{aligned} \tag{24}$$

[8] This follows from the fact that $\arctan(x) + \arctan\left(\frac{1}{x}\right) = \frac{\pi}{2}$ for $x > 0$.

suggest that the estimator of the impulse response under NP may be sensitive to the realisations of the non-target shocks in the period in which the sign of the target shock is known.

The sensitivity of the estimator of $\tilde{\eta}_{21}$ to the realisation of the shocks is also immediately clear by considering the form of the estimator as a function of the data. Given the conditional identified set for θ, the conditional identified set for $\tilde{\eta}_{21}$ is

$$\frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}} \tan\left(\arctan\left(\frac{\sigma_{11}y_{2k} - \sigma_{21}y_{1k}}{\sigma_{22}y_{1k}}\right)\right) = \frac{y_{2k}}{y_{1k}}. \tag{27}$$

This expression coincides with the analogue estimator one would obtain based on the identification arguments in Stock and Watson (2016, 2018) when the proxy variable is a NP. The estimator (or conditional identified set) for the impulse response is simply the ratio of the realisations of the data in the period in which the sign of the shock is known. Importantly, this is the case even when the reduced-form parameters are known with certainty (i.e. given a time series of infinite length). Any estimate obtained in this setting would therefore be sensitive to the period-k realisation of the data or, equivalently, the shocks.

4 Frequentist Properties of Approaches to Inference

This section discusses the theoretical properties of the inferential procedures described above when there is a single shock-sign NR or a single narrative proxy. Building on results from GKR, we argue that the robust credible region for an impulse response will have asymptotically valid frequentist coverage of the true impulse response under a fixed number of shock-sign restrictions. We then discuss whether the asymptotic validity of the weak-proxy robust Anderson-Rubin (AR) confidence intervals shown in MSW holds or not under NP. One of the key assumptions required for the asymptotic-validity of the standard AR confidence intervals is that the sample covariance between the proxy and the reduced-form VAR innovations is $\sqrt{T}$-asymptotically normal. This assumption is violated under NP if the number of available shock sign restrictions are small relative to T. Consequently, the frequentist validity of the MSW confidence intervals is not guaranteed in this setting. Numerical analysis of the next section indeed shows that the standard AR confidence intervals have distorted coverage.

4.1 Robust Bayesian Inference Under NR

GKR show that their robust Bayesian approach to inference under NR is asymptotically valid from a frequentist perspective when the number of NR is fixed with the sample size. In particular, the robust credible interval has at least the nominal level of coverage for the true impulse response to a standard-deviation shock asymptotically under some high-level assumptions. These assumptions are that the conditional identified set for the impulse response is closed and convex, and has lower and upper bounds that are differentiable in $\boldsymbol{\phi}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ with non-zero derivatives.[9] In the current context, we are interested in the impulse response to a unit shock rather than a standard-deviation shock. This shift of parameter of interest gives rise an unbounded conditional identified set depending on $\boldsymbol{\phi}$ and the realization of $\mathbf{y}_k$, which complicates an application of the asymptotic validity claim of the robust Bayes credible region of GKR. A feasible modification to get around this complication is to first construct the robust credible region for θ and project it to the parameter of interest $\tilde{\eta}_{21}$

4.2 Weak-proxy Robust Inference Under NP

As discussed in MSW, standard approaches to inference (e.g. based on asymptotic normality of the reduced-form parameters and the delta-method) in proxy SVARs are invalid when the proxy is only weakly correlated with the target structural shock. This is likely to be the case in applications using NP. For example, under the assumption that the structural shocks are normally distributed and the sign of the shock is known in K periods, the expected value of the covariance between the proxy variable and the target structural shock is $(K/T)\sqrt{2/\pi}$, where T is the number of observations.[10] For K small relative to T, this covariance will be close to zero on average. Furthermore, for K fixed as T approaches infinity, this covariance will converge to zero at rate T, which is faster than the $\sqrt{T}$ rate of convergence typically considered under weak-instrument asymptotics.

[9] See Propositions B.1–B.3 in GKR provide primitive conditions under which these assumptions hold when there are shock-sign restrictions.

[10] Here the 'expected value' we refer to is the average over random realisations of a time series with fixed length T.

Given that the NP is likely to be weakly correlated with the target structural shock, it seems natural to turn to the weak-proxy robust approach to inference developed in MSW. Building on the Anderson-Rubin approach to weak-instrument robust inference in microeconometrics, they develop weak-proxy robust confidence intervals for the impulse responses in a proxy SVAR. MSW show that their weak-proxy robust confidence intervals have correct coverage of the true impulse response asymptotically under the assumption that the reduced-form parameters – appropriately centered and scaled – are normally distributed asymptotically. Importantly, this result does not depend on the strength of the covariance between the proxy and the target structural shock.

Define the vector of covariances between the NP and the data by $\mathbf{\Gamma}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_T(z_t\mathbf{y}_t)$, where $\mathbb{E}_T(.)$ is the expectation operator for fixed sample size (i.e. taking expectations over alternative realisations of a time series of length T). Since NP instrument $\{z_t : t = 1 \dots T\}$ is not weakly stationary if the periods that the shock sign restrictions are imposed are deterministic, covariace $\mathbb{E}_T(z_t\mathbf{y}_t)$ differs in t. For instance, if we impose a shock sign restriction only for the first K-th periods, then we have

$$\mathbb{E}_T(z_t\mathbf{y}_t) = \begin{cases} \mathbb{E}_T(\text{sign}(\varepsilon_{1t})\mathbf{y}_t) & \text{for } 1 \leq t \leq K, \\ 0, & \text{for } t > K, \end{cases}$$

and $\mathbf{\Gamma}_T = \frac{1}{T}\sum_{t=1}^{K}\mathbb{E}_T(\text{sign}(\varepsilon_{1t})\mathbf{y}_t)$.

Let $\hat{\mathbf{\Gamma}}_T$ denote a sample analogue estimator for $\mathbf{\Gamma}_T$. Assuming that we can know the signs of the first structural shocks for the first K periods, we have

$$\hat{\mathbf{\Gamma}}_T = \frac{1}{T}\sum_{t=1}^{T} z_t\mathbf{y}_t = \frac{1}{T}\sum_{t=1}^{K}\text{sign}(\varepsilon_{1t})\mathbf{y}_t. \tag{28}$$

Assumption 2 of MSW requires that the reduced-form covariance estimator $\sqrt{T}(\hat{\mathbf{\Gamma}}_T - \mathbf{\Gamma}_T)$ is normally distributed asymptotically. However, in the NP setting with fixed K,

$$\begin{aligned}\sqrt{T}(\hat{\mathbf{\Gamma}}_T - \mathbf{\Gamma}_T) &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\left[z_t\mathbf{y}_t - \mathbb{E}_T(z_t\mathbf{y}_t)\right]\right) \\ &= \frac{1}{\sqrt{T}}\mathbf{A}_0^{-1}\sum_{t=1}^{K}\left[\text{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t - \mathbb{E}_T(\text{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t)\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{T}} \mathbf{A}_0^{-1} \sum_{t=1}^{K} \left[\text{sign}(\varepsilon_{1t}) \boldsymbol{\varepsilon}_t - \frac{K}{T} \sqrt{\frac{2}{\pi}} \mathbf{A}_0^{-1} \mathbf{e}_1 \right] \\
&\to_p \mathbf{0}_{2\times 1} \quad \text{as } T \to \infty,
\end{aligned}
$$

where $\mathbf{e}_1$ is the first column of $\mathbf{I}_2$. The reduced-form covariance between the NP and the data therefore has a degenerate distribution asymptotically after rescaling by $\sqrt{T}$. Consequently, Assumption 2 in MSW is not satisfied. It is therefore unclear whether the weak-proxy robust confidence intervals treating $(z_t : t = 1, \ldots, T)$ as an instru will have valid coverage of the true impulse response asymptotically in the NP setting.

To examine the coverage property of the AR confidence intervals, express the parameter of interest $\tilde{\eta}_{21}$ in terms of $\boldsymbol{\Gamma}_T$. Combining (6) and (22), $\tilde{\eta}_{21}$ is identified by the ratio of the second element of $\boldsymbol{\Gamma}_T$ to the first element of $\boldsymbol{\Gamma}_T$,

$$
\tilde{\eta}_{21} = \frac{T^{-1} \sum_{t=1}^{T} \mathbb{E}_T(z_t y_{2t})}{T^{-1} \sum_{t=1}^{T} \mathbb{E}_T(z_t y_{1t})} = \frac{T^{-1} \sum_{t=1}^{K} \mathbb{E}_T(\text{sign}(\varepsilon_{1t}) y_{2t})}{T^{-1} \sum_{t=1}^{K} \mathbb{E}_T(\text{sign}(\varepsilon_{1t}) y_{1t})}. \tag{29}
$$

Following MSW, we view the AR confidence intervals as inference for the ratio of the two means based on their unbiased statistics $\hat{\boldsymbol{\Gamma}}_T = \frac{1}{T} \cdot \sum_{t=1}^{K} \text{sign}(\varepsilon_{1t}) \mathbf{y}_t$ whose elements we denote by $(\hat{\gamma}_{1T}, \hat{\gamma}_{2T})'$. As discussed above, with fixed K, $\sqrt{T}(\hat{\boldsymbol{\Gamma}}_T - \boldsymbol{\Gamma}_T)$ is not asymptotically normal. In addition, even with the Gaussian specification for structural shocks, the distribution of $\hat{\boldsymbol{\Gamma}}_T$ is not jointly normal due to the multiplier term of $\text{sign}(\varepsilon_{1t})$.

The standard AR confidence intervals are constructed by inverting the Wald χ^2 test statistic. Let $\tilde{\eta}_{21}$ be the true impulse response specified under the null and $\hat{\boldsymbol{\Sigma}}_{tr}$ be a consistent estimator for $\boldsymbol{\Sigma}_{tr}$ (e.g. the maximum likelihood estimator). Given $\tilde{\eta}_{21}$ at the null and $\hat{\boldsymbol{\Sigma}}_{tr}$, let $\tilde{\mathbf{Q}}$ be the orthonormal matrix pinned down by the value of θ be the value of θ solving (6) with $\hat{\boldsymbol{\Sigma}}_{tr}$ plugged in and under the sign normalizations. We accordingly define $\hat{\mathbf{A}}_0^{-1} \equiv \hat{\boldsymbol{\Sigma}}_{tr} \hat{\mathbf{Q}}$.

If we naively apply the weak instrument robust AR confidence intervals, the Wald statistic for the tests to be inverted is constructed by

$$
W_T(\tilde{\eta}_{21}) \equiv \frac{T \left(\hat{\gamma}_{2T} - \tilde{\eta}_{21} \hat{\gamma}_{1T} \right)^2}{\begin{pmatrix} -\tilde{\eta}_{21} & 1 \end{pmatrix} \hat{\boldsymbol{\Omega}}_T \begin{pmatrix} -\tilde{\eta}_{21} & 1 \end{pmatrix}'}, \tag{30}
$$

where $\hat{\mathbf{\Omega}}_T$ is the sample variance-covariance matrix of $(z_t \mathbf{y}_t : t = 1, \ldots, T)$,

$$\begin{aligned}\hat{\mathbf{\Omega}}_T &= \frac{1}{T}\sum_{t=1}^{T}(z_t\mathbf{y}_t - \hat{\mathbf{\Gamma}}_T)(z_t\mathbf{y}_t - \hat{\mathbf{\Gamma}}_T)' \\ &= \mathbf{A}_0^{-1}\left[\frac{1}{T}\sum_{t=1}^{K}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \left(\frac{1}{T}\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right)\left(\frac{1}{T}\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right)'\right](\mathbf{A}_0^{-1})'. \quad (31)\end{aligned}$$

Noting

$$T^2\left(\hat{\gamma}_{2T} - \tilde{\eta}_{21}\hat{\gamma}_{1T}\right)^2 = \begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}\mathbf{A}_0^{-1}\left[\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right]\left[\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right]'(\mathbf{A}_0^{-1})'\begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}', \quad (32)$$

the Wald statistic (30) can be expressed as

$$W_T(\tilde{\eta}_{21}) = \frac{\begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}\mathbf{A}_0^{-1}\left[\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right]\left[\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right]'(\mathbf{A}_0^{-1})'\begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}'}{\begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}\mathbf{A}_0^{-1}\left[\sum_{t=1}^{K}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{K}\operatorname{sign}(\varepsilon_{1t})\boldsymbol{\varepsilon}_t\right)'\right](\mathbf{A}_0^{-1})'\begin{pmatrix}-\tilde{\eta}_{21} & 1\end{pmatrix}'} \quad (33)$$

This equivalent expression of the Wald statistic shows that $W_T(\tilde{\eta}_{21})$ indeed does not follow the χ^2 distribution with the degree of freedom 1 under the null even asymptotically as $T \to \infty$ with fixed K. First, the numerator term is not a squared Gaussian random variable when $K \geq 2$. In the special case of $K = 1$, $(\operatorname{sign}(\varepsilon_{11}))^2 = 1$ leads to the numerator term of (33) to be squared Gaussian. Second, the denominator term remains random even asymptotically and does not converge in probability to the variance covariance matrix of the numerator random variable under the null. In the special case of $K = 1$, the denominator is $\frac{T-1}{T}$ times the numerator term, so the Wald statistic is constant equal to $\frac{T}{T-1}$ under the null.

For these reasons, performing the χ^2 test with the degree of freedom 1 using statistics $W_T(\tilde{\eta}_{21})$ does not generally control the type I error of the test when K is small relative to T. A naive application of the weak instrument robust AR confidence intervals are therefore not recommended to perform inference for NP in such setting.

While the asymptotic validity of the weak instrument robust AR confidence intervals does not carry over to NP with a few shock-sign restrictions, the analysis for their invalidity shown in the previous section offers some approach to perform an asymptotically valid hypothesis tests for the null hypothesis of the impulse response being equal to $\tilde{\eta}_{21}$. One way is to use $T^2 \left(\hat{\gamma}_{2T} - \tilde{\eta}_{21}\hat{\gamma}_{1T}\right)^2$ as a test statistic and obtain the null distribution of the statistic numerically by simulating a feasible version of the right-hand side of (32),

$$\begin{pmatrix} -\tilde{\eta}_{21} & 1 \end{pmatrix} \hat{\mathbf{A}}_0^{-1} \left[\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t}) \boldsymbol{\varepsilon}_t \right] \left[\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t}) \boldsymbol{\varepsilon}_t \right]' (\hat{\mathbf{A}}_0^{-1})' \begin{pmatrix} -\tilde{\eta}_{21} & 1 \end{pmatrix}', \tag{34}$$

with $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$, $t = 1, \ldots, K$, i.i.d. A practical drawback of this approach is computationally heavy as the null distribution of the statistic is not pivotal so that the critical value of the test needs to be simulated at every null value of $\tilde{\eta}_{21}$. We leave how this approach can be compared with the robust Bayes inference approach of GKR for a future question.

5 Monte Carlo

This section describes a Monte Carlo simulation designed to compare the coverage properties of the weak-proxy robust confidence intervals in the NP setting against that of the robust Bayesian credible intervals under a corresponding set of shock-sign NR.

5.1 Design

For consistency with the theoretical results obtained above, we maintain the bivariate SVAR(0) described in Section 2. We set

$$\mathbf{A}_0^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0.2 & 1.8 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.5 & -0.8 \\ -0.8 & 3.28 \end{bmatrix}. \tag{35}$$

The parameter of interest is the impulse response of y_{2t} to a unit shock in y_{1t}, which is $0.2/0.5 = 0.4$

We proceed under the assumption that $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{2\times 1}, \mathbf{I}_n)$ for $t \notin \{1, \ldots, K\}$. For $t \in \{1, \ldots, K\}$, we assume that $\varepsilon_{1t} \sim N^+(0,1)$, where $N^+(0,1)$ is the standard normal distribution truncated to be positive, and that $\varepsilon_{2t} \sim N(0,1)$. We are therefore conditioning on ε_{1t} being positive in the first K periods.[11] We fix the sample length $T = 500$, which is consistent with empirical applications using monthly data, and compare the two approaches to identification and inference under different assumptions about K.

Under the NP approach, the proxy z_t is generated according to $z_t = 1(t \leq K)$, where $1(.)$ is the indicator function. For each Monte Carlo sample, we compute weak-proxy robust confidence intervals by applying the replication code from MSW.[12]

Under the NR approach, we impose the restriction $\varepsilon_{1t} \geq 0$ for $t \in \{1, \ldots, K\}$ along with the sign normalisation $\mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{e}_{1,n} \geq 0$. We conduct inference using the robust Bayesian approach from GKR. We assume an improper Jeffreys' prior over $\boldsymbol{\Sigma}$, which implies that the posterior is inverse-Wishart. In each Monte Carlo sample, we obtain 1,000 draws of $\boldsymbol{\Sigma}$ such that the conditional identified set – the set of values of $\mathbf{q}_1$ that are consistent with the NR and sign normalisation – is nonempty.[13] When the conditional identified set is nonempty, we compute its bounds by intersecting the bounds that would be obtained under each separate shock-sign restriction (for which we have analytical expressions; see Appendix). We construct a robust credible interval with credibility α by computing the $(1-\alpha)/2$ quantile of the posterior distribution of the lower bound and the $(1+\alpha)/2$ quantile of the posterior distribution of the upper bound.

5.2 Results

To summarise the performance of the weak-proxy robust approach to inference in the NP setting, we compute five statistics of interest. The first is the coverage probability of the weak-proxy robust confidence interval, which is the share of Monte Carlo replications in which the confidence interval includes the true value of the impulse response. The second is the proportion of Monte Carlo samples in which the confidence interval is unbounded. We also present the average and median width of the

[11] For $t \in \{1, \ldots, K\}$, we draw ε_{1t} from a truncated normal distribution using the inverse-CDF transform. Assuming that ε_{1t} is known to be positive in K random periods rather than the first K periods does not change the results, since the shocks and data are iid over time. The results are similar if we do not condition on the the shocks being positive and instead assume that the sign of the first shock is revealed in the first K periods.

[12] The replication code was obtained from José Luis Montiel Olea's website: `http://www.joseluismontielolea.com`.

[13] We check whether the identified set is nonempty by applying the 'Chebyshev criterion' from Amir-Ahmadi and Drautzburg (2021).

Table 1: Weak-proxy Robust Inference – Monte Carlo Results for $T = 500$ and $\alpha = 0.95$

K	Coverage prob.	Unbounded	Average width	Median width	Average W
1	1.0000	1.0000	∞	∞	1.002
2	1.0000	1.0000	∞	∞	1.201
3	1.0000	1.0000	∞	∞	1.463
4	0.9953	0.9825	∞	∞	1.741
5	0.9793	0.8897	∞	∞	2.048
10	0.9557	0.5567	∞	∞	3.601
15	0.9544	0.3207	∞	14.996	5.194
20	0.9546	0.1739	∞	8.2676	6.849
25	0.9490	0.0910	∞	6.2054	8.477
30	0.9508	0.0459	∞	5.0615	10.113
35	0.9555	0.0207	∞	4.3371	11.824
40	0.9508	0.0105	∞	3.9729	13.453
45	0.9508	0.0051	∞	3.6197	15.108
50	0.9525	0.0028	∞	3.3551	16.821
100	0.9481	0.0000	2.3191	2.1343	34.578

Notes: 'Coverage prob.' is the coverage probability of the 95 per cent weak-proxy robust confidence interval; 'Unbounded' is the proportion of Monte Carlo samples in which the confidence interval is unbounded; W is the Wald statistic for testing the null hypothesis that the covariance between the proxy and y_{1t} is zero.

confidence interval across the Monte Carlo replications. Finally, we present the average value of the Wald statistic for testing the null hypothesis that the covariance between the proxy and y_{1t} is zero (i.e. $\hat{\gamma}_{1T} = 0$), which is a measure of the strength of the proxy.

The 95 per cent weak-proxy robust confidence interval is always unbounded when $K = 1, 2, 3$, so the coverage probability is trivially equal to one (Table 1). As K increases, the confidence intervals are bounded with higher probability. Coverage probabilities remain higher than the nominal level for small-to-moderate values of K. Only for larger values of K do the confidence intervals possess approximately correct coverage. The improving coverage properties of the confidence intervals as K increases reflects the fact that null distribution of W_T is better approximated by the $\chi^2(1)$ distribution. The declining width of the confidence intervals as K increases reflects the increasing strength of the proxy for the target structural shock, as evident from the increasing average value of the Wald statistic for the test $\hat{\gamma}_{1,T} = 0$.

Table 2 repeats this exercise when $\alpha = 0.68$. Here, the results are essentially the

Table 2: Weak-proxy Robust Inference – Monte Carlo Results for $T = 500$ and $\alpha = 0.68$

K	Coverage prob.	Unbounded	Average width	Median width	Average W
1	0.0000	0.0000	0	0	1.002
2	0.4962	0.3698	∞	7.5814	1.2005
3	0.5638	0.3565	∞	8.1493	1.4632
4	0.5987	0.3247	∞	7.3015	1.7408
5	0.6311	0.2761	∞	6.1407	2.0481
10	0.6575	0.1091	∞	3.6805	3.6013
15	0.6652	0.0502	∞	2.8572	5.1943
20	0.6752	0.0164	∞	2.392	6.8491
25	0.6663	0.0075	∞	2.1274	8.4766
30	0.6687	0.0023	∞	1.908	10.1131
35	0.6668	0.001	∞	1.7403	11.8237
40	0.6686	0.0002	∞	1.6412	13.4533
45	0.675	0.0002	∞	1.5379	15.1077
50	0.6747	0.0002	∞	1.4571	16.8211
100	0.678	0.0000	1.0803	1.0104	34.5785

Notes: 'Coverage prob.' is the coverage probability of the 68 per cent weak-proxy robust confidence interval; 'Unbounded' is the proportion of Monte Carlo samples in which the confidence interval is unbounded; W is the Wald statistic for testing the null hypothesis that the covariance between the proxy and y_{1t} is zero.

Table 3: Monte Carlo Results for $T = 500$ – Robust Bayesian Inference

K	Coverage prob.	Unbounded	Average width	Median width
1	0.9998	0.7785	∞	∞
5	0.9994	0.2873	∞	9.860
10	0.9991	0.0804	∞	6.949
15	0.9981	0.0218	∞	6.365
20	0.9974	0.0052	∞	6.109
25	0.9973	0.0017	∞	5.982
30	0.9973	0.0005	∞	5.893
35	0.9954	0.0001	∞	5.828
40	0.9942	0	5.911	5.776
45	0.9944	0	5.828	5.737
50	0.9935	0	5.781	5.705
100	0.9895	0	5.581	5.544

Notes: 'Coverage prob.' is the coverage probability of the 95 per cent robust credible interval; 'Unbounded' is the proportion of Monte Carlo samples in which the robust credible interval is unbounded.

reverse of the case where $\alpha = 0.95$. In particular, the confidence intervals possess coverage probabilities *lower* than the nominal level for small values of K. Again, the coverage probabilities approach the nominal level only for larger values of K.

Table 3 presents analogous statistics for the robust Bayesian approach to inference under the shock-sign NR (with the exception of the average Wald statistic, which is not applicable in this setting). Under the shock-sign NR, the conditional identified set is not necessarily unbounded at every value of the reduced-form parameters, so the robust credible intervals are not necessarily unbounded. However, when the restrictions are imposed in only a handful of periods, the robust credible intervals are unbounded in a large proportion of the Monte Carlo replications (this occurs in any particular sample when the conditional identified set is unbounded with high posterior probability). The robust credible interval contains the true impulse response with probability greater than the nominal level at all values of K, which is consistent with the theoretical results about the asymptotic behaviour of these intervals in GKR. However, the robust credible interval is conservative; even when $K = 100$, the coverage probability is close to 99 per cent and the credible interval is twice as wide on average as the weak-proxy robust credible intervals obtained under the narrative proxy.

References

Amir-Ahmadi, P. and T. Drautzburg (2021): "Identification and Inference with Ranking Restrictions," *Quantitative Economics*, 12, 1–39.

Antolín-Díaz, J. and J. Rubio-Ramírez (2018): "Narrative Sign Restrictions for SVARs," *American Economic Review*, 108, 2802–29.

Arias, J., J. Rubio-Ramírez, and D. Waggoner (2018): "Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685720.

Baumeister, C. and J. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

Ben Zeev, N. (2018): "What Can We Learn About News Shocks from the Late 1990s and Early 2000s Boom-bust Period?" *Journal of Economic Dynamics and Control*, 87, 94–105.

Boer, L. and H. Lütkepohl (in press): "Qualitative versus Quantitative External Information for Proxy Vector Autoregressive Analysis," *Journal of Economic Dynamics and Control*.

Budnik, K. and G. Rünstler (2020): "Identifying SVARs from Sparse Narrative Instruments: Dynamic Effects of U.S. Macroprudential Policies," European Central Bank Working Paper No. 2353.

Cheng, K. and Y. Yang (2020): "Revisiting the Effects of Monetary Policy Shocks: Evidence from SVAR with Narrative Sign Restrictions," *Economics Letters*, 196, 109598.

Furlanetto, F. and Ø. Robstad (2019): "Immigration and the Macroeconomy: Some New Empirical Evidence," *Review of Economic Dynamics*, 34, 1–19.

Giacomini, R. and T. Kitagawa (2021): "Robust Bayesian Inference for Set-identified Models," *Econometrica*, 89, 1519–1556.

Giacomini, R., T. Kitagawa, and M. Read (2021): "Identification and Inference Under Narrative Restrictions," arXiv: 2102.06456 [econ.EM].

Inoue, A. and L. Kilian (in press): "Joint Bayesian Inference about Impulse Responses in VAR Models," *Journal of Econometrics*.

Kilian, L. and X. Zhou (2020a): "Does Drawing Down the US Strategic Petroleum Reserve Help Stabilize Oil Prices?" *Journal of Applied Econometrics*, 35, 673–691.

——— (2020b): "Oil Prices, Exchange Rates and Interest Rates," Center for financial studies working paper series no. 646.

Laumer, S. (2020): "Government Spending and Heterogeneous Consumption Dynamics," *Journal of Economic Dynamics and Control*, 114, 103868.

Ludvigson, S., S. Ma, and S. Ng (2018): "Shock Restricted Structural Vector-Autoregressions," National Bureau of Economic Research Working Paper No. 23225.

——— (2021): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics*, 13, 369–410.

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

Montiel-Olea, J., J. Stock, and M. Watson (2021): "Inference in Structural Vector Autoregressions Identified with an External Instrument," *Journal of Econometrics*, 225, 74–87.

Plagborg-Møller, M. and C. Wolf (2021b): "Supplement to "Local Projections and VARs Estimate the Same Impulse Responses"," *Econometrica*, 89.

Redl, C. (2020): "Uncertainty Matters: Evidence from Close Elections," *Journal of International Economics*, 103296.

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

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Zhou, X. (2020): "Refining the Workhorse Oil Market Model," *Journal of Applied Econometrics*, 35, 130–140.