Identification and Inference Under Narrative Restrictions[*]

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

October 26, 2020

Abstract

We consider semi-structural time series models subject to 'narrative restrictions', which are inequality restrictions on functions of the structural shocks in specific time periods (as in Antolín-Díaz and Rubio-Ramírez (2018) and Ludvigson, Ma and Ng (2018)). These restrictions do not fit into the existing framework for studying set-identification and there is no known frequentist procedure for conducting inference in these models. We provide a formal framework for estimation and inference under narrative restrictions by: 1) formalizing the identification problem under narrative restrictions; 2) showing some undesirable properties of the Bayesian approach in Antolín-Díaz and Rubio-Ramírez (2018); 3) proposing an alternative robust Bayesian approach to estimation and inference that overcomes these problems; and 4) showing that our approach has frequentist validity in large samples. We illustrate the method by analyzing the effects of monetary policy in the United States using a novel restriction on the relative magnitudes of the same structural shock in different time periods.

[†]University College London, Department of Economics/Cemmap. Email: r.giacomini@ucl.ac.uk
[‡]University College London, Department of Economics/Cemmap. Email: t.kitagawa@ucl.ac.uk
[§]University College London, Department of Economics. Email: matthew.read.16@ucl.ac.uk

1 Introduction

Antolín-Díaz and Rubio-Ramírez (2018) (AR18) propose a new class of 'narrative sign restrictions' in structural vector autoregressions (SVARs), which are inequality restrictions that involve the structural shocks hitting the economy in particular time periods. Ludvigson, Ma and Ng (2018) and Ludvigson, Ma and Ng (forthcoming) also consider restrictions on the sign or magnitude of structural shocks in particular periods. These restrictions are potentially useful for empirical researchers, as demonstrated by a burgeoning empirical literature that adopts the approach in AR18, including Furlanetto and Robstad (2019), Kilian and Zhou (2019, 2020), Cheng and Yang (2020), Inoue and Kilian (2020), Laumer (2020), Redl (2020), Zhou (2020) and Antolín-Díaz, Petrella and Rubio-Ramírez (In Press). The nature of the restrictions raises a nonstandard estimation problem and novel econometric questions. This paper clarifies the nature of the problem and offers a solution that is valid from both Bayesian and frequentist perspectives.

Henceforth, we simply call 'narrative restrictions' (NR) any restrictions that can be written as an inequality involving structural shocks in particular periods. An example of NR are 'shock-sign restrictions', such as the restriction in AR18 that the US economy was hit by a positive monetary policy shock in October 1979. This is when the Federal Reserve markedly increased the federal funds rate following Paul Volcker becoming chairman, and is widely considered an example of a positive monetary policy shock (e.g., Romer and Romer 1989). AR18 also consider the 'historical decomposition restriction' that the change in the federal funds rate in October 1979 was overwhelmingly due to a monetary policy shock. This is an inequality restriction that simultaneously constrains the historical decomposition of the federal funds rate with respect to all structural shocks. We further consider novel 'shock-rank' NR and assume that the monetary policy shock in October 1979 was the largest positive realisation of the monetary policy shock in the sample period.

From the perspective of identification analysis, NR are fundamentally different from traditional restrictions such as the sign restrictions on impulse responses proposed in Uhlig (2005). Under the Gaussian specification for the structural shock distribution, traditional sign restrictions induce set-identification since they generate a set-valued mapping from the SVAR's reduced-form parameters to its structural parameters. NR also generate a set-valued mapping from the reduced-form parameters to the structural parameters, but the mapping

depends on the realisation of the data independently of the reduced-form parameters.[1] This novel feature of NR has important implications for identification and inference. In particular, one cannot directly apply the existing framework for analyzing identification and performing frequentist inference in set-identified models. This means that there is no known frequentist inference procedure for models subject to NR.

From the perspective of Bayesian analysis, there is little apparent difference in the algorithms that impose traditional or narrative restrictions, as demonstrated by the Bayesian approach proposed by AR18. However, we illustrate two features of the approach of AR18 that can spuriously affect inference. First, the conditional likelihood considered by AR18 implies that for some types of NR the prior is updated only in the direction that makes the NR unlikely to hold a-priori. Second, standard Bayesian inference under NR is sensitive to the choice of prior.

In this paper, we propose a formal framework for studying identification and for conducting estimation and inference under NR that is potentially appealing to both Bayesians and frequentists. We proceed in three steps. We first formalize the identification problem under NR. We then propose a robust Bayesian approach for estimation and inference in SVAR models subject to NR that overcomes the potential pitfalls of the approach considered by AR18. Finally, we show that our approach has frequentist validity in large samples.

The use in AR18 of a conditional likelihood can potentially cause problems when the conditioning event (the NR holding) is not ancillary – that is, its probability of occurring depends on the model's parameters. While the shock-sign restriction considered by AR18 is ancillary, the historical decomposition restriction is not. In SVARs this means that the numerator of the conditional likelihood is flat with respect to the orthonormal matrix that maps reduced-form VAR innovations into structural shocks, whereas the denominator can depend on this matrix. Given the reduced-form parameters, the conditional likelihood in these cases is therefore maximised at the value of the orthonormal matrix that minimises the probability that the NR are satisfied. Consequently, the posterior places more weight on values of the orthonormal matrix that yield a lower probability of the NR being satisfied. This lower posterior probability does not reflect any prior information about the plausibility of the NR. It also does not reflect any new information in the data, since the probability that the NR are satisfied does not depend on the data. Accordingly, we advocate using the unconditional likelihood – the joint probability of observing the data and the NR being

---

[1]Applications using SVARs with NR typically assume Gaussian structural shocks to facilitate Bayesian inference. Notable exceptions are Ludvigson, Ma and Ng (2018) and Ludvigson, Ma and Ng (forthcoming), who conduct inference using a bootstrap (although the frequentist validity of this bootstrap is unknown). In a single-equation setting, Petterson, Seim and Shapiro (2020) consider a likelihood-free framework to derive bounds for a slope parameter given restrictions on the magnitude of the errors.

satisfied – when constructing the posterior. Using the unconditional likelihood rather than the conditional likelihood is also computationally less demanding.

A standard Bayesian approach to inference remains problematic even when considering the unconditional likelihood. To see why, note that NR truncate the unconditional likelihood so that it is flat, with the points of truncation depending on the realisations of the data that enter the NR.[2] This implies that the posterior distribution of the orthonormal matrix will be proportional to the prior distribution whenever the likelihood is nonzero. Posterior inference may therefore be sensitive to the choice of prior for the orthonormal matrix. This is a problem that also occurs in set-identified models under standard restrictions (e.g. Poirier 1998). Moreover, this issue is not necessarily alleviated by imposing a prior that is uniform over the orthonormal matrix, since a prior that is noninformative for some parameters may be informative for functions of these parameters, such as the impulse responses (see, for example, Baumeister and Hamilton (2015)).

To address these issues, we adapt the robust Bayesian approach of Giacomini and Kitagawa (2018) (GK18) to models with NR. In the context of an SVAR under standard restrictions, this approach involves decomposing the prior for the structural parameters into a revisable prior for the reduced-form parameters and an unrevisable prior for the orthonormal matrix. Considering the class of all priors for the orthonormal matrix that are consistent with the identifying restrictions generates a class of posteriors, which can be summarised by a set of posterior means (an estimator of the identified set) and a robust credible region. This removes the source of posterior sensitivity.[3] We show that the approach also applies under NR, with modifications to account for the novel features of the restrictions. In particular, one cannot write down a conditional prior for the orthonormal matrix to represent the NR, because the mapping between the reduced-form parameters and the structural parameters induced by the restrictions depends on the realisation of the data independently of the reduced-form parameters, and a prior cannot depend on the realisation of the data. However, by considering the class of all priors consistent with any standard identifying restrictions (if present), one can trace out all possible posteriors that are consistent with the standard restrictions *and* the NR (given a prior on the reduced-form parameters). This is because standard restrictions truncate the support of the conditional prior, while NR truncate the support of the likelihood, so the posterior given any particular conditional prior is only supported on the common support of the conditional prior and the likelihood. We describe algorithms to implement this approach.

---

[2]This is also true for the conditional likelihood when the NR are ancillary.

[3]Giacomini, Kitagawa and Read (2019) extend this approach to SVARs where the parameters of interest are set-identified using so-called 'external instruments' (also known as 'proxy SVARs').

We also explore the frequentist properties of our robust Bayesian procedure under NR. First, we introduce the notion of a 'conditional identified set' that extends the standard notion of an identified set to a setting where identification is defined in a repeated sampling experiment conditional on the set of observations entering the NR. Under the assumption that there is a fixed number of NR, we provide conditions under which our robust Bayesian approach provides asymptotically valid frequentist inference about the conditional identified set for the impulse response and for the impulse response itself.

We illustrate our methods by estimating the effect of monetary policy shocks in the United States. We find that inferences about the effect of monetary policy shocks on output obtained under the NR considered in AR18 may be sensitive to the choice of conditional prior for the orthonormal matrix. We also apply a novel 'shock-rank' restriction, which constrains the structural shocks to be consistent with views about the magnitude of a particular shock hitting the economy in a particular period relative to the same shock in other periods. Specifically, we assume that the monetary policy shock in October 1979 was the largest positive realisation of the monetary policy shock in the sample period. The large number of constraints generated by this restriction poses numerical challenges for existing algorithms that have been used to conduct inference under traditional sign restrictions, so we adapt algorithms recently developed by Amir-Ahmadi and Drautzburg (2019). We find that the shock-rank restriction substantially tightens inference about the effects of US monetary policy relative to the restrictions on the historical decomposition used in AR18.

Outline. The remainder of the paper is structured as follows. Section 2 explores the econometric issues that arise when imposing NR using a simple bivariate example. Section 3 sets these issues out in a general SVAR(p) framework and introduces our novel shock-rank restrictions. Section 4 explains how to conduct prior-robust Bayesian inference under NR and explores the frequentist properties of this approach. Section 5 describes numerical algorithms that can be used to conduct single-prior Bayesian inference using the unconditional likelihood to construct the posterior and to conduct prior-robust Bayesian inference under NR. Section 6 contains the empirical application.

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

2   A Bivariate Example

This section sets out the econometric issues that arise when imposing NR, using the simplest possible SVAR – a bivariate SVAR(0) – as an example. We abstract from dynamics for ease of exposition, but this is without loss of generality.

Consider the bivariate SVAR(0) $\mathbf{A}_0\mathbf{y}_t = \boldsymbol{\varepsilon}_t$, for $t = 1,\ldots,T$, where $\mathbf{y}_t = (y_{1t}, y_{2t})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ with $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{2\times1}, \mathbf{I}_2)$. The orthogonal reduced form of the model sets $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$, where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor (with positive diagonal elements) of the innovation covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{y}_t\mathbf{y}_t') = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$. We parameterise $\boldsymbol{\Sigma}_{tr}$ directly as

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

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

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

where the first set is the set of 'rotation' matrices and the second set is the set of 'reflection' matrices. We henceforth leave the restriction $\theta \in [-\pi, \pi]$ implicit.

In the absence of any restrictions, the set of values for $\mathbf{A}_0$ that are consistent with the reduced-form parameters is

$$\mathbf{A}_0 \in \left\{ \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ -\sigma_{21}\cos\theta - \sigma_{22}\sin\theta & \sigma_{11}\cos\theta \end{bmatrix}, \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ \sigma_{22}\sin\theta - \sigma_{21}\cos\theta & -\sigma_{11}\cos\theta \end{bmatrix} \right\}. \tag{1}$$

We impose the 'sign normalisation' that the diagonal elements of $\mathbf{A}_0$ are nonnegative.

Typically the object of interest in analyses using SVARs is the impulse response rather than the structural parameters themselves. In the absence of restrictions, the set of admissible values for the matrix of contemporaneous impulse responses is

$$\mathbf{A}_0^{-1} \in \left\{ \begin{bmatrix} \sigma_{11}\cos\theta & -\sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{22}\cos\theta - \sigma_{21}\sin\theta \end{bmatrix}, \begin{bmatrix} \sigma_{11}\cos\theta & \sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{21}\sin\theta - \sigma_{22}\cos\theta \end{bmatrix} \right\}. \tag{2}$$

We denote the contemporaneous impulse response of y_{1t} to a standard-deviation structural shock to y_{1t} by $\eta \equiv \sigma_{11}\cos\theta$.

2.1 Restrictions on the sign of a structural shock

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

$$\varepsilon_{1k} = \mathbf{e}'_{1,2}\mathbf{A}_0\mathbf{y}_k = (\sigma_{11}\sigma_{22})^{-1}(\sigma_{22}y_{1k}\cos\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\sin\theta) \geq 0. \tag{3}$$

We refer to this type of restriction as a 'shock-sign' restriction. Under the sign normalisation and the shock-sign restriction, θ is restricted to the set

$$\theta \in \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \geq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}$$
$$\cup \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \leq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}. \tag{4}$$

Since y_{1k} and y_{2k} enter the inequalities characterising this set, the shock-sign restriction induces a set-valued mapping from the reduced-form parameter $\boldsymbol{\phi}$ to the parameter θ that depends on the realisation of the data in the period in which the shock-sign restriction is imposed. For example, if $\sigma_{21} < 0$, $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$ and $y_{1k} > 0$,

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right)\right]. [4] \tag{5}$$

Assume that the econometrician observes $\mathbf{y}^T = (\mathbf{y}'_1, \ldots, \mathbf{y}'_T)'$. For simplicity, we assume for now that the econometrician knows $\boldsymbol{\phi}$.[5]

When conducting Bayesian inference, AR18 construct the posterior using the conditional likelihood, which is the likelihood of observing the data conditional on the NR holding. Given the realisation of the data in period k, Equation 3 implies that the restricted structural shock can be written as a function $\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k)$. The conditional likelihood is then

$$p\left(\mathbf{y}^T|\theta, \boldsymbol{\phi}, \varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0\right) = \frac{\prod_{t=1}^{T}(2\pi)^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\mathbf{y}'_t\boldsymbol{\Sigma}^{-1}\mathbf{y}_t\right)}{\Pr(\varepsilon_{1k} \geq 0|\theta, \boldsymbol{\phi})}\mathbf{1}\left(\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0\right), \tag{6}$$

where $n = 2$ is the dimension of the SVAR. The numerator in the first term is a function only of the reduced-form parameter $\boldsymbol{\phi}$ and the data $\mathbf{y}^T$, and is therefore independent of θ. Also, $\Pr(\varepsilon_{1k} \geq 0|\theta, \boldsymbol{\phi}) = 1/2$, since the marginal distribution of ε_{1k} is standard normal. The conditional likelihood therefore depends on θ only through the indicator function $\mathbf{1}(\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0)$. The conditional likelihood function is flat over the region for θ satisfying the shock-sign restriction and is zero outside this region; that is, the likelihood is truncated, with the truncation points depending on the realisation of the data entering the

---

[4]See Appendix A for the full analytical characterisation of this mapping.
[5]$\boldsymbol{\phi}$ is point-identified from the Cholesky decomposition of $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{y}_t\mathbf{y}'_t)$.

narrative sign restriction.

To illustrate, the top-left panel of Figure 1 plots the likelihood function given different realisations of the data drawn from a data-generating process with $\sigma_{21} < 0$.[6] The likelihood is clearly flat, with the support of the nonzero region depending on the particular realisation of the data. The flat likelihood function implies that the posterior distribution will be proportional to the prior in the regions where the likelihood function is nonzero, and it will be zero outside these regions. The standard algorithm for inference in Bayesian SVARs identified via sign restrictions implies a uniform (or Haar) prior over $\mathbf{Q}$, as does the algorithm proposed in AR18.[7] In the bivariate example, this is equivalent to a prior for θ that is uniform over the interval $[-\pi, \pi]$. Clearly, this prior implies that the posterior for θ is also uniform over the interval for θ where the likelihood function is nonzero.

Figure 1: Shock-sign Restriction
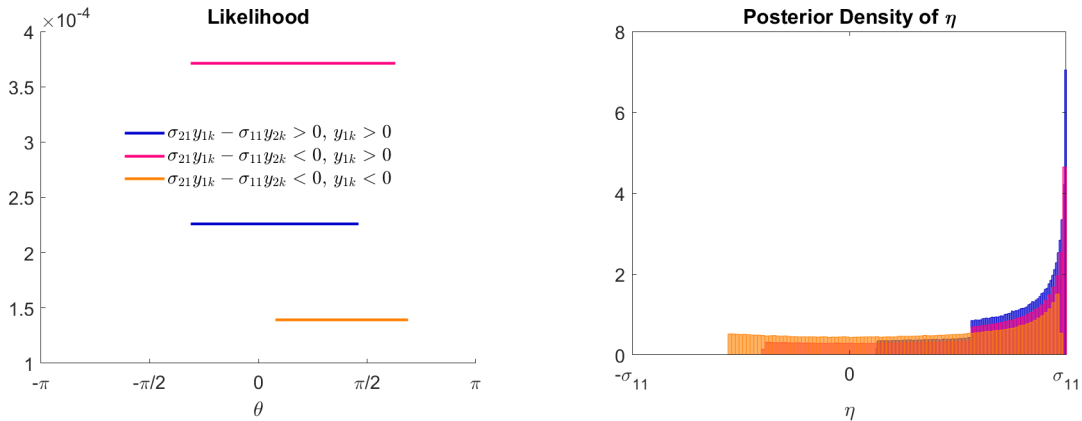
Notes: $T = 3$, ϕ is known and $\varepsilon_{1k}(\theta, \phi, \mathbf{y}_k) \geq 0$ is the narrative sign restriction; likelihood in top-left panel is zero outside of plotted intervals; posterior density of η is approximated using 1,000,000 draws of θ from its (uniform) posterior distribution.

The top-right panel of Figure 1 plots the posterior distribution for the impulse response η induced by a uniform prior over θ given the same realisations of the data for which the likelihood was plotted in the top-left panel. The uniform posterior for θ induced by the flat likelihood and uniform prior induces a posterior for η that is clearly non-uniform. In particular, the posterior for η assigns more probability mass to more-extreme values of η. This is also the case in standard set-identified SVARs (see Baumeister and Hamilton (2015)).

---

[6]The data-generating process assumes $\mathbf{A}_0 = \begin{bmatrix} 1 & 0.5 \\ 0.2 & 1.2 \end{bmatrix}$, which implies that $\theta = \arcsin(0.5\sigma_{22})$ with $\mathbf{Q}$ equal to the rotation matrix. We assume the time series is of length $T = 3$ and draw sequences of structural shocks such that $\varepsilon_{1,1} \geq 0$. We set T to a small number to control Monte Carlo sampling error in exercises below without needing to resort to extremely large Monte Carlo sample sizes.

[7]See, for example, Rubio-Ramírez, Waggoner and Zha (2010), Baumeister and Hamilton (2015) and Arias, Rubio-Ramírez and Waggoner (2018).

One difference here is that the support and shape of the posterior for the impulse response depends on the realisation of the data in period k through its effect on the truncation points of the likelihood, whereas under standard sign restrictions the posterior does not depend on the realisation of the data given the reduced-form parameters. For example, when $\sigma_{21} < 0$, $\sigma_{21} y_{1k} - \sigma_{11} y_{2k} > 0$ and $y_{1k} > 0$,

$$\eta \in \left[\sigma_{11} \cos \left(\arctan \left(\max \left\{ -\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22} y_{1k}}{\sigma_{21} y_{1k} - \sigma_{11} y_{2k}} \right\} \right) \right), \sigma_{11} \right]. \tag{7}$$

2.2 Restrictions on the historical decomposition

The historical decomposition is the contribution of a particular structural shock to the observed change in a particular variable over some horizon. Let $H_{i,j,t}$ represent the contribution of the jth structural shock to the unexpected change in the ith variable in period t. AR18 consider two broad types of restrictions on the historical decomposition. One imposes that the jth shock was the 'overwhelming contributor' to the observed change in the ith variable, which requires that $|H_{i,j,t}| \geq \sum_{k \neq j} |H_{i,k,t}|$. The other imposes that the jth shock was the 'most important contributor' to the observed change in the ith variable, which requires that $|H_{i,j,t}| \geq \max_{k \neq j} |H_{i,k,t}|$. In a bivariate SVAR, these two restrictions are identical. AR18 impose these restrictions alongside shock-sign restrictions.

The contribution of the first shock to the change in the first variable in the kth period is equal to the impact impulse response of the first variable to the first shock multiplied by the realisation of the first shock in the kth period. In the bivariate example, this is

$$H_{1,1,k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{22}^{-1} \left(\sigma_{22} y_{1k} \cos^2 \theta + (\sigma_{11} y_{2k} - \sigma_{21} y_{1k}) \cos \theta \sin \theta \right). \tag{8}$$

The contribution of the second shock to the change in the first variable in the kth period is

$$H_{2,1,1}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{22}^{-1} \left(\sigma_{22} y_{1k} \sin^2 \theta - (\sigma_{11} y_{2k} + \sigma_{21} y_{1k}) \cos \theta \sin \theta \right). \tag{9}$$

Consider the restriction that the first structural shock was positive and that it was the most important (or overwhelming) contributor to the change in the first variable in the kth period. Under these restrictions and the sign normalisation, the set of values that θ can take will be defined by a set of inequalities that depend on the reduced-form parameters and the realisation of the data in period k. As in the case of the shock-sign restriction, this set of restrictions will generate a mapping from the reduced-form parameters to the set of

9

admissible values of θ that will depend on the realisation of the data in period k.[8]

Let $\mathcal{D}$ represent the event $\{\varepsilon_{1k} \geq 0, |\tilde{H}_{1,1,k}(\theta, \phi, \varepsilon_{1k})| \geq |\tilde{H}_{2,1,k}(\theta, \phi, \varepsilon_{2k})|\}$, where $\tilde{H}_{i,j,k}(\theta, \phi, \varepsilon_k)$ is the historical decomposition as a function of the period-k structural shocks rather than the data. The conditional likelihood function given these restrictions is

$$
\begin{aligned}
&p\left(\mathbf{y}^T | \theta, \phi, \mathcal{D}\right) = \\
&\frac{\prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{y}_t' \mathbf{\Sigma}^{-1} \mathbf{y}_t\right)}{\Pr(\mathcal{D}|\theta, \phi)} \times 1\left(\varepsilon_{1k}(\theta, \phi, \mathbf{y}_k) \geq 0, |H_{1,1,k}(\theta, \phi, \mathbf{y}_k)| \geq |H_{2,1,k}(\theta, \phi, \mathbf{y}_k)|\right).
\end{aligned}
$$
(10)

As in the case of the shock-sign restriction, the numerator of the first term is independent of θ. In contrast, when the historical decomposition is restricted, the probability that the shocks satisfy the restriction depends on θ through the historical decomposition. The conditional likelihood therefore depends on θ both through this probability and through the indicator function determining the truncation points of the likelihood. Consequently, the likelihood function is not necessarily flat when it is nonzero.

To illustrate, the left panel of Figure 2 plots the likelihood function for a random re-alisation of the data satisfying the restrictions using the same data-generating process as above and assuming that ϕ is known. The probability in the denominator of the likelihood is estimated by drawing 1,000,000 realisations of ε_k (a bivariate standard normal random variable) and computing the proportion of draws satisfying the restriction at each value of θ.[9] This probability is plotted in the right panel of Figure 2. The likelihood is again truncated depending on the realisation of the data, but within the range where it is nonzero it is no longer flat. Despite there being a set-valued mapping from ϕ and $\mathbf{y}_k$ to θ, the likelihood has a unique maximum occurring at the value of θ that minimises the probability that the NR are satisfied (within the set of values of θ that are consistent with the restrictions). The posterior distribution for θ induced by a uniform prior will clearly assign greater posterior probability to values of θ that yield a lower probability of satisfying the NR.

It is well-known in statistics that when conducting likelihood-based inference, one should condition only on events whose probability of occurring does not depend on the parameter of interest, or 'ancillary' events. When the only NR is a shock-sign restriction, the probability that the restriction is satisfied is independent of the parameters; that is, the event that the

---

[8]See Appendix A for this set of inequalities. It is more difficult to analytically characterise the mapping from reduced-form parameters and realisations of the data to θ than in the shock-sign example, so we do not pursue this.

[9]One could potentially compute this probability without recourse to Monte Carlo methods, since $|\tilde{H}_{1,1,k}(\theta, \phi, \varepsilon_{1k})| - |\tilde{H}_{2,1,k}(\theta, \phi, \varepsilon_{2k})|$ is the difference between two independent half-normal distributions, and the event $\varepsilon_{1k} \geq 0$ is independent of the event $|\tilde{H}_{1,1,k}(\theta, \phi, \varepsilon_{1k})| \geq |\tilde{H}_{2,1,k}(\theta, \phi, \varepsilon_{2k})|$.

Figure 2: Restriction on the Historical Decomposition – Conditional Likelihood
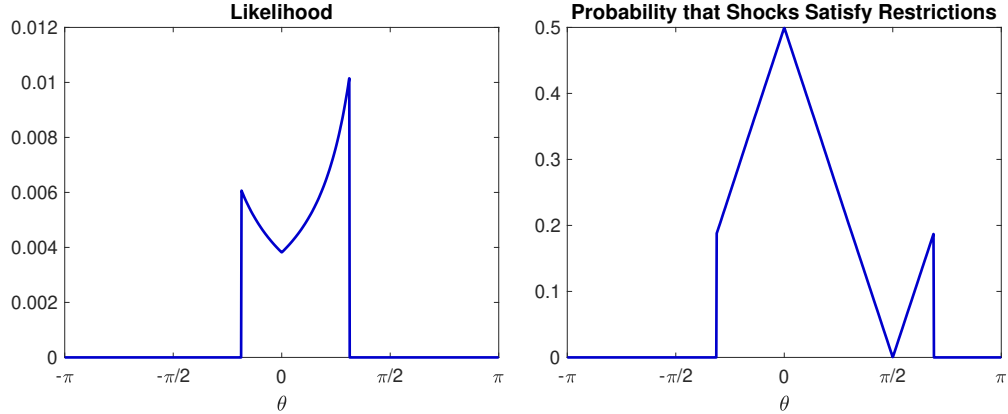
NR is satisfied is ancillary, and the conditional and unconditional likelihoods are identical up to a scale factor. In the case where there is also a restriction on the historical decomposition, the probability that the NR are satisfied depends on the parameter of interest. The event that the NR are satisfied is therefore not ancillary. Using the conditional likelihood to construct the posterior distribution will put more weight on values of θ that yield lower probabilities that the NR are satisfied. This lower posterior probability does not reflect any prior information about the probability that the NR are satisfied. It also does not reflect any new information in the data, since the probability that the NR are satisfied does not depend on the data. We therefore advocate forming the likelihood without conditioning on the restrictions holding.

The joint (unconditional) likelihood of observing the data and the NR holding is obtained by multiplying the conditional likelihood by the probability that the NR are satisfied:

$$p\left(\mathbf{y}^T, 1(D) = 1 | \theta, \phi\right) =$$

$$\prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y}_t'\boldsymbol{\Sigma}^{-1}\mathbf{y}_t\right)\right) \times 1\left(\varepsilon_{1k}(\theta, \phi, \mathbf{y}_k) \geq 0, |H_{1,1,k}(\theta, \phi, \mathbf{y}_k)| \geq |H_{2,1,k}(\theta, \phi, \mathbf{y}_k)|\right).$$

$$(11)$$

Conditional on being nonzero, the unconditional likelihood function is flat with respect to θ; θ only affects the unconditional likelihood through the points of truncation. To illustrate, Figure 3 plots the unconditional likelihood function given the same realisation of the data used in Figure 2. As in the case of the shock-sign restriction, the flat unconditional likelihood implies that posterior inference may be sensitive to the choice of prior.

11

Figure 3: Restriction on the Historical Decomposition – Unconditional Likelihood
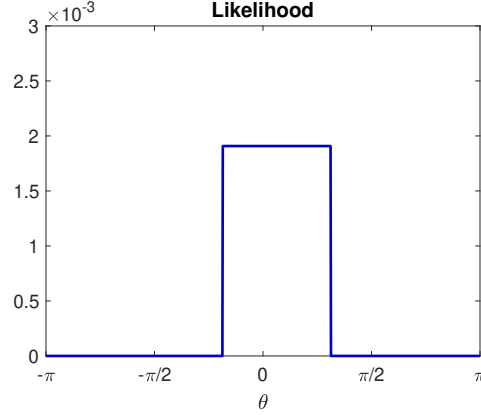
Notes: $T = 3$ and $\boldsymbol{\phi}$ is known; $\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0$ and $|H_{1,1,1}(\boldsymbol{\phi}, \theta, \mathbf{y}_k)| \geq |H_{2,1,1}(\boldsymbol{\phi}, \theta, \mathbf{y}_k)|$ are the narrative sign restrictions.

3 General Framework

This section sets out the issues discussed above in the general framework of an SVAR(p) and formalises the sense in which NR can be considered identifying restrictions.

3.1 The Structural Vector Autoregression

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

$$\mathbf{A}_0 \mathbf{y}_t = \sum_{l=1}^{p} \mathbf{A}_l \mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{12}$$

where $\mathbf{A}_0$ has positive diagonal elements (a sign normalisation) and is invertible, and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n \times 1}, \mathbf{I}_n)$ are structural shocks. 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 straight-forward to include. Letting $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ and $\mathbf{A}_+ = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, rewrite the SVAR($p$) as

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{A}_+ \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T. \tag{13}$$

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

$$\mathbf{y}_t = \mathbf{B} \mathbf{x}_t + \mathbf{u}_t, \quad t = 1, ..., T, \tag{14}$$

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 $\mathbf{u}_t = \mathbf{A}_0^{-1} \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n \times 1}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$. $\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})')' \in \boldsymbol{\Phi}$ 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.[10]

As is standard in the literature that considers set-identified SVARs, we reparameterise the model into its 'orthogonal reduced form':

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{15}$$

where $\mathbf{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of $\mathbf{\Sigma}$ (i.e. $\mathbf{\Sigma}_{tr}\mathbf{\Sigma}'_{tr} = \mathbf{\Sigma}$) with diagonal elements normalized to be non-negative, $\mathbf{Q}$ is an $n \times n$ orthonormal matrix and $\mathcal{O}(n)$ is the set of all such matrices. The parameterisations are related through the mapping $\mathbf{B} = \mathbf{A}_0^{-1}\mathbf{A}_+$, $\mathbf{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$ and $\mathbf{Q} = \mathbf{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$, or $\mathbf{A}_0 = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}$ and $\mathbf{A}_+ = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}\mathbf{B}$.

The VMA(∞) representation of the model is

$$\mathbf{y}_t = \sum_{h=0}^{\infty} \mathbf{C}_h \mathbf{u}_{t-h} = \sum_{h=0}^{\infty} \mathbf{C}_h \mathbf{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{16}$$

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.[11] The (i,j)th element of the matrix $\mathbf{C}_h\mathbf{\Sigma}_{tr}\mathbf{Q}$, which we denote by $\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q})$, is the impulse response of the ith variable to the jth structural shock at the hth horizon:

$$\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{e}'_{i,n}\mathbf{C}_h\mathbf{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n} = \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}_j, \tag{17}$$

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

The historical decomposition is the cumulative contribution of the jth shock to the observed unexpected change in the ith variable between periods t and $t+h$:

$$H_{i,j,t,t+h} = \sum_{l=0}^{h} \mathbf{e}'_{i,n}\mathbf{C}_l\mathbf{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n}\mathbf{e}'_{j,n}\boldsymbol{\varepsilon}_{t+h-l} = \sum_{l=0}^{h} \mathbf{c}'_{i,l}(\boldsymbol{\phi})\mathbf{q}_j\mathbf{q}'_j\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_{t+h-l}. \tag{18}$$

3.2 Narrative restrictions

In the absence of any restrictions on $\mathbf{A}_0$, it is well-known that $\mathbf{A}_0$ (and thus $\mathbf{A}_+$) is set-identified. Since any $\mathbf{A}_0 = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}$ satisfies $\mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})' = \mathbf{\Sigma}$, the identified set for $\mathbf{A}_0$ is $\{\mathbf{A}_0 = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1} : \mathbf{Q} \in \mathcal{O}(n)\}$. Imposing traditional identifying restrictions on the SVAR is equivalent to restricting $\mathbf{Q}$ to lie in a subspace of $\mathcal{O}(n)$. It is conventional to impose a 'sign normalisation' on the structural shocks. We normalise the diagonal elements of $\mathbf{A}_0$ to be

---

[10]The VAR(p) is invertible into a VMA(∞) process when the eigenvalues of the companion matrix lie inside the unit circle. See Hamilton (1994) or Kilian and Lütkepohl (2017).

[11]$\mathbf{C}_h$ is defined recursively by $\mathbf{C}_h = \sum_{l=1}^{\min\{k,p\}} \mathbf{B}_l \mathbf{C}_{h-l}$ for $h \geq 1$ with $\mathbf{C}_0 = \mathbf{I}_n$.

non-negative, so a positive value of ε_{it} is a positive shock to the ith equation in the SVAR at time t. The sign normalisation implies that $\text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}$.

The sign restrictions proposed by Uhlig (2005) restrict the impulse responses of particular variables to particular shocks. The restriction that the horizon-h impulse response of the ith variable to the jth shock is nonnegative is $c'_{i,h}(\boldsymbol{\phi})\mathbf{q}_j \geq 0$, which is a linear inequality restriction on a single column of $\mathbf{Q}$ that depends only on the reduced-form parameter $\boldsymbol{\phi}$. Restrictions on elements of $\mathbf{A}_0$ take a similar form.

In contrast, the NR proposed by AR18 constrain the sign of the structural shocks and/or the historical decomposition in particular periods. The structural shocks are

$$\boldsymbol{\varepsilon}_t = \mathbf{A}_0 \mathbf{u}_t = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_t. \tag{19}$$

The ith structural shock at time t is therefore

$$\varepsilon_{it}(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{u}_t) = \mathbf{e}'_{i,n}\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_t = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_t)'\mathbf{q}_i. \tag{20}$$

Given knowledge of the reduced-form VAR parameters $\boldsymbol{\phi}$ and the reduced-form VAR innovations $\mathbf{u}_t$ (which follows from knowledge of $\boldsymbol{\phi}$ and the data $(\mathbf{y}_t, \mathbf{x}_t)$), a shock-sign restriction is a linear inequality restriction on a single column of $\mathbf{Q}$. In contrast with traditional sign restrictions, a shock-sign restriction depends on the data $(\mathbf{y}_t, \mathbf{x}_t)$ through the reduced-form VAR innovations independently of the reduced-form parameters $\boldsymbol{\phi}$.

AR18 consider two types of inequality restrictions that constrain the historical decomposition. An example of their 'Type A' restrictions is that the jth structural shock is the 'most important contributor' to the change in the ith variable between periods t and $t + h$, which is taken to mean that the absolute cumulative contribution of the jth shock to the change in the ith variable is larger than the contribution of any other shock, or $|H_{i,j,t,t+h}| \geq \max_{k \neq j} |H_{i,k,t,t+h}|$. Another example of a Type A restriction is that the jth structural shock is the 'least important contributor', in which case $|H_{i,j,t,t+h}| \leq \min_{k \neq j} |H_{i,k,t,t+h}|$. An example of their 'Type B' restrictions is that the jth structural shock is the 'overwhelming contributor' to the change in the ith variable between periods t and $t + h$, which is taken to mean that the absolute cumulative contribution of the jth structural shock to the change in the ith variable is larger than the sum of the contributions of all other shocks, or $|H_{i,j,t,t+h}| \geq \sum_{k \neq j} |H_{i,k,t,t+h}|$. Conversely, the jth shock is a 'negligible contributor' when $|H_{i,j,t,t+h}| \leq \sum_{k \neq j} |H_{i,k,t,t+h}|$. From Equation 18, it is clear that Type A and Type B restrictions are nonlinear inequality restrictions that simultaneously constrain every column of $\mathbf{Q}$ and that depend on the realisations of the data in particular periods independently of the reduced-form parameters.

A restriction not considered in AR18 is a restriction on the relative magnitudes of a particular structural shock in different periods. We refer to this type of restriction as a 'shock-rank' restriction. For example, in addition to requiring that the ith shock was positive in the tth period (a shock-sign restriction), one could impose that the ith structural shock in period t was the largest positive realisation of this shock. This requires that $\varepsilon_{it}(\phi, \mathbf{Q}, \mathbf{u}_t) \geq \max_{k \neq t}\{\varepsilon_{ik}(\phi, \mathbf{Q}, \mathbf{u}_k)\}$, which can be expressed as a system of $T - 1$ linear inequality restrictions on a single column of $\mathbf{Q}$: $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_t - \mathbf{u}_k))'\mathbf{q}_i \geq 0$ for $k \neq t$. Similarly, one could impose that the ith shock in the tth period was the largest negative realisation of this shock. These restrictions could also be applied to a subset of the observations rather than the full sample. For example, one could impose that the ith shock in the tth period was the largest positive realisation of the shock in periods $t_1, t_2, \ldots, t_K$. One could also impose that the ith shock in period t was the largest-magnitude realisation of that shock, or $|\varepsilon_{it}(\phi, \mathbf{Q}, \mathbf{u}_t)| \geq \max_{k \neq t}\{|\varepsilon_{ik}(\phi, \mathbf{Q}, \mathbf{u}_k)|\}$. If $\varepsilon_{it} > 0$, this would require that $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_t - \mathbf{u}_k))'\mathbf{q}_i \geq 0$ and $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_t + \mathbf{u}_k))'\mathbf{q}_i \geq 0$ for $k \neq t$, which is a system of $2(T - 1)$ linear inequalities constraining $\mathbf{q}_i$.

Assume that there are s NR that constrain the structural shocks in K distinct periods, $t_1, \ldots, t_K$. Let $\mathbf{U} = (\mathbf{u}'_{t_1}, \ldots, \mathbf{u}'_{t_K})'$ collect the reduced-form VAR innovations in these periods. The collection of NR are represented in the general form $N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}$. As an example, consider the case where there is a single shock-sign restriction in period k, $\varepsilon_{1k}(\phi, \mathbf{Q}, \mathbf{u}_k) \geq 0$, as well as a Type A restriction that the first structural shock is the most important contributor to the change in the first variable in period k. Then

$$N(\phi, \mathbf{Q}, \mathbf{U}) = \begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k)'\mathbf{q}_1 \\ |\mathbf{e}'_{1,n}\boldsymbol{\Sigma}_{tr}\mathbf{q}_1\mathbf{q}'_1\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k| - \max_{j \neq 1}|\mathbf{e}'_{1,n}\boldsymbol{\Sigma}_{tr}\mathbf{q}_j\mathbf{q}'_j\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k| \end{bmatrix} \geq \mathbf{0}_{2 \times 1}. \quad (21)$$

The set of NR and the sign normalisation will generate a set-valued mapping from ϕ to $\mathbf{Q}$ that depends on the realisation of the data through $\mathbf{U}$. In general, the set of values of $\mathbf{Q}$ satisfying the NR and the sign normalisation is $\{\mathbf{Q} : N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}, \text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}, \mathbf{Q} \in \mathcal{O}(n)\}$.

Traditional sign and zero restrictions can also be applied alongside NR. In what follows, we follow AR18 by explicitly allowing for sign restrictions on impulse responses and on elements of $\mathbf{A}_0$. We denote such sign restrictions by $S(\phi, \mathbf{Q}) \geq \mathbf{0}_{\tilde{s} \times 1}$, where $\tilde{s}$ is the number of traditional sign restrictions. It is straightforward to additionally allow for zero restrictions, including 'short-run' zero restrictions (as in Sims (1980) and Christiano, Eichenbaum and Evans (1999)), 'long-run' zero restrictions (as in Blanchard and Quah (1989)), or restrictions arising from external instruments (as in Stock and Watson (2012) and Mertens and Ravn (2013)); for example, see GK18 and Giacomini, Kitagawa and Read (2019).

3.3 Conditional and unconditional likelihoods

Let $\mathbf{Y}^T = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ collect the data, let $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_{t_1}', \ldots, \boldsymbol{\varepsilon}_{t_K}')'$ collect the structural shocks in the periods in which the NR are imposed, and let $\tilde{N}(\boldsymbol{\phi}, \mathbf{Q}, \boldsymbol{\varepsilon}) \geq \mathbf{0}_{s \times 1}$ represent the collection of NR as a function of the structural shocks rather than the data. The likelihood conditional on the NR holding is

$$p(\mathbf{Y}^T | N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}, \boldsymbol{\phi}, \mathbf{Q}) =$$
$$\left[\frac{\prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left((\mathbf{y}_t - \mathbf{Bx}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{Bx}_t)\right)\right)}{\Pr(\tilde{N}(\boldsymbol{\phi}, \mathbf{Q}, \boldsymbol{\varepsilon}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q})} \right] \times 1\left(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\right), \quad (22)$$

where we have suppressed conditioning on the initial conditions. The numerator of the term in square brackets depends only on $\boldsymbol{\phi}$ and the data, and is thus independent of $\mathbf{Q}$. The indicator function determines the truncation points of the likelihood and depends on $\mathbf{U}$. The truncation points therefore depend on the data that enter the NR through $\mathbf{U}$. The denominator of the first term, which is the probability that the NR are satisfied, will be a constant when there are only shock-sign or shock-rank restrictions.[12] However, when there are restrictions on the historical decomposition, this probability will depend on $\boldsymbol{\phi}$ and $\mathbf{Q}$. Importantly, even if $\boldsymbol{\phi}$ is known, the conditional likelihood will depend on $\mathbf{Q}$ through this probability. Moreover, given $\boldsymbol{\phi}$, the conditional likelihood will be maximised at the value of $\mathbf{Q}$ that minimises the probability that the NR are satisfied.

As discussed in Section 2 for the bivariate case, these issues suggest that one should construct the likelihood without conditioning on the NR holding when the probability that the restrictions are satisfied depends on the structural parameters. The joint (or unconditional) likelihood of observing the data and the NR being satisfied is

$$p(\mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q}) =$$
$$\left[\prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left((\mathbf{y}_t - \mathbf{Bx}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{Bx}_t)\right)\right) \right] \times 1\left(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\right). \quad (23)$$

The term in square brackets depends only on $\boldsymbol{\phi}$ and the data, and is thus independent of $\mathbf{Q}$. As in the case of the conditional likelihood, the truncation points of the likelihood depend on $\mathbf{U}$ independently of $\boldsymbol{\phi}$. As in the bivariate case (where $\mathbf{Q}$ was summarised by the scalar parameter θ), the likelihood function will be flat with respect to $\mathbf{Q}$ in a particular region of the parameter space and will be zero outside this region. In other words, at every value of $\boldsymbol{\phi}$ there will be a set of values of $\mathbf{Q}$ that satisfy the restrictions, which depend on the

---

[12]If there are s shock-sign restrictions, $\Pr(\tilde{N}(\boldsymbol{\phi}, \mathbf{Q}, \boldsymbol{\varepsilon}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q}) = (1/2)^s$.

data, but the value of the likelihood will be the same for all such values of $\mathbf{Q}$. The posterior of $\mathbf{Q}|\phi, \mathbf{Y}^T$ will therefore be proportional to the prior for $\mathbf{Q}|\phi$ in these regions. Given a fixed number of NR, the likelihood will possess flat regions even with a time-series of infinite length, so posterior inference may be sensitive to the choice of prior, even asymptotically. This motivates considering Bayesian inferential procedures that are robust to the choice of unrevisable prior for $\mathbf{Q}$, as discussed in GK18.

3.4 Point identification under NR

As argued in Sections 3.2 and 3.3, the mapping from the reduced-form parameters to the structural parameters is set-valued, whereas an important distinction from the standard set-identified SVARs is that the set-valued map under NR depends on the data. This raises a theoretical question of whether the NR-SVAR model is point-identified or not in a formal frequentist sense. This subsection examines point-identification of the structural parameters under NR for both unconditional and conditional likelihood models.

To suppress the notation, let $\mathbf{U}(\mathbf{Y}^T; \phi)$ be the value of the reduced-form residuals pinned down by given $\mathbf{Y}^T$ and ϕ, and define

$$D_N = D_N(\phi, \mathbf{Q}, \mathbf{Y}^T) \equiv 1\{N(\phi, \mathbf{Q}, \mathbf{U}(\mathbf{Y}^T; \phi)) \geq \mathbf{0}_{s \times 1}\},$$

$$r(\phi, \mathbf{Q}) \equiv \Pr(D_N(\phi, \mathbf{Q}, \mathbf{Y}^T) = 1 | \phi, \mathbf{Q}),$$

$$f(\mathbf{Y}^T | \phi) \equiv \prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)' \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)\right).$$

Then, the unconditional likelihood (the joint distribution of $\mathbf{Y}^T$ and D_N) can be expressed as

$$p(\mathbf{Y}^T, D_N = d | \phi, \mathbf{Q}) = \left[f(\mathbf{Y}^T | \phi) D_N(\phi, \mathbf{Q}, \mathbf{Y}^T)\right]^d \cdot \left[f(\mathbf{Y}^T | \phi) \left(1 - D_N(\phi, \mathbf{Q}, \mathbf{Y}^T)\right)\right]^{1-d}$$

$$= f(\mathbf{Y}^T | \phi) \cdot \left[D_N(\phi, \mathbf{Q}, \mathbf{Y}^T)\right]^d \cdot \left[1 - D_N(\phi, \mathbf{Q}, \mathbf{Y}^T)\right]^{1-d}. \quad (24)$$

Denoting the true parameter value by $(\phi_0, \mathbf{Q}_0)$, point-identification for the parametric model (24) requires that there is no other parameter value $(\phi, \mathbf{Q}) \neq (\phi_0, \mathbf{Q}_0)$ that is observationally equivalent to $(\phi_0, \mathbf{Q}_0)$.[13]

To assess the existence or non-existence of observationally equivalent parameter points, we analyze a statistical distance between $p(\mathbf{Y}^T, D_N = d | \phi, \mathbf{Q})$ and $p(\mathbf{Y}^T, D_N = d | \phi_0, \mathbf{Q}_0)$. Specifically, in the current setting where the support of the distribution of observables can

---

[13]$(\phi, \mathbf{Q}) \neq (\phi_0, \mathbf{Q}_0)$ is observationally equivalent to $(\phi_0, \mathbf{Q}_0)$ if $p(\mathbf{Y}^T, D_N = d | \phi, \mathbf{Q}) = p(\mathbf{Y}^T, D_N = d | \phi_0, \mathbf{Q}_0)$ holds for all $\mathbf{Y}^T$ and $d \in \{0, 1\}$.

depend on the parameters, it is convenient to work with the Hellinger distance:

$$HD(\phi, \mathbf{Q}) \equiv \sum_{d=0,1} \int_{\mathbf{Y}^T} \left(p^{1/2}(\mathbf{Y}^T, D_N = d|\phi, \mathbf{Q}) - p^{1/2}(\mathbf{Y}^T, D_N = d|\phi_0, \mathbf{Q}_0) \right)^2 d\mathbf{Y}^T$$

$$= 2\left(1 - \mathcal{H}(\phi, \mathbf{Q})\right), \quad \text{where}$$

$$\mathcal{H}(\phi, \mathbf{Q}) \equiv \sum_{d=0,1} \int_{\mathbf{Y}^T} p^{1/2}(\mathbf{Y}^T, D_N = d|\phi, \mathbf{Q}) \cdot p^{1/2}(\mathbf{Y}^T, D_N = d|\phi_0, \mathbf{Q}_0) d\mathbf{Y}^T. \quad (25)$$

As is known in the literature on minimum distance estimation (see, e.g., Basu, Shioya and Park 2011), $(\phi, \mathbf{Q})$ and $(\phi_0, \mathbf{Q}_0)$ are observationally equivalent if and only if $HD(\phi, \mathbf{Q}) = 0$, or equivalently, $\mathcal{H}(\phi, \mathbf{Q}) = 1$.

The conditional likelihood given $D_N = 1$ can be written as

$$p(\mathbf{Y}^T, |D_N = 1, \phi, \mathbf{Q}) = \frac{f(\mathbf{Y}^T|\phi)}{r(\phi, \mathbf{Q})} \cdot D_N(\phi, \mathbf{Q}, \mathbf{Y}^T). \quad (26)$$

Accordingly, we define the Hellinger distance for the conditional likelihood as

$$HD(\phi, \mathbf{Q}) \equiv 2\left(1 - \mathcal{H}_c(\phi, \mathbf{Q})\right), \quad \text{where}$$

$$\mathcal{H}_c(\phi, \mathbf{Q}) \equiv \int_{\mathbf{Y}^T} p^{1/2}(\mathbf{Y}^T|D_N = 1, \phi, \mathbf{Q}) \cdot p^{1/2}(\mathbf{Y}^T|D_N = 1, \phi_0, \mathbf{Q}_0) d\mathbf{Y}^T. \quad (27)$$

The next proposition analyzes the conditions for $\mathcal{H}(\phi, \mathbf{Q}) = 1$ and $\mathcal{H}_c(\phi, \mathbf{Q}) = 1$, and shows that the observational equivalence between $(\phi, \mathbf{Q})$ and $(\phi_0, \mathbf{Q}_0)$ boils down to the geometric equivalence of the set of reduced-form residuals satisfying the narrative restrictions, $\{\mathbf{U} : N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\}$.

Proposition 3.1. *Let $(\phi_0, \mathbf{Q}_0)$ be the true parameter value. Define*

$$\mathcal{Q}^* \equiv \left\{ \begin{array}{l} \mathbf{Q} \in \mathcal{O}(n): \quad \{\mathbf{U} : N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\} = \{\mathbf{U} : N(\phi_0, \mathbf{Q}_0, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\} \text{ up to } f(\mathbf{Y}^T|\phi_0)\text{-null set,} \\ \qquad\qquad\qquad\qquad\qquad \text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1} \end{array} \right\}.$$

The unconditional likelihood model (24) and the conditional likelihood model (26) are globally identified (i.e., no observationally equivalent parameter points to $(\phi_0, \mathbf{Q}_0)$) if and only if $\mathcal{Q}^*$ is a singleton. If the parameter of interest is an impulse response to the jth structural shock, $\eta_{i,j,h}(\phi, \mathbf{Q}) = \mathbf{c}'_{i,h}(\phi)\mathbf{q}_j$, as defined in (17), then $\eta_{i,j,h}(\phi, \mathbf{Q})$ is point-identified if the projection of $\mathcal{Q}^*$ onto the jth column vector is a singleton.

Proof. See Appendix B. □

This proposition provides a necessary and sufficient condition for global identification

of SVARs by NR. As shown in the proof in Appendix B, $\mathcal{Q}^*$ defined in this proposition corresponds to the observationally equivalent $\mathbf{Q}$ matrices given $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, but, importantly, it does not correspond to any flat region of the observed likelihood (conditional identified sets in Definition 4.1 below).

To illustrate this point, consider the simple bivariate example of Section 2 with the NR (3), where $\mathbf{y}_t$ itself is the reduced-form error, so $\mathbf{U}$ in Proposition 3.1 can be set to $\mathbf{y}_k$. Given $\boldsymbol{\phi}_0 = (\sigma_{11}, \sigma_{21}, \sigma_{22})$, the set of $\mathbf{y}_k \in \mathbb{R}^2$ satisfying the NR is the half-space given by

$$\left\{ \mathbf{y}_k \in \mathbb{R}^2 : (\sigma_{11}\sigma_{22})^{-1} \left(\sigma_{22}\cos\theta - \sigma_{21}\sin\theta, \quad \sigma_{11}\sin\theta \right) \mathbf{y}_k \geq 0 \right\}. \tag{28}$$

The condition for point-identification shown in Proposition 3.1 is satisfied if no $\theta' \neq \theta$ can generate the half-space of $\mathbf{y}_k$ identical to (28). Such θ' cannot exist, since a half-space passing through the origin $(a_1, a_2)\mathbf{y}_k \geq 0$ can be indexed uniquely by the slope a_1/a_2 and (28) implies the slope $\sigma_{11}^{-1}(\sigma_{22}(\tan\theta)^{-1} - \sigma_{21})$ is a bijective map of θ on a constrained domain due to the sign normalization restriction.

Figure 4 plots the Hellinger distances in this bivariate example under the shock-sign restriction (3) and the historical decomposition restriction. For both the conditional and unconditional likelihood, the Hellinger distances are minimized uniquely at the true θ, confirming our point-identification claim for θ.[14]

Figure 4: Hellinger Distance
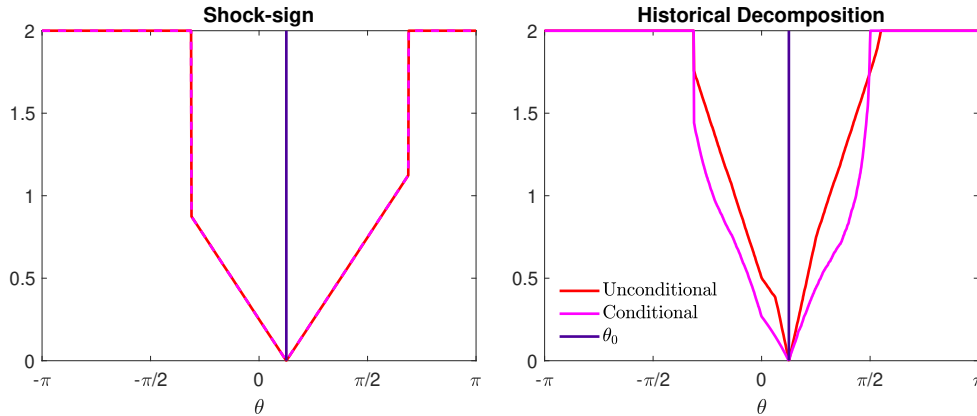
Notes: $T = 3$ and $\boldsymbol{\phi}$ is known; Hellinger distances are approximated using Monte Carlo.

Although a single NR can deliver point-identification of θ in the frequentist sense, the practical implication of this theoretical claim is not obvious. The observed unconditional

---

[14]For the historical decomposition case, a notable difference between the conditional and unconditional likelihood cases is in the slope of the Hellinger distance around the minimum. The Hellinger distance of the unconditional likelihood yields a steeper slope than the conditional likelihood. This indicates the loss of information for θ in the conditional likelihood due to conditioning on the non-ancillary event.

likelihood is almost always flat at the maximum, so we cannot obtain a unique maximum likelihood estimator for the structural parameter. As a result, the standard asymptotic approximation of the sampling distribution of the maximum likelihood estimator is not applicable. The SVAR model with NR possesses features of set-identified models from the Bayesian standpoint (i.e., flat regions of the likelihood). However, strictly speaking, it can be classified as a globally identified model in the frequentist sense when the condition of Proposition 3.1 holds.

4 Robust Bayesian Inference Under NR

This section explains how to conduct prior-robust Bayesian inference about a scalar-valued function of the structural parameters under narrative and traditional sign restrictions. The approach is an extension of that in GK18 with some modifications to account for the novel features of the NR. We assume that the object of interest is a particular impulse response η, although the discussion in this section also applies to any other scalar-valued function of the structural parameters, such as the forecast error variance decomposition or the historical decomposition.

We introduce our proposal primarily for the purpose of performing a global sensitivity analysis, where one wants to assess what posterior conclusion is robust to the choice of prior on the flat regions of the likelihood. We subsequently establish approximate frequentist validity of our robust Bayes proposal from the conditional frequentist perspective, assuming that the number of time periods in which the shock restrictions are imposed is small relative to the sample size.

4.1 Assessing posterior sensitivity

Let $\mathcal{Q}(\phi|S) = \{\mathbf{Q} : S(\phi, \mathbf{Q}) \geq \mathbf{0}_{\tilde{s} \times 1}, \text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}, \mathbf{Q} \in \mathcal{O}(n)\}$ represent the set of orthonormal matrices satisfying any traditional sign restrictions and the sign normalisation, and let π_{ϕ} be a prior over the reduced-form parameter ϕ. A joint prior for $\boldsymbol{\theta} = (\phi', \text{vec}(\mathbf{Q})')' \in \boldsymbol{\Phi} \times \text{vec}(\mathcal{O}(n))$ can be written as $\pi_{\boldsymbol{\theta}} = \pi_{\mathbf{Q}|\phi}\pi_{\phi}$, where $\pi_{\mathbf{Q}|\phi}$ is supported only on $\mathcal{Q}(\phi|S)$. When there are only traditional identifying restrictions, the prior for $\mathbf{Q}|\phi$ is not updated by the data, because the likelihood function is not a function of $\mathbf{Q}$. Posterior inference may therefore be sensitive to the choice of conditional prior, even asymptotically. As discussed above, a similar issue arises under NR. The difference is that the prior is updated by the data through the truncation points of the likelihood. However, within the set of values of $\mathbf{Q}$ given ϕ and satisfying the NR, the likelihood is flat. Consequently, the posterior

for $\mathbf{Q}$ given $\boldsymbol{\phi}$ and satisfying the NR is proportional to the prior at each $\boldsymbol{\phi}$.

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

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}} = \left\{ \pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}(\mathcal{Q}(\boldsymbol{\phi}|S)) = 1 \right\}. \tag{29}$$

Notice that we cannot impose the NR using a particular conditional prior on $\mathbf{Q}|\boldsymbol{\phi}$. This is because the NR generate a mapping from the reduced-form parameters to the structural parameters that depends on the realisation of the data independently of the reduced-form parameters, and the prior clearly cannot depend on the realisation of the data. However, by considering all possible priors for $\mathbf{Q}|\boldsymbol{\phi}$ that are consistent with the traditional identifying restrictions, we can trace out all possible posteriors for $\mathbf{Q}|\boldsymbol{\phi}, \mathbf{Y}^T$ that are consistent with the traditional identifying restrictions *and* the NR. This is because the NR truncate the likelihood function and the traditional identifying restrictions truncate the prior for $\mathbf{Q}|\boldsymbol{\phi}$, so the posterior for $\mathbf{Q}|\boldsymbol{\phi}, \mathbf{Y}^T$ is supported only on the values of $\mathbf{Q}$ that satisfy both sets of restrictions.

Let $\pi_{\mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q}}$ represent the unconditional likelihood of observing the data and the NR being satisfied. From Equation (23), it is clear that $\pi_{\mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q}} = \pi_{\mathbf{Y}^T | \boldsymbol{\phi}} 1(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1})$. Given a particular joint prior for $\boldsymbol{\theta}$, the joint posterior is thus

$$\pi_{\boldsymbol{\theta} | \mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}} \propto \pi_{\mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1} | \boldsymbol{\phi}, \mathbf{Q}} \pi_{\mathbf{Q} | \boldsymbol{\phi}} \pi_{\boldsymbol{\phi}}$$
$$\propto \pi_{\mathbf{Y}^T | \boldsymbol{\phi}} \pi_{\boldsymbol{\phi}} \pi_{\mathbf{Q} | \boldsymbol{\phi}} 1(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1})$$
$$\propto \pi_{\boldsymbol{\phi} | \mathbf{Y}^T} \pi_{\mathbf{Q} | \boldsymbol{\phi}} 1(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}).$$

The first line applies Bayes' rule and decomposes the joint prior for $\boldsymbol{\theta}$ into a prior for $\boldsymbol{\phi}$ and a conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$, the second line uses Equation (23) and the third line applies Bayes' rule again. The final expression for the posterior makes it clear that any prior for $\mathbf{Q}|\boldsymbol{\phi}$ that is consistent with the traditional identifying restrictions is in effect further truncated by the NR (through the likelihood) once the data are realised. Generating this posterior using every prior within the class of priors for $\mathbf{Q}|\boldsymbol{\phi}$ generates a class of posteriors for $\boldsymbol{\theta}$:

$$\Pi_{\boldsymbol{\theta} | \mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}} = \left\{ \pi_{\boldsymbol{\theta} | \mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}} = \pi_{\boldsymbol{\phi} | \mathbf{Y}^T} \pi_{\mathbf{Q} | \boldsymbol{\phi}} 1(N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}) : \pi_{\mathbf{Q} | \boldsymbol{\phi}} \in \Pi_{\mathbf{Q} | \boldsymbol{\phi}} \right\}. \tag{30}$$

Marginalising each posterior for $\boldsymbol{\theta}$ in this class of posteriors induces a class of posteriors for η, $\Pi_{\eta | \mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}}$. Each prior within the class of priors $\Pi_{\mathbf{Q} | \boldsymbol{\phi}}$ therefore induces a posterior for η. Associated with each of these posteriors will be quantities such as the posterior mean, median and other quantiles. For example, as we consider each possible prior within $\Pi_{\mathbf{Q}|\boldsymbol{\phi}}$,

we trace out the set of all possible posterior means for η. This will always be an interval, so we can summarise this 'set of posterior means' by its endpoints:

$$\left[\int_{\boldsymbol{\Phi}} l(\boldsymbol{\phi}, \mathbf{U}) d\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}, \int_{\boldsymbol{\Phi}} u(\boldsymbol{\phi}, \mathbf{U}) d\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}\right], \tag{31}$$

where $l(\boldsymbol{\phi}, \mathbf{U}) = \inf\{\eta(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \tilde{\mathcal{Q}}(\boldsymbol{\phi}, \mathbf{U}|N, S)\}$, $u(\boldsymbol{\phi}, \mathbf{U}) = \sup\{\eta(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \tilde{\mathcal{Q}}(\boldsymbol{\phi}, \mathbf{U}|N, S)\}$ and

$$\tilde{\mathcal{Q}}(\boldsymbol{\phi}, \mathbf{U}|N, S) = \{\mathbf{Q} : N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}, \mathbf{Q} \in \mathcal{Q}(\boldsymbol{\phi}|S)\} \tag{32}$$

is the set of values of $\mathbf{Q}$ that are consistent with the traditional identifying restrictions *and* the NR. In contrast, in GK18 the set of posterior means is obtained by finding the infimum and supremum of $\eta(\boldsymbol{\phi}, \mathbf{Q})$ over $\mathcal{Q}(\boldsymbol{\phi}|S)$. The important difference from GK18 is that the current set of posterior means depends on the data not only through the posterior for $\boldsymbol{\phi}$ but also through the set of admissible values of $\mathbf{Q}$ incorporating the NR. As a result, being different from GK18, we cannot interpret the set of posterior means (31) as a consistent estimator for the identified set for η (which is not well-defined, as we discuss below). Nevertheless, the set of posteriors means (31) still carries a robust Bayes interpretation similar to GK18 such that it clarifies posterior results that are robust to a choice of prior on the non-updated part of the parameter space (i.e., on the flat regions of the likelihood).

As in GK18, we can also report a robust credible region with credibility level α, which is the shortest interval estimate for η such that the posterior probability put on the interval is greater than or equal to α uniformly over the posteriors in $\Pi_{\eta|\mathbf{Y}^T, N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}}$ (see Proposition 1 of GK18). One may also be interested in posterior lower and upper probabilities, which are the infimum and supremum, respectively, of the probability for a hypothesis over all posteriors in the class. GK18 provide conditions under which their robust Bayesian approach has a valid frequentist interpretation, in the sense that the robust credible region is an asymptotically valid confidence set for the true identified set. For the same reason as mentioned above, however, frequentist validity of the robust credible region does not immediately extend to the NR case.

4.2 Conditional identified set

To formalize a certain frequentist validity of the robust Bayes credible region under NR, we require an important refinement of the concept of an identified set. Specifically, we introduce the 'conditional identified set', which extends the standard identified set in set-identified SVARs to the setting where identification is defined in a repeated sampling experiment conditional on the set of observations entering the NR.

Let $(\mathbf{Y}, \mathcal{Y})$ and $(\boldsymbol{\Theta}, \mathcal{A})$ be measurable spaces of a sample $\mathbf{Y}^T \in \mathbf{Y}$ and a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, respectively. Assume that the conditional distribution of $\mathbf{Y}^T$ given $\boldsymbol{\theta}$ exists and has a probability density $p(\mathbf{y}^T|\boldsymbol{\theta})$ at every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with respect to a σ-finite measure on $(\mathbf{Y}, \mathcal{Y})$, where $\mathbf{y}^T$ indicates a realisation of $\mathbf{Y}^T$. Traditionally, set-identification of $\boldsymbol{\theta}$ occurs when there are multiple observationally equivalent values of $\boldsymbol{\theta}$, so that there exists $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ such that $p(\mathbf{y}^T|\boldsymbol{\theta}) = p(\mathbf{y}^T|\boldsymbol{\theta}')$ for every $\mathbf{y}^T \in \mathbf{Y}$ (e.g., Rothenberg 1971). Observational equivalence can be represented by a many-to-one function $g : (\boldsymbol{\Theta}, \mathcal{A}) \to (\boldsymbol{\Phi}, \mathcal{B})$ such that $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}')$ if and only if $p(\mathbf{y}^T|\boldsymbol{\theta}) = p(\mathbf{y}^T|\boldsymbol{\theta}')$ for every $\mathbf{y}^T \in \mathbf{Y}$. $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ is the 'reduced-form' parameter, which carries all the information about the structural parameter $\boldsymbol{\theta}$ contained in the data. The identified set for $\boldsymbol{\theta}$ is then the inverse image of $g(.)$, $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}) = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : g(\boldsymbol{\theta}) = \boldsymbol{\phi}\}$.

The complication in applying this definition of the identified set in SVARs when there are NR is that the reduced-form parameters no longer represent all information about the structural parameters contained in the data; by truncating the likelihood, the realisations of the data entering the NR contain additional information about the structural parameters. To address this, we introduce a refinement of the definition of an identified set.

Definition 4.1. *Let $N(\boldsymbol{\theta}, \mathbf{y}^T) \geq \mathbf{0}_{s \times 1}$ represent a set of NR in terms of the structural parameters and the data.*

*(i) **Conditional identified set for $\boldsymbol{\theta}$ under NR** is*

$$IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{y}^T) = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : g(\boldsymbol{\theta}) = \boldsymbol{\phi}, N(\boldsymbol{\theta}, \mathbf{y}^T) \geq \mathbf{0}_{s \times 1}\}, \tag{33}$$

where $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ maps the structural parameters to the reduced-form parameters; i.e., $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}')$ if and only if $p(\mathbf{y}^T|\boldsymbol{\theta}) = p(\mathbf{y}^T|\boldsymbol{\theta}')$ for every $\mathbf{y}^T \in \mathbf{Y}$. The conditional identified set for impulse response $\eta = h(\boldsymbol{\theta})$ under NR is defined by projecting $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{y}^T)$ via $h(\boldsymbol{\theta})$,

$$IS_{\eta}(\boldsymbol{\phi}, \mathbf{y}^T) = \{\eta = h(\boldsymbol{\theta}) : \boldsymbol{\theta} \in IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{y}^T)\}. \tag{34}$$

*(ii) Let $\mathbf{s} : \mathbf{Y} \to \mathbb{R}^S$ be a statistic. We call $\mathbf{s}(\mathbf{Y}^T)$ a **sufficient statistic for the conditional identified set** $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{y}^T)$ if the conditional identified set for $\boldsymbol{\theta}$ depends on sample $\mathbf{y}^T$ through $\mathbf{s}(\mathbf{y}^T)$; i.e., there exists $\widetilde{IS}_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \cdot)$ such that*

$$IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{y}^T) = \widetilde{IS}_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{s}(\mathbf{y}^T)) \tag{35}$$

holds for all $\boldsymbol{\phi} \in \Phi$ and $\mathbf{y}^T \in \mathbf{Y}$.

Unlike the standard identified set $IS_{\boldsymbol{\theta}}(\phi)$, the conditional identified set under NR $IS_{\boldsymbol{\theta}}(\phi, \mathbf{y}^T)$

depends on the sample $\mathbf{y}^T$ because of the aforementioned data-dependent support of the likelihood. In terms of the observed likelihood, however, they share the property that the likelihood is flat on the identified set. Hence, given the sample $\mathbf{y}^T$ and the reduced-form parameters $\boldsymbol{\phi}$, any structural parameter values in $IS_{\boldsymbol{\theta}}(\phi, \mathbf{y}^T)$ fit the data equally well and, in this particular sense, they are observationally equivalent.

When NR concern shocks in only a subset of the time periods in the data, the conditional identified set under these NR depends on the sample only through a few observations entering the narrative restrictions. The sufficient statistics $\mathbf{s}(\mathbf{y}^T)$ defined in Definition 4.1 (ii) represent such observations. For instance, in the toy example of Section 2.1 with a single NR (3), the conditional identified set depends only on the observations at period $k \in \{1, \ldots, T\}$, so $\mathbf{s}(\mathbf{y}^T) = \mathbf{y}_k$. If we extend the example of Section 2.1 to the SVAR(p), the narrative sign restriction (3) can be expressed as

$$\varepsilon_{1k} = \mathbf{e}'_{1,2}\mathbf{A}_0\mathbf{u}_k = \mathbf{e}'_{1,2}\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{y}_k - \mathbf{B}\mathbf{x}_k) \geq 0. \tag{36}$$

Hence, the conditional identified set $IS_{\boldsymbol{\theta}}(\phi, \mathbf{y}^T)$ depends on the data only through $(\mathbf{y}'_k, \mathbf{x}'_k)' = (\mathbf{y}'_k, \mathbf{y}'_{k-1}, \cdots, \mathbf{y}'_{k-p})'$, so we can set $\mathbf{s}(\mathbf{y}^T) = (\mathbf{y}'_k, \mathbf{y}'_{k-1}, \cdots, \mathbf{y}'_{k-p})'$.

If the conditional distribution of $\mathbf{Y}^T$ given $\mathbf{s}(\mathbf{Y}^T) = \mathbf{s}(\mathbf{y}^T)$ is nondegenerate, we can consider a frequentist experiment (repeated sampling of $\mathbf{Y}^T$) conditional on the sufficient statistics set to the observed value. In this conditional experiment, we can view the conditional identified set $\widetilde{IS}_{\boldsymbol{\theta}}(\phi, \mathbf{s}(\mathbf{y}^T))$ as the standard identified set in set-identified models since it no longer depends on the data in the conditional experiment where $\mathbf{s}(\mathbf{y}^T)$ is fixed. This is the reason that we refer to $IS_{\boldsymbol{\theta}}(\phi, \mathbf{y}^T)$ as the conditional identified set. In Section 4.3 below, we show the frequentist validity of the robust-Bayes credible region by establishing the conditional coverage of the conditional identified set for an impulse response.

4.3 Frequentist coverage under a few narrative restrictions

In this section, we show that the robust Bayes credible regions attain asymptotically valid frequentist coverage in the setting that the number of NR is small relative to the length of the sampled periods in a sense that we make precise in the next assumption.

Assumption 4.1. (fixed-dimensional $\mathbf{s}(\mathbf{y}^T)$): The conditional identified set under NR have sufficient statistics $\mathbf{s}(\mathbf{y}^T)$, as defined in Definition 4.1 (ii), and the dimension of $\mathbf{s}(\mathbf{y}^T)$ does not depend on T.

Let $\boldsymbol{\theta}_0$ be the true structural parameters and $\boldsymbol{\phi}_0 = g(\boldsymbol{\theta}_0)$ be the corresponding reduced-form parameters. We view the sample $\mathbf{Y}^T$ being drawn from $p(\mathbf{Y}^T|\boldsymbol{\phi}_0)$. Let $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$ be

the conditional distribution of sample $\mathbf{Y}^T$ given the sufficient statistics for the conditional identified set $\mathbf{s} = \mathbf{s}(\mathbf{Y}^T)$ at the reduced-form parameters set to the truth $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. We denote by $p(\mathbf{s}|\boldsymbol{\phi}_0)$ the distribution of the sufficient statistics $\mathbf{s}(\mathbf{Y}^T)$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. The next assumption assumes that in the conditional experiment given $\mathbf{s}(\mathbf{Y}^T)$, the sampling distribution for the maximum likelihood estimator $\hat{\boldsymbol{\phi}} \equiv \arg\max_\phi p(\mathbf{Y}^T|\phi)$ centered at $\boldsymbol{\phi}_0$ and the posterior for ϕ centered at $\hat{\boldsymbol{\phi}}$ asymptotically coincide.

Assumption 4.2. (Conditional Bernstein-von Mises property for ϕ): For $p(\mathbf{s}|\boldsymbol{\phi}_0)$-almost every $\mathbf{s}$ and $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$-almost every sampling sequence $\mathbf{Y}^T$, the posterior distribution for $\sqrt{T}(\phi - \hat{\boldsymbol{\phi}})$ asymptotically coincides with the sampling distribution of $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ with respect to $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$, as $T \to \infty$, in the sense stated in Assumption 5 (i) in GK18.

This is a key assumption for establishing the asymptotic frequentist validity of the robust credible region under NR. It holds, for instance, when $\mathbf{s}(\mathbf{y}^T)$ corresponds to one or a few observations in the whole sample, as we had in the toy example of Section 2.1. In this case, the influence of $\mathbf{s}(\mathbf{y}^T)$ vanishes in the conditional sampling distribution of $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ as $T \to \infty$, as the latter asymptotically agrees with the asymptotically normal sampling distribution for the maximum likelihood estimator with the variance-covariance matrix given by the inverse of the Fisher information matrix. By the well-known Bernstein-von Mises theorem for regular parametric models, the posterior distribution for $\sqrt{T}(\phi - \hat{\boldsymbol{\phi}})$ asymptotically agrees with this sampling distribution.

The last assumption we require is convexity and smoothness of the conditional identified set, analogous to Assumption 5 (ii) of GK18 for standard partially identified models.

Assumption 4.3. (Almost-sure convexity and smoothness of the impulse response identified set): Let $\widetilde{IS}_\eta(\boldsymbol{\phi}, \mathbf{s}(\mathbf{Y}^T))$ be the conditional identified set for $\eta = h(\boldsymbol{\theta})$ with the sufficient statistics $\mathbf{s}(\mathbf{Y}^T)$. For $p(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-almost every $\mathbf{Y}^T$, $\widetilde{IS}_\eta(\boldsymbol{\phi}, \mathbf{s}(\mathbf{y}^T))$ is closed and convex, $\widetilde{IS}_\eta(\boldsymbol{\phi}, \mathbf{s}(\mathbf{y}^T)) = [\tilde{\ell}(\boldsymbol{\phi}, \mathbf{s}(\mathbf{Y}^T)), \tilde{u}(\boldsymbol{\phi}, \mathbf{s}(\mathbf{Y}^T))]$, and its lower and upper bounds are differentiable in $\boldsymbol{\phi}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ with nonzero derivatives.

Propositions 4.1–4.3 below present primitive conditions for Assumption 4.3. Imposing these three assumptions altogether, we obtain the following theorem.

Theorem 4.4. *For $\gamma \in (0, 1)$, let $\widehat{C}_\alpha^*$ be the volume-minimizing robust credible region for η*

with credibility α,[15] *which satisfies*

$$\inf_{\pi \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}^T, N(\boldsymbol{\theta}, \mathbf{Y}^T) \geq \mathbf{0}_{s \times 1}}} \pi(\widehat{C}_{\alpha}^*) = \pi_{\boldsymbol{\phi}}(IS_{\eta}(\boldsymbol{\phi}, \mathbf{Y}^T) \subset \hat{C}_{\alpha}^* | \mathbf{Y}^T, N(\boldsymbol{\theta}, \mathbf{Y}^T) \geq \mathbf{0}_{s \times 1}) = \alpha. \qquad (37)$$

Under Assumptions 4.1, 4.2, and 4.3, $\widehat{C}_{\alpha}^*$ *attains asymptotically valid coverage for the true impulse response conditional on* $\mathbf{s}(\mathbf{Y}^T)$.

$$\liminf_{T \to \infty} P_{\mathbf{Y}^T | \mathbf{s}, \boldsymbol{\phi}}(\eta_0 \in \widehat{C}_{\alpha}^* | \mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0) \geq \lim_{T \to \infty} P_{\mathbf{Y}^T | \mathbf{s}, \boldsymbol{\phi}}(\widetilde{IS}_{\eta}(\boldsymbol{\phi}_0, \mathbf{s}(\mathbf{Y}^T)) \subset \widehat{C}_{\alpha}^* | \mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0) = \alpha.$$
$$(38)$$

Accordingly, $\widehat{C}_{\alpha}^*$ *attains an asymptotically valid coverage for* η_0 *unconditionally,*

$$\liminf_{T \to \infty} P_{\mathbf{Y}^T | \boldsymbol{\phi}}(\eta_0 \in \widehat{C}_{\alpha}^* | \boldsymbol{\phi}_0) \geq \lim_{T \to \infty} P_{\mathbf{Y}^T | \boldsymbol{\phi}}(\widetilde{IS}_{\eta}(\boldsymbol{\phi}_0, \mathbf{s}(\mathbf{Y}^T)) \subset \widehat{C}_{\alpha}^* | \boldsymbol{\phi}_0) = \alpha. \qquad (39)$$

Proof. See Appendix B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

This theorem shows that the robust credible region of GK18 applied to the SVAR model with NR attains asymptotically valid frequentist coverage for the true impulse response as well as the conditional impulse response identified set. Even if the point-identification condition of Proposition 3.1 holds for the impulse response, it is not obvious if the standard Bayesian credible region can attain frequentist coverage. This is because the Bernstein-von Mises theorem does not seem to hold for the impulse response due to the non-standard features of the models with NR.

In what follows, we present sufficient conditions for convexity, continuity, and differentiability (both in $\boldsymbol{\phi}$) of the conditional impulse response identified set under the assumption that there is a fixed number of shock-sign restrictions constraining the first structural shock only (possibly in multiple periods). The proofs are collected in Appendix B.

Proposition 4.1. *Convexity.* *Let the parameter of interest be* $\eta_{i,1,h}$, *the impulse response of the ith variable at the hth horizon to the first structural shock. Assume that there are shock-sign restrictions on* $\varepsilon_{1,t}$ *for* $t = t_1, \ldots, t_K$, *so* $N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) = (\boldsymbol{\Sigma}_{tr}^{-1} \mathbf{u}_{t_1}, \ldots, \boldsymbol{\Sigma}_{tr}^{-1} \mathbf{u}_{t_K})' \mathbf{q}_1 \geq \mathbf{0}_{K \times 1}$. *Then the set of values of* $\eta_{i,1,h}$ *satisfying the shock-sign restrictions and sign normalisation,* $\{\eta_{i,1,h}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{c}_{i,h}(\boldsymbol{\phi}) \mathbf{q}_1 : N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{K \times 1}, \operatorname{diag}(\mathbf{Q}' \boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}, \mathbf{Q} \in \mathcal{O}(n)\}$ *is*

---

[15]The volume-minimizing robust credible region $\widehat{C}_{\alpha}^*$ is defined as a shortest interval among the connected intervals C_{α} satisfying

$$P_{\mathbf{Y}^T | \mathbf{s}, \boldsymbol{\phi}}(\widetilde{IS}_{\eta}(\boldsymbol{\phi}_0, \mathbf{s}(\mathbf{Y}^T)) \subset C_{\alpha} | \mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0) \geq \alpha.$$

See Proposition 1 in GK18 for a procedure to compute the volume-minimizing credible region.

convex for all i and h if there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ satisfying

$$\begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \dots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K})' \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(K+1)\times 1}. \tag{40}$$

Proposition 4.2. ***Continuity.*** *Let the parameter of interest and restrictions be as in Proposition 4.1, and assume that the conditions in the proposition are satisfied. If there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ such that, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$,*

$$\begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \dots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K})' \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} >> \mathbf{0}_{(K+1)\times 1}, \tag{41}$$

then $u(\boldsymbol{\phi}, \mathbf{U})$ and $l(\boldsymbol{\phi}, \mathbf{U})$ are continuous at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ for all i and h.[16]

Proposition 4.3. ***Differentiability.*** *Let the parameter of interest and restrictions be as in Proposition 4.1, and assume that the conditions in the proposition are satisfied. If, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the set of solutions to the optimisation problem*

$$\max_{\mathbf{q} \in \mathcal{S}^{n-1}} \left(\min_{\mathbf{q} \in \mathcal{S}^{n-1}} \right) \quad \mathbf{c}_{i,h}'(\boldsymbol{\phi})\mathbf{q} \quad s.t. \quad \left[(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \dots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K}), \; \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n} \right]' \mathbf{q} \geq \mathbf{0}_{(K+1)\times 1} \tag{42}$$

is singleton, the optimised value $u(\boldsymbol{\phi}, \mathbf{U})$ $(l(\boldsymbol{\phi}, \mathbf{U}))$ is nonzero, and the number of binding inequality restrictions at the optimum is at most $n-1$, then $u(\boldsymbol{\phi}, \mathbf{U})$ $(l(\boldsymbol{\phi}, \mathbf{U}))$ is almost-surely differentiable at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

5 Numerical Algorithms for Posterior Inference

In this section, we describe algorithms to conduct posterior inference under NR using the unconditional likelihood to construct the posterior. We first describe an algorithm that can be used to conduct posterior inference under a uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$. We then describe algorithms that can be used to conduct prior-robust Bayesian inference.

5.1 Single Prior

AR18 propose an algorithm for drawing from the uniform-normal-inverse-Wishart posterior of $\boldsymbol{\theta}$ given a set of traditional and NR. This is the posterior induced by a normal-inverse-Wishart prior over $\boldsymbol{\phi}$ and an unconditionally uniform prior over $\mathbf{Q}$. As discussed above, the

---

[16]For a vector $\mathbf{x} = (x_1, \dots, x_m)'$, $\mathbf{x} >> \mathbf{0}_{m \times 1}$ means that $x_i > 0$ for all $i = 1, \dots, m$.

likelihood that they use to construct the posterior is conditional on the NR holding. The algorithm proceeds by drawing ϕ from a normal-inverse-Wishart distribution and $\mathbf{Q}$ from a uniform distribution over $\mathcal{O}(n)$, and checking whether the traditional and NR are satisfied. If the restrictions are not satisfied, the joint draw is discarded and another draw is made. If the restrictions are satisfied, the probability that the NR are satisfied at the drawn parameter values (under the joint distribution of the restricted shocks) is approximated via Monte Carlo simulation. Once the desired number of draws are obtained satisfying the restrictions, the draws are resampled with replacement using as importance weights the inverse of the probability that the NR are satisfied.

As discussed above, using the conditional likelihood to construct the posterior places higher posterior probability on values of $\mathbf{Q}$ that yield a lower probability of the NR being satisfied, so we advocate using the *unconditional* likelihood to construct the posterior. The algorithm in AR18 essentially draws from the posterior under the unconditional likelihood and then uses importance sampling to draw from the posterior given the conditional likelihood. To draw from the uniform-normal-inverse-Wishart posterior using the unconditional likelihood to construct the posterior, one therefore simply needs to omit the importance-sampling step from this algorithm. Approximating the probability used to construct the importance weights requires Monte Carlo integration, which can be computationally expensive, particularly when the NR constrain the structural shocks in multiple periods. Omitting the importance-sampling step can therefore greatly ease the computational burden of drawing from the posterior.

By rejecting draws that do not satisfy the traditional restrictions and NR, the algorithm described above places more weight on draws of the reduced-form parameters that are less likely to satisfy the restrictions under the uniform distribution on $\mathbf{Q}$. As discussed in Uhlig (2017), one may instead prefer to use as a prior a distribution that is *conditionally* uniform over $\mathbf{Q}|\phi$.[17] Algorithm 1 below describes how to draw from the posterior of $\boldsymbol{\theta}$ given an arbitrary prior over ϕ and a conditionally uniform prior over $\mathbf{Q}|\phi$, using the unconditional likelihood to construct the posterior.

Algorithm 1. *Let $N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}$ be the set of narrative restrictions and let $S(\phi, \mathbf{Q}) \geq \mathbf{0}_{\tilde{s} \times 1}$ be the set of traditional sign restrictions (excluding the sign normalisation).*

- ***Step 1***: *Specify a prior for ϕ, π_ϕ, and obtain the posterior $\pi_{\phi|\mathbf{Y}^T}$.*

---

[17]Based on the results in Arias, Rubio-Ramírez and Waggoner (2018), AR18 argue that their algorithm draws from a normal-generalised-normal posterior distribution over the SVAR's structural parameters $(\mathbf{A}_0, \mathbf{A}_+)$ induced by a conjugate normal-generalised-normal prior, conditional on the NR. Practitioners who wish to draw from this posterior under the unconditional likelihood could simply omit the importance sampling step from the algorithm in AR18.

- **Step 2**: *Draw ϕ from $\pi_{\phi|\mathbf{Y}^T}$, compute the reduced-form VAR innovations $\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ for $t = t_1, \ldots, t_K$, and attempt to draw $\mathbf{Q}$ from the uniform distribution over $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$ using the subroutine below.*

  - **Step 2.1**: *Draw an $n \times n$ matrix of independent standard normal random variables, $\mathbf{Z}$, and let $\mathbf{Z} = \tilde{\mathbf{Q}}\mathbf{R}$ be the QR decomposition of $\mathbf{Z}$.*[18]

  - **Step 2.2**: *Define*

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

    where $\tilde{\mathbf{q}}_j$ is the jth column of $\tilde{\mathbf{Q}}$.

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

The algorithm relies on the fact that if $\mathbf{Q}|\phi$ is uniformly distributed over $\mathcal{Q}(\phi|S)$, then $\mathbf{Q}|\phi, \mathbf{Y}^T$ is also uniformly distributed over $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$. The single prior for $\mathbf{Q}|\phi$ induces a single posterior for $\boldsymbol{\theta}$ and thus a single posterior for any function of the structural parameters. One can obtain draws from the posterior of such an object by transforming the draws of $\boldsymbol{\theta}$.

When the restrictions substantially truncate $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$, it may take very many draws of $\mathbf{Q}$ from $\mathcal{O}(n)$ to obtain a single draw satisfying the restrictions. Amir-Ahmadi and Drautzburg (2019) introduce algorithms for inference in set-identified SVARs that may be useful in this case. Specifically, they propose checking whether the identified set is empty by solving a simple linear program; we discuss this approach below. They also suggest drawing $\mathbf{Q}$ directly from the space of orthonormal matrices satisfying the sign restrictions by using a Gibbs sampler, which avoids rejection sampling. However, these approaches are applicable only when there are linear inequality restrictions on $\mathbf{Q}$, which will not be the case when there are restrictions on the historical decomposition.

5.2 Multiple priors

GK18 propose numerical algorithms for conducting robust Bayesian inference in SVARs identified using traditional sign and zero restrictions. Their Algorithm 1 uses a numerical

---

[18]This is the algorithm used by Rubio-Ramírez, Waggoner and Zha (2010) to draw from the uniform distribution over $\mathcal{O}(n)$, except that we do not normalise the diagonal elements of $\mathbf{R}$ to be positive. This is because we impose a sign normalisation based on the diagonal elements of $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ in Step 2.2.

optimisation routine to obtain the lower and upper bounds of the identified set at each draw of ϕ. Obtaining the bounds via numerical optimisation is not generally applicable under the class of NR considered in AR18, since the constraints on the historical decomposition are not differentiable everywhere in $\mathbf{Q}$. We therefore adapt Algorithm 2 of GK18, which approximates the bounds of the identified set at each draw of ϕ using Monte Carlo simulation. After describing the algorithm, we discuss alternative algorithms that are more computationally efficient in some scenarios but are less-widely applicable.

Algorithm 2. *Let $N(\phi, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}$ be the set of narrative restrictions and let $S(\phi, \mathbf{Q}) \geq \mathbf{0}_{\tilde{s} \times 1}$ be the set of traditional sign restrictions (excluding the sign normalisation). Assume the object of interest is $\eta_{i,j^*,h} = c_{i,h}(\phi)' \mathbf{q}_{j^*}$.*

- ***Step 1**: Specify a prior for ϕ, π_ϕ, and obtain the posterior $\pi_{\phi|\mathbf{Y}^T}$.*

- ***Step 2**: Draw ϕ from $\pi_{\phi|\mathbf{Y}^T}$, compute the reduced-form VAR innovations $\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ for $t = t_1, \ldots, t_K$, and check whether $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$ is empty using Steps 2.1–2.3 of Algorithm 1. If $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$ is empty, repeat Step 2. Otherwise, proceed to Step 3.*

- ***Step 3**: Repeat Steps 2.1–2.3 of Algorithm 1 until K draws of $\mathbf{Q}$ are obtained. Let $\{\mathbf{Q}_k, k = 1, ..., K\}$ be the K draws of $\mathbf{Q}$ that satisfy the restrictions and let $\mathbf{q}_{j^*,k}$ be the j^*th column of $\mathbf{Q}_k$. Approximate $[l(\phi, \mathbf{U}), u(\phi, \mathbf{U})]$ by $[\min_l c_{i,h}(\phi)' \mathbf{q}_{j^*,l}, \max_l c_{i,h}(\phi)' \mathbf{q}_{j^*,l}]$.*

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

- ***Step 5**: To obtain an approximation of the smallest robust credible region with credibility $\alpha \in (0, 1)$, define $d(\eta, \phi, \mathbf{U}) = \max\{|\eta - l(\phi, \mathbf{U})|, |\eta - u(\phi, \mathbf{U})|\}$ and let $\hat{z}_\alpha(\eta)$ be the sample α-th quantile of $\{d(\eta, \phi_m, \mathbf{U}_m), m = 1, ..., M\}$. 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 $\min_\eta \hat{z}_\alpha(\eta)$.*

5.2.1 Remarks

Algorithm 2 approximates $[l(\phi, \mathbf{U}), u(\phi, \mathbf{U})]$ at each draw of ϕ via Monte Carlo simulation. The approximated set will lie strictly within the true set, but will converge to the true set as K goes to infinity. The algorithm may be computationally demanding when there are sign restrictions that substantially truncate $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$, because many draws of $\mathbf{Q}$ from $\mathcal{O}(n)$ may be rejected at each draw of ϕ. However, the same draws of $\mathbf{Q}$ can be used to compute the posteriors of $l(\phi, \mathbf{U})$ and $u(\phi, \mathbf{U})$ for different objects of interest, which cuts

down on computation time. For example, the same draws of $\mathbf{Q}$ can be used to compute the impulse responses of all variables to all shocks at all horizons of interest. They can also be used to compute other parameters by replacing $\eta_{i,j^*,h}$ with some other function. For example, $\eta_{i,j^*,h}$ can be replaced with the forecast error variance decomposition or an element of $\mathbf{A}_0$.[19] η may also be a function of the data, so one can use the algorithm to conduct robust Bayesian inference about the historical decomposition or the structural shocks themselves in particular periods. Step 3 is parallelizable, so large reductions in computing time are possible by distributing computation across multiple processors. Other algorithms may be computationally more efficient than Algorithm 2 in particular cases. We discuss these below.

Other algorithms. Assume that the object of interest is an impulse response to the first structural shock. The upper bound of the set of admissible values for the horizon-h impulse response of the ith variable to this shock given ϕ and $\mathbf{U}$ is the value function associated with the optimisation problem

$$u(\phi, \mathbf{U}) = \max_{\mathbf{Q} \in \tilde{Q}(\phi, \mathbf{U}|N,S)} \mathbf{c}'_{i,h}(\phi)\mathbf{q}_1. \tag{43}$$

$l(\phi, \mathbf{U})$ is obtained by minimising the same objective function subject to the same constraints. When $N(\phi, \mathbf{Q}, \mathbf{U})$ and $S(\phi, \mathbf{Q})$ only constrain $\mathbf{q}_1$, applying the change of variables $\mathbf{x} = \mathbf{\Sigma}_{tr}\mathbf{q}_1$ yields the optimisation problem in Gafarov, Meier and Montiel-Olea (2018) with additional inequality restrictions that are functions of $\mathbf{U}$.[20] Given a set of active inequality restrictions, Gafarov *et al.* (2018) provide an analytical expression for the value function and solution of this optimisation problem. To find the bounds of the identified set, they compute these quantities for every possible combination of active restrictions and check which pair solves the optimisation problem. Since the bounds are computed analytically at each set of active restrictions, this algorithm is computationally inexpensive as long as there is not a very large number of inequality restrictions. However, if $N(\phi, \mathbf{Q}, \mathbf{U})$ contains restrictions on the historical decomposition, all columns of $\mathbf{Q}$ are (nonlinearly) constrained and the analytical results are longer applicable. Similarly, the approach is not applicable when there are shock-sign or shock-rank restrictions on different structural shocks, or traditional sign restrictions on other columns of $\mathbf{Q}$. This approach may also be prohibitively slow when there are shock-rank restrictions, since the number of sign restrictions may be of the same order

---

[19]Impulse responses to a unit shock – rather than a standard-deviation shock – can be computed as in Algorithm 3 of Giacomini, Kitagawa and Read (2019).

[20]In their notation, the jth structural shock at time t is

$$\mathbf{e}'_{j,n}(\mathbf{B}^{-1}\mathbf{u}_t) = (\mathbf{B}^{-1}\mathbf{u}_t)'\mathbf{e}_{j,n} = \mathbf{u}'_t(\mathbf{B}^{-1}\mathbf{\Sigma})^{-1}\mathbf{e}_{j,n} = \mathbf{u}'_t\mathbf{\Sigma}^{-1}\mathbf{B}_j,$$

where $\mathbf{B}_j = \mathbf{B}\mathbf{e}_{j,n}$ is the jth column of $\mathbf{B}$ ($\mathbf{A}_0^{-1}$ in our notation) and we have used that $\mathbf{B}' = \mathbf{B}^{-1}\mathbf{\Sigma}$.

as the sample size. At most $n-1$ inequality constraints may be active at an optimum of the program in (43), so the number of combinations of active constraints that must be checked when there are s constraints is $\sum_{k=0}^{n-1} \binom{s}{k}$. For example, in the empirical application below, when we consider a shock-rank restriction alongside traditional sign restrictions, there are 1.7515×10^{14} possible combinations of active restrictions to consider.

As mentioned above, Amir-Ahmadi and Drautzburg (2019) propose an algorithm to determine whether the set of admissible values for $\mathbf{Q}$ is nonempty without recourse to random sampling from $\mathcal{O}(n)$. This algorithm can be more accurate and efficient than the simulation-based approach used in Algorithms 1 and 2, but it is applicable only when the columns of $\mathbf{Q}$ are linearly restricted by sign restrictions, which is not the case when there are restrictions on the historical decomposition. However, practitioners may not always wish to impose restrictions on the historical decomposition. We therefore describe an algorithm that can be used to conduct robust Bayesian inference when there are shock-rank, shock-sign and/or traditional sign restrictions that restrict only the first column of $\mathbf{Q}$, and which makes use of the approach in Amir-Ahmadi and Drautzburg (2019) to determine whether $\tilde{\mathcal{Q}}_1(\boldsymbol{\phi}, \mathbf{U}|N, S)$ is nonempty.

Algorithm 3. *Let $N(\boldsymbol{\phi}, \mathbf{U})\mathbf{q}_1 \geq \mathbf{0}_{s \times 1}$ be the set of narrative restrictions and let $S(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{(\tilde{s}+1) \times 1}$ be the set of traditional sign restrictions including the sign normalisation. Also, assume the object of interest is $\eta_{i,1,h} = c_{i,h}(\boldsymbol{\phi})'\mathbf{q}_1$. Replace Steps 2 and 3 of Algorithm 2 with the following.*

- ***Step 2***: *Draw $\boldsymbol{\phi}$ from $\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}$, compute the reduced-form VAR innovations $\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ for $t = t_1, \ldots, t_K$, and check whether $\tilde{\mathcal{Q}}_1(\boldsymbol{\phi}, \mathbf{U}|N, S)$ is empty by using the following subroutine.*

    - ***2.1*** *Solve for the Chebychev center $\{R, \tilde{\mathbf{q}}\}$ of the set $\{\tilde{\mathbf{q}} : (N(\boldsymbol{\phi}, \mathbf{U})', S(\boldsymbol{\phi})')'\tilde{\mathbf{q}} \geq \mathbf{0}_{(s+\tilde{s}+1) \times 1}, |\tilde{q}_i| \leq 1, i = 1, \ldots, n\}$. If $R > 0$, $\tilde{\mathcal{Q}}_1(\boldsymbol{\phi}, \mathbf{U}|N, S)$ is nonempty, so proceed to Step 3. Otherwise, repeat Step 2.*

- ***Step 3***: *Compute $l(\boldsymbol{\phi}, \mathbf{U})$ by solving the following constrained optimisation problem with initial value $\mathbf{q}^0 = \tilde{\mathbf{q}}/\|\tilde{\mathbf{q}}\|$:*

$$l(\boldsymbol{\phi}) = \min_{\mathbf{q}} c_{i,h}(\boldsymbol{\phi})'\mathbf{q} \quad s.t. \quad N(\boldsymbol{\phi}, \mathbf{U})\mathbf{q}_1 \geq \mathbf{0}_{s \times 1}, S(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{(\tilde{s}+1) \times 1}, \mathbf{q}'\mathbf{q} = 1. \quad (44)$$

Similarly, obtain $u(\boldsymbol{\phi}, \mathbf{U})$ by maximising $c_{i,h}(\boldsymbol{\phi})'\mathbf{q}$ subject to the same set of constraints.

Step 2.1 requires solving for the Chebychev center of the set satisfying the narrative and traditional sign restrictions. The Chebychev center $\tilde{\mathbf{q}}$ is the center of the largest ball with

32

radius R that can be inscribed within the set $\{\tilde{\mathbf{q}} : (N(\boldsymbol{\phi}, \mathbf{U})', S(\boldsymbol{\phi})')'\tilde{\mathbf{q}} \geq \mathbf{0}_{(s+\tilde{s}+1)\times 1}, |\tilde{q}_i| \leq 1, i = 1, \dots, n\}$, which is the intersection of the halfspaces generated by the inequality restrictions and the unit n-cube.[21] Letting $\mathbf{Z}_k'$ be the kth row of $(N(\boldsymbol{\phi}, \mathbf{U})', S(\boldsymbol{\phi})')'$, the Chebychev center and radius can be obtained as the solution to the following problem (see, for example, Boyd and Vandenberghe (2004)):

$$\max_{\{R \geq 0, \tilde{\mathbf{q}}\}} R \tag{45}$$

subject to

$$\mathbf{Z}_k'\tilde{\mathbf{q}} + R\|\mathbf{Z}_k\| \geq 0, \quad k = 1, \dots, s + \tilde{s} + 1 \tag{46}$$

$$\tilde{q}_i + R \leq 1, \quad i = 1, \dots, n. \tag{47}$$

$$\tilde{q}_i - R \geq -1, \quad i = 1, \dots, n. \tag{48}$$

This is a linear program, which can be solved efficiently. If $R > 0$, then $\tilde{\mathcal{Q}}_1(\boldsymbol{\phi}, \mathbf{U}|N, S)$ is nonempty. If $\tilde{\mathbf{q}}$ is a Chebychev center with $R > 0$, then $\tilde{\mathbf{q}}$ satisfies the sign restrictions and $\|\tilde{\mathbf{q}}\| > 0$. $\mathbf{q}_1^0 = \tilde{\mathbf{q}}/\|\tilde{\mathbf{q}}\|$ then has unit norm and satisfies the sign restrictions, so we can use it as an initial value in the optimisation problem of Step 3. In practice, we solve the optimisation problem in Step 3 using an interior-point algorithm within Matlab's 'fmincon' optimiser. Algorithm 3 can also be used to conduct robust Bayesian inference about other objects by replacing the objective function in Step 3.

6 Empirical Application: The Dynamic Effects of a Monetary Policy Shock

AR18 estimate the effects of monetary policy shocks on the US economy using a combination of sign restrictions on impulse responses and NR. The reduced-form VAR is the same as that used in Christiano, Eichenbaum and Evans (1999) and Uhlig (2005). The model's endogenous variables are real GDP, the GDP deflator, a commodity price index, total reserves, non-borrowed reserves and the federal funds rate. The data are monthly and run from January 1965 to November 2007. The VAR includes 12 lags and we include a constant.

As NR, AR18 impose that the monetary policy shock in October 1979 was positive and that it was the overwhelming contributor to the unexpected change in the federal funds rate in that month (a Type B restriction on the historical decomposition). This was the

---

[21]The restriction that $\tilde{q}$ lies within the unit n-cube ensures that the problem is well-defined.

month in which the Federal Reserve markedly and unexpectedly increased the federal funds rate following the appointment of Paul Volcker as chairman of the Federal Reserve, and is widely considered to be an example of a positive monetary policy shock. The traditional sign restrictions considered in Uhlig (2005) are also imposed. Specifically, the response of the federal funds rate is restricted to be non-negative for $h = 0, 1, \ldots, 5$ and the responses of the GDP deflator, the commodity price index and nonborrowed reserves are restricted to be nonpositive for $h = 0, 1, \ldots, 5$.

We assume a Jeffreys' (improper) prior over the reduced-form parameters, $\pi_{\phi} = \pi_{\mathbf{B}, \mathbf{\Sigma}} \propto |\mathbf{\Sigma}|^{-\frac{n+1}{2}}$. The posterior for the reduced-form parameters, $\pi_{\phi|\mathbf{Y}^T}$, is then a normal-inverse-Wishart distribution, from which it is straightforward to obtain independent draws (for example, see Del Negro and Schorfheide (2011)). We obtain 1,000 draws from the posterior of ϕ such that the VAR is stable and $\tilde{Q}(\phi, \mathbf{U}|N, S)$ is non-empty. We use Algorithm 2 with $K = 10,000$ draws of $\mathbf{Q}$ at each draw of ϕ to approximate $l(\phi, \mathbf{U})$ and $u(\phi, \mathbf{U})$. If we cannot obtain a draw of $\mathbf{Q}$ satisfying the restrictions after 100,000 draws of $\mathbf{Q}$, we approximate $\tilde{Q}(\phi, \mathbf{U}|N, S)$ as being empty at that draw of ϕ.

We first consider the effect of using the conditional likelihood rather than the unconditional likelihood on posterior inference under a conditionally uniform prior for $\mathbf{Q}|\phi$. The posterior corresponding to the unconditional likelihood is obtained using Algorithm 1, while the posterior corresponding to the conditional likelihood is obtained by resampling the draws obtained using Algorithm 1, using as importance weights the probability that the NR hold at each draw of $\boldsymbol{\theta}$. This probability is approximated by Monte Carlo simulation with 10,000 replications.[22] For brevity, we report only the impulse responses of the federal funds rate and real GDP to a positive standard-deviation monetary policy shock (Figure 5). In this application, the posterior distributions of the two sets of impulse responses are quite similar regardless of whether the conditional or unconditional likelihood is used to construct the posterior.[23]

We then explore the sensitivity of posterior inference to the choice of prior for $\mathbf{Q}|\phi$ when the unconditional likelihood is used to construct the posterior. Figure 6 plots the full set of impulse responses obtained using both the conditionally uniform prior and the robust Bayesian approach. The results suggest that posterior inference about the effect of a monetary policy shock can be sensitive to the choice of prior for $\mathbf{Q}|\phi$. For example, under

---

[22]The importance weights imply that the effective sample size is 866 (see Arias, Rubio-Ramírez and Waggoner (2018)).

[23]The results are not directly comparable to those presented in Figure 6 of AR18. First, we present responses to a standard-deviation shock, whereas AR18 describe their responses as being to a 25 basis point shock (although, from close inspection of their Figure 6, it is evident that this normalisation is not imposed correctly, because the impact response of the federal funds rate fans out around zero). Second, we use a prior for $\mathbf{Q}$ that is conditionally uniform given ϕ, whereas AR18 use a prior that is unconditionally uniform.

Figure 5: Impulse Responses to a Monetary Policy Shock – Conditional vs Unconditional Likelihood
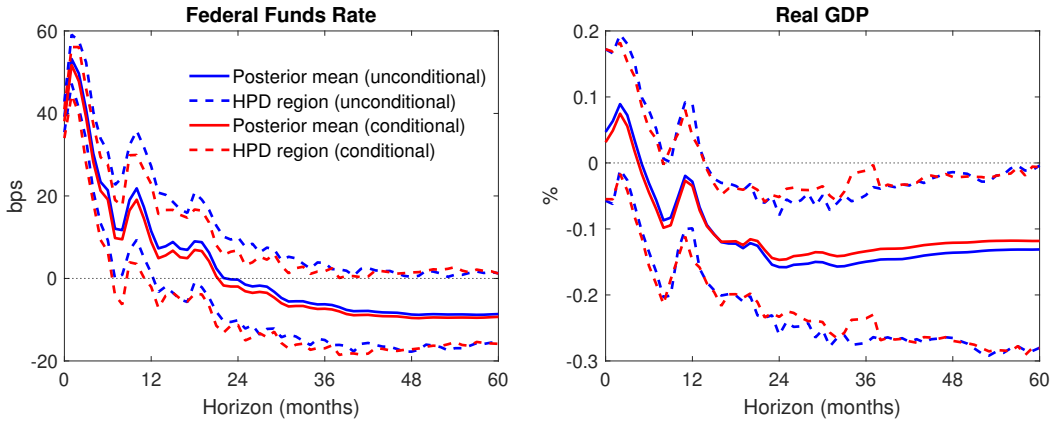
the conditionally uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$, the 68 per cent highest posterior density credible intervals for the response of real GDP exclude zero at horizons greater than a year or so, whereas the 68 per cent robust credible intervals include zero at all horizons. Under the single prior, the posterior probability that the shock results in output falling after two years is 95 per cent. In contrast, the posterior lower probability – the smallest probability over the class of posteriors generated by the class of priors – that output falls after two years is only 55 per cent. The choice of single prior shrinks the credible intervals by about 60 per cent on average across the variables and horizons considered.

AR18 also consider an alternative set of identifying restrictions. Specifically, they impose that the monetary policy shock was: positive in April 1974, October 1979, December 1988 and February 1994; negative in December 1990, October 1998, April 2001 and November 2002; and the most important contributor to the observed unexpected change in the federal funds rate in these months. The choice of these dates is based on a synthesis of information from different sources, including the chronology of monetary policy actions from Romer and Romer (1989), an updated series of the monetary policy shocks constructed using Greenbook forecasts in Romer and Romer (2004), the high-frequency monetary policy surprises from Gürkaynak, Sack and Swanson (2005), and minutes from Federal Open Markets Committee meetings. Building on this chronology, we impose a novel shock-rank restriction. Specifically, we impose that the monetary policy shock in October 1979 was the largest positive realisation of the monetary policy shock in the sample period. This restriction is motivated by the fact that the change in the federal funds rate in October 1979 was more positive than the change in the federal funds rate in the other periods identified by AR18 as containing notable

Figure 6: Impulse Responses to a Monetary Policy Shock – Single Prior vs Robust
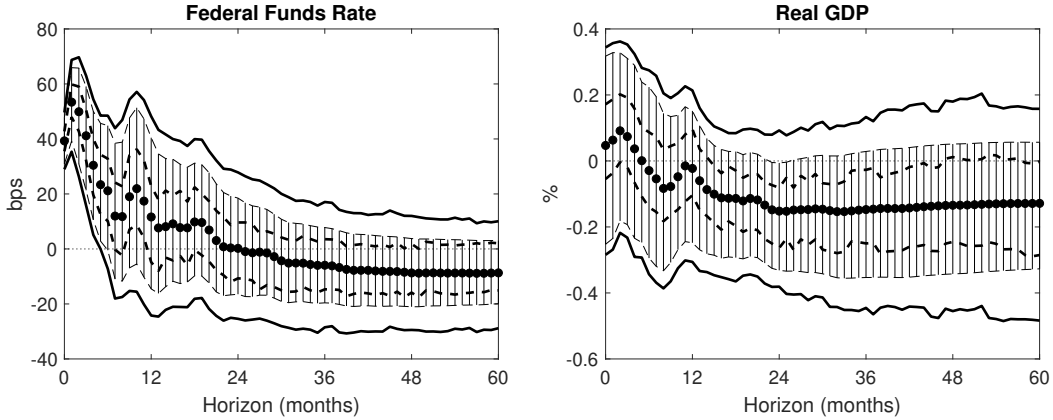
Notes: Circles and dashed lines are, respectively, posterior means and 68 per cent (point-wise) highest posterior density intervals under the uniform prior for $\mathbf{Q}|\phi$; vertical bars are sets of posterior means and solid lines are 68 per cent (point-wise) robust credible regions obtained using Algorithm 2 with 10,000 draws from $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$; results are based on 1,000 draws from the posterior of ϕ with nonempty $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$; impulse responses are to a standard-deviation shock.

monetary policy shocks (Table 1).[24]

Table 1: Monthly Change in Federal Funds Rate (ppt)

| Oct 79	Apr 74	Dec 88	Feb 94	Dec 90	Oct 98	Apr 01	Nov 02
2.34	1.16	0.41	0.2	−0.5	−0.44	−0.51	−0.41

Source: FRED

Figure 7 plots the estimates under these restrictions. Since the shock-rank restriction and the traditional sign restrictions constrain one column of $\mathbf{Q}$ only and generate a large number of inequality constraints, we use Algorithm 3 to obtain the results. For comparison, we also plot the results under the restrictions on the sign of the monetary policy shock and the historical decomposition in October 1979. The shock-rank restriction substantially shrinks the set of posterior means and robust credible regions relative to those obtained under the restriction on the historical decomposition. The posterior lower probability that output falls two years after the shock is 73 per cent, compared with 55 per cent under the restriction on

---

[24]The posterior mean of the reduced-form VAR innovation to the federal funds rate in October 1979 is also more positive than in the other periods.

the historical decomposition.[25]

Figure 7: Impulse Responses to a Monetary Policy Shock – Shock-rank Restriction
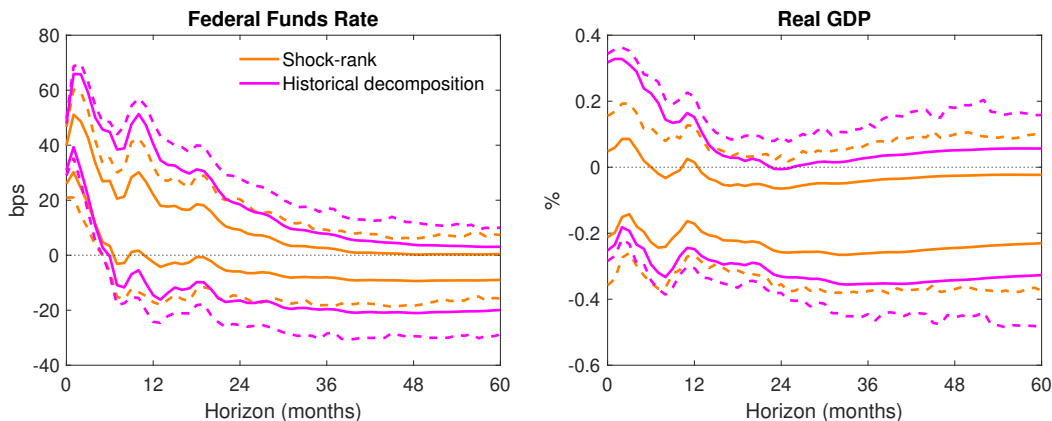
Notes: Solid lines represent set of posterior means and dashed lines represent 68 per cent (pointwise) robust credible regions; results are based on 1,000 draws from the posterior of ϕ with nonempty $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$; results under shock-rank restriction are obtained using Algorithm 3; results under restriction on the historical decomposition are obtained using Algorithm 2 with 10,000 draws from $\tilde{\mathcal{Q}}(\phi, \mathbf{U}|N, S)$; impulse responses are to a standard-deviation shock.

7   Conclusion

Restricting structural shocks to be consistent with the historical narrative offers a potentially useful approach to disciplining structural vector autoregressions, but raises new issues related to identification and inference. We show that these restrictions generate a set-valued mapping from the model's reduced-form parameters to its structural parameters that depends on the realisation of the data entering the restrictions independently of the reduced-form parameters. Conditioning on the restrictions holding may result in the posterior distribution placing more weight on parameters that yield a lower *ex ante* probability that the restrictions are satisfied. We therefore advocate using the unconditional likelihood when constructing the posterior. However, this likelihood will possess flat regions in any particular sample, which implies that the posterior will be proportional to the prior in some region of the parameter space. Posterior inference may therefore be sensitive to the choice of prior. To address this,

---

[25]In general, $\tilde{Q}(\phi, \mathbf{U}|N, S)$ may be empty at particular values of ϕ. The proportion of draws of ϕ where $\tilde{Q}(\phi, \mathbf{U}|N, S)$ is empty can therefore be used to assess the plausibility of the restrictions (see GK18). In this application, under the restriction on the historical decomposition, $\tilde{Q}(\phi, \mathbf{U}|N, S)$ is nonempty at every draw of ϕ, so the posterior plausibility of the restrictions is one. In contrast, under the shock-rank restriction, the posterior plausibility of the restrictions is only 17 per cent. The results in Figure 7 are largely unchanged when based on draws of ϕ satisfying both sets of restrictions.

we provide tools to assess posterior sensitivity to the choice of conditional prior over the orthonormal matrix. We provide conditions under which these tools will have a valid frequentist interpretation. In particular, given a fixed number of NR, these tools provide valid frequentist inference about the identified set under a refinement of the concept of identified set that allows the mapping from reduced-form to structural parameters to depend on the realisation of the data independently of the reduced-form parameters.

References

Amir-Ahmadi, P. and T. Drautzburg (2019). *Identification And Inference With Ranking Restrictions*. Unpublished manuscript.

Antolín-Díaz, J., I. Petrella and J.F. Rubio-Ramírez (In Press). "Structural Scenario Analysis with SVARs". *Journal of Monetary Economics*.

Antolín-Díaz, J. and J.F. Rubio-Ramírez (2018). "Narrative Sign Restrictions for SVARs". *American Economic Review* 108 (10), pp. 2802–29.

Arias, J.E., J.F. Rubio-Ramírez and D.F. Waggoner (2018). "Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications". *Econometrica* 86 (2), pp. 685–720.

Basu, A., H. Shioya and C. Park (2011). *Statistical Inference: The Minimum Distance Approach*. CRC Press.

Baumeister, C. and J.D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information". *Econometrica* 83 (5), pp. 1963–1999.

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.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

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

Christiano, L.J., M. Eichenbaum and C.L. Evans (1999). "Chapter 2 – Monetary Policy Shocks: What Have We Learned and to What End?" In: *Handbook of Macroeconomics*. Ed. by J.B. Taylor and M. Woodford. Vol. 1. Elsevier, pp. 65–148.

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.

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

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.

Giacomini, R. and T. Kitagawa (2018). "Robust Bayesian Inference for Set-identified Models". cemmap Working Paper CWP61/18.

Giacomini, R., T. Kitagawa and M. Read (2019). "Robust Bayesian Inference in Proxy SVARs". (CWP23/19).

Gürkaynak, R. S, B. Sack and E. Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". *International Journal of Central Banking* 1, pp. 55–93.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

Inoue, A. and L. Kilian (2020). "Joint Bayesian Inference about Impulse Responses in VAR Models". unpublished manuscript.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.

Kilian, L. and X. Zhou (2019). "Does Drawing Down the U.S. Strategic Petroleum Reserve Help Stabilize Oil Prices?". CEPR Discussion Paper No. DP13849.

— (2020). "Oil Prices, Exchange Rates and Interest Rates". unpublished manuscript.

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

Ludvigson, S.C., S. Ma and S. Ng (forthcoming). "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics*.

— (2018). "Shock Restricted Structural Vector-Autoregressions". National Bureau of Economic Research Working Paper No. 23225.

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.

Petterson, Marco, David Seim and Jesse Shapiro (2020). "Bounds on a Slope from Size Restrictions on Economic Shocks". *unpublished manuscript.*

Poirier, D.J. (1998). "Revising Beliefs in Nonidentified Models". *Econometric Theory* 14 (4), pp. 483–509.

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

Romer, C.D. and D.H. Romer (1989). "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz". In: *NBER Macroeconomics Annual*. Ed. by O.J. Blanchard and S. Fischer. Vol. 4. MIT Press, Cambridge MA, pp. 121–84.

— (2004). "A New Measure of Monetary Shocks: Derivation and Implications". *American Economic Review* 94 (4), pp. 1055–1084.

Rothenberg, T.J. (1971). "Identification in Parametric Models". *Econometrica* 39 (3), pp. 577–591.

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.

Sims, C.A. (1980). "Macroeconomics and Reality". *Econometrica* 48 (1), pp. 1–48.

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.

Uhlig, H. (2005). "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics* 52 (2), pp. 381–419.

— (2017). "Shocks, Sign Restrictions, and Identification". In: *Advances in Economics and Econometrics: Eleventh World Congress.* Ed. by B. Honoré A. Pakes, M. Piazzesi and L. Samuelson. Vol. 2. Cambridge University Press, pp. 95–127.

Zhou, X. (2020). "Refining the Workhorse Oil Market Model". *Journal of Applied Econometrics* 35 (1), pp. 130–140.

8 Appendix A: Bivariate Example

This section presents analytical expressions for the set of values of θ consistent with the NR in the bivariate example of Section 2. When determining the set of values of θ consistent with the restrictions, there are four main cases to consider, which differ in terms of the signs of σ_{21} and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k}$. Within each of these cases, the set of values of θ consistent with the restrictions also depends on the sign of y_{1k}.[26]

1. $\sigma_{21} > 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$

    (a) $y_{1k} > 0$:

    $$\theta \in \left[-\pi + \arctan\left(\max\left\{ \frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right\} \right), \right.$$
    $$\left. \arctan\left(\min\left\{ \frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right\} \right) \right]. \quad (49)$$

    (b) $y_{1k} < 0$:
    $$\theta \in \left[-\pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right) \right]. \quad (50)$$

2. $\sigma_{21} > 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} < 0$

    (a) $y_{1k} > 0$:
    $$\theta \in \left[\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right) \right]. \quad (51)$$

    (b) $y_{1k} < 0$:

    $$\theta \in \begin{cases} \left[\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right) \right] & \text{if } \frac{\sigma_{22}}{\sigma_{21}} > \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \\ \left[-\pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right), -\pi + \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right) \right] & \text{otherwise.} \end{cases}$$
    $$(52)$$

3. $\sigma_{21} < 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$

    (a) $y_{1k} > 0$:
    $$\theta \in \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \right) \right]. \quad (53)$$

---

[26]It may be useful to recall that $\arctan(x) \in (-\pi/2, \pi/2)$ and $-\arctan(x) = \arctan(-x)$.

(b) $y_{1k} < 0$:

$$\theta \in \begin{cases} \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right)\right] & \text{if } \frac{\sigma_{22}}{\sigma_{21}} < \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}} \\ \left[\pi + \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right), \pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right] & \text{otherwise.} \end{cases} \tag{54}$$

4. $\sigma_{21} < 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} < 0$

(a) $y_{1k} > 0$:

$$\theta \in \left[\arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right\}\right),\right.$$

$$\left. \pi + \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right\}\right)\right]. \tag{55}$$

(b) $y_{1k} < 0$:

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right), \pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]. \tag{56}$$

Under the restrictions that the first structural shock is positive in period k and was the most important (or overwhelming) contributor to the change in the first variable, θ is restricted to lie in the set

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

$$\left. |\sigma_{22}y_{1k}\cos^2\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\cos\theta\sin\theta| \geq |\sigma_{22}y_{1k}\sin^2\theta - (\sigma_{11}y_{2k} + \sigma_{21}y_{1k})\cos\theta\sin\theta|\right\}$$

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

$$\left. |\sigma_{22}y_{1k}\cos^2\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\cos\theta\sin\theta| \geq |\sigma_{22}y_{1k}\sin^2\theta - (\sigma_{11}y_{2k} + \sigma_{21}y_{1k})\cos\theta\sin\theta|\right\}. \tag{57}$$

As in the case of the shock-sign restriction, this set also depends on the data $\mathbf{y}_k$ independently of the reduced-form parameters.

9   Appendix B: Omitted Proofs

Proof of Proposition 3.1

Proof. $\mathcal{H}(\boldsymbol{\phi}, \mathbf{Q})$ can be written as

$$\mathcal{H}(\boldsymbol{\phi}, \mathbf{Q}) = \int_{\mathbf{Y}^T} f^{1/2}(\mathbf{Y}^T|\boldsymbol{\phi}) f^{1/2}(\mathbf{Y}^T|\boldsymbol{\phi}_0) \cdot D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{Y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T) d\mathbf{Y}^T$$
$$+ \int_{\mathbf{Y}^T} f^{1/2}(\mathbf{Y}^T|\boldsymbol{\phi}) f^{1/2}(\mathbf{Y}^T|\boldsymbol{\phi}_0) \cdot (1 - D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{Y}^T))(1 - D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)) d\mathbf{Y}^T.$$

Note that the likelihood for the reduced-form parameters $f(\mathbf{Y}^T|\boldsymbol{\phi})$ point-identifies $\boldsymbol{\phi}$, so $f(\cdot|\boldsymbol{\phi}) = f(\cdot|\boldsymbol{\phi}_0)$ holds only at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. Hence, we set $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ and consider $\mathcal{H}(\boldsymbol{\phi}_0, \mathbf{Q})$,

$$\mathcal{H}(\boldsymbol{\phi}_0, \mathbf{Q}) = \int_{\{\mathbf{Y}^T : D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)\}} f(\mathbf{Y}^T|\boldsymbol{\phi}_0) d\mathbf{Y}^T.$$

Hence, $\mathcal{H}(\boldsymbol{\phi}_0, \mathbf{Q}) = 1$ if and only if $D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)$ holds $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-a.s. In terms of the reduced-form residuals entering the narrative restrictions, the latter condition is equivalent to $\{\mathbf{U} : N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\} = \{\mathbf{U} : N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{U}) \geq \mathbf{0}_{s \times 1}\}$ up to $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-null set. Hence, $\mathcal{Q}^*$ defined in the proposition collects observationally equivalent values of $\mathbf{Q}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ in terms of the unconditional likelihood.

Next, consider the conditional likelihood and consider

$$\mathcal{H}_c(\boldsymbol{\phi}_0, \mathbf{Q}) = \frac{1}{r^{1/2}(\boldsymbol{\phi}, \mathbf{Q}) r^{1/2}(\boldsymbol{\phi}_0, \mathbf{Q}_0)} \int_{\mathbf{Y}^T} f(\mathbf{Y}^T|\boldsymbol{\phi}_0) \cdot D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{Y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T) d\mathbf{Y}^T$$
$$= \frac{E_{\mathbf{Y}^T|\boldsymbol{\phi}_0} \left[D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T) \right]}{r^{1/2}(\boldsymbol{\phi}, \mathbf{Q}) r^{1/2}(\boldsymbol{\phi}_0, \mathbf{Q}_0)}$$
$$\leq 1,$$

where the inequality follows by the Cauchy-Schwartz inequality, and it holds with equality if and only if $D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)$ holds $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-a.s. Hence, by repeating the argument for the unconditional likelihood case, we conclude that $\mathcal{Q}^*$ consists of observationally equivalent values of $\mathbf{Q}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ in terms of the conditional likelihood. $\square$

Proof of Theorem 4.4. Since $\boldsymbol{\theta}_0$ satisfies the imposed narrative restrictions $N(\boldsymbol{\theta}_0, \mathbf{y}^T) \geq \mathbf{0}$ and the other sign restrictions if any imposed, $\eta_0 \in \widetilde{IS}_\eta(\boldsymbol{\phi}_0, \mathbf{s}(\mathbf{y}^T))$ holds for any $\mathbf{y}^T$. Hence, for all T,

$$P_{\mathbf{Y}^T|\mathbf{s}, \boldsymbol{\phi}}(\eta_0 \in \widehat{C}_\alpha^* | \mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0) \geq P_{\mathbf{Y}^T|\boldsymbol{\phi}}(\widetilde{IS}_\eta(\boldsymbol{\phi}_0, \mathbf{s}(\mathbf{Y}^T)) \subset \widehat{C}_\alpha^* | \mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0). \tag{58}$$

Hence, to prove the claim, it suffices to focus on the asymptotic behavior of the coverage probability for the conditional identified set shown in the right-hand side.

Under Assumption 4.2 and 4.3, the asymptotically correct coverage for the conditional identified set can be obtained by applying Proposition 2 in GK18. □

Proof of Proposition 4.1. If there exists a unit-length vector $\mathbf{q}$ satisfying the inequality in (40), it must lie within the intersection of the K halfspaces defined by the inequalities $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_k})'\mathbf{q} \geq 0$, $k = 1, \ldots, K$, the halfspace defined by the sign normalisation, $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q} \geq 0$, and the unit sphere in $\mathbb{R}^n$. The intersection of these $K+1$ halfspaces and the unit sphere is a path-connected set. Since $\eta_{i,1,h}(\boldsymbol{\phi}, \mathbf{Q})$ is a continuous function of $\mathbf{q}_1$, the set of values of $\eta_{i,1,h}$ satisfying the restrictions is an interval and is thus convex, because the set of a continuous function with a path-connected domain is always an interval. □

Proof of Proposition 4.2. After noting that $\mathbf{U}$ is (implicitly) continuous in $\boldsymbol{\phi}$, continuity of $u(\boldsymbol{\phi}, \mathbf{U})$ and $l(\boldsymbol{\phi}, \mathbf{U})$ follows by the same logic as in the proof of Proposition 4 of GK18. We omit the detail for brevity. □

Proof of Proposition 4.3. One-to-one differentiable reparameterization of the optimisation problem in Equation 42 using $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$ yields the optimisation problem in Equation (2.5) of Gafarov *et al.* (2018) with a set of inequality restrictions that are now a function of the data through $\mathbf{U}$. Noting that $\mathbf{U}$ is (implicitly) differentiable in $\boldsymbol{\phi}$, differentiability of $u(\boldsymbol{\phi}, \mathbf{U})$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ follows from their Theorem 2 under the assumptions that, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the set of solutions to the optimisation problem is singleton, the optimized value $u(\boldsymbol{\phi}, \mathbf{U})$ is nonzero, and the number of binding sign restrictions at the optimum is at most $n - 1$. Differentiability of $l(\boldsymbol{\phi}, \mathbf{U})$ follows similarly. Note that Theorem 2 of Gafarov *et al.* (2018) additionally requires that the column vectors of $\left[(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \ldots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K}), \ \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n}\right]$ are linearly independent, but this occurs almost-surely under the probability law for $\mathbf{U}$. □